



---

**A construção do corpus no estudo de fenômenos em plataformas: Limites e interseções entre a semiótica, a etnografia digital e a ciência de dados<sup>1</sup>**  
**Building the corpus in the study of platform phenomena: Limits and intersections between semiotics, digital ethnography, and data science**

Santiago Videla

**Palavras-chave:** Mídia; Sociosemiótica; Metodologia; Plataformas.

En respuesta a las propias dificultades del trabajo de investigación sobre plataformas mediáticas este trabajo se propone dar cuenta de los desafíos de construir un corpus, que constituya una muestra representativa cuando nos enfrentamos a un universo particularmente cambiante como es el que nos ocupa.

Paralelamente buscamos exponer la complejidad que supone establecer los atributos del objeto de investigación, de modo que puedan ser procesados por sistemas automatizados y que arrojen resultados significativos en el análisis de fenómenos sociales mediatizados.

Para eso recorreremos los avances de las ciencias de datos, las etnografías de redes y los estudios culturales y semióticos a fin de establecer las diferencias y un marco de acuerdos comunes para este tipo de investigaciones.

Buena parte de los muy buenos avances del amplio espectro de las ciencias de datos tienen en común que se problematiza la búsqueda de eficiencia del procesamiento, en el volumen creciente de datos generados por los que entienden como *la vida fuertemente*

---

<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.



---

*digitalizada de la sociedad*. En esa línea, el pasaje del modelo de Big Data al de Fast Data (Sánchez Piccardí y Palomo, 2021) y el análisis de flujos permanentes e ilimitados de datos lleva al camino de soluciones como el apilamiento de software de lectura de bases de datos. Trabajos como la tesis de Fajardo (2023) discuten la efectividad de diferentes alternativas, Stark y Flink. Ambas opciones permiten al analista acceder a volúmenes incalculables de información. Sin embargo, la naturaleza de la captura de los datos y su ponderación y representabilidad no son parte de las problematizaciones.

Los trabajos en el campo de las etnografías de redes parecen estar en el otro extremo. Los datos que se trabajan son de profundo detalle micro (Pink, 2019). La escala es de decenas de observaciones o entrevistas y en algunos casos, cientos de posts o comentarios en redes. Parte de la semiótica y los estudios culturales van en esa línea: estudios de caso único o pocos casos de cosas que pasan en plataformas (AAVV 2023). Sin desconocer la riqueza de estos aportes y su validez metodológica, el volumen y la variedad del contenido producido en plataformas como Tik Tok o Instagram requieren también de una mirada que, mientras no pierda de vista la riqueza de las observaciones de nivel *micro* no deje de atender a su inevitable escalabilidad.

Lejos de pretender ser una solución definitiva, en trabajos anteriores (Autor, 2023) se dio cuenta de la complejidad de construir muestras representativas de, por caso, posts en Redes Sociales. Verón (1985) proponía para el estudio de medios gráficos tomar períodos largos (dos años) en los que la coyuntura afectara de manera pareja la muestra y descansaba en la premisa (permítaseme la simplificación) de que el diario o la revista es *más o menos igual* en el tiempo. En cambio, al músico que postea o al usuario que comenta le cambia su vida entre los 14 y los 18, entre los 19 y los 25, entre los 26 y los 35. Se enamora, se pelea, se entristece, cambia a su equipo de trabajo, fracasa, tiene éxito. Además, por caso Facebook (Meta) cambia cada 2 años el entorno visual y las cosas que se pueden o no hacer en todas sus plataformas. Aun sin negar que existen rasgos que permanecen en el tiempo. Pero la variabilidad individual en ambos polos es



---

un factor que no debe ser menospreciado. No porque deba ser tenido en cuenta a nivel del caso, pero si como ordenador de la metodología de recolección.

En el mismo trabajo habíamos propuesto que lo más cercano a una muestra representativa es primero asumir que solo se tendrá una *foto* de un momento determinado. Para eso se habrá de elegir ya sea un conjunto de hashtags o de posteos o de cuentas (a modo de conglomerados) lo suficientemente amplios y representativos del segmento a estudiar y de la totalidad en la que se inscriben, y seleccionar un corpus de unidades por algún sistema de azar que representen de la manera más fiable posible una instantánea de un momento del sistema de intercambio que se analiza.

En *Una Mecánica metodológica*, José Luis Fernández (2023) propone que el primer paso necesario es el de llevar adelante la descripción del objeto/corpus. Esto es algo tan sencillo, como sofisticado y riesgoso: acercarse al objeto, mirarlo y *anotar que tiene*. La ventaja: aquello de lo que hablará el investigador será efectivamente del objeto y no de un prejuicio de conversación de bar. El riesgo es el de generar categorías ramplonas o pomposas denominaciones innecesarias. Con esa descripción la propuesta es la de desarrollar grillas de análisis que son a su vez, para ese autor, el primer análisis del objeto.

Desde allí, cuando la observación del objeto se realiza aparecen atributos difíciles de capturar. Por caso, tomemos un simple emoji de una carita con una lágrima en la caja de comentarios ¿Da cuenta de que el posteo lo apena al comentador? (*algo como yo vi esto y me puso triste a mi*) ¿Es una muestra de emoción genérica ante el posteo? ¿Es una muestra de llanto moderado por congoja? ¿Es una muestra de empatía y acompaña el sentir del otro? (*Yo vi tu posteo y te acompaño el sentir a vos*) Así surge la necesidad de grillar posteos y comentarios para entender el sistema de intercambio (Autor, 2025) pero así y todo no alcanza.

Con estas herramientas en mente nos proponemos avanzar en la discusión de la necesidad de articular niveles y técnicas de observación que permitan avanzar en la comprensión más acabada de estos fenómenos.



---

## Referências

Fajardo, H. Estudio comparativo entre Apache Spark y Apache Flink en el procesamiento de streaming en entornos Big Data. Trabajo Final presentado para obtener el grado de Especialista en Inteligencia de Datos Orientada a Big Data. UNLP. 2023

Fernández, J. L. Una Mecánica Metodológica. La Crujía. 2023

Sánchez Piccardi, M. L., e Palomo, L. E. . (2021). Del big data al fast data: enfoques modernos de streaming de datos para el procesamiento de datos masivos en tiempo real. *Difusiones*, 21(21), 38–58. Recuperado a partir de <http://ediciones.ucse.edu.ar/ojsucse/index.php/difusiones/article/view/401> 2021

Autor. Referencia oculta. 2023

Autor. Referencia oculta. 2025