Anais de Resumos Expandidos
VII Seminário Internacional de Pesquisas em Midiatização e Processos Sociais
ISSN 2675-4169 Vol. 1, N. 7 (2025)

A construção e analise do corpus no estudo de fenômenos em plataformas: Limites e interseções entre a semiótica, a etnografia digital e a ciência de dados[1]

Building and analyzing the corpus in the study of platform phenomena: Limits and intersections between semiotics, digital ethnography, and data science

Santiago Videla

Palavras-chave: Midiatização; Sociossemiótica; Metodología; Plataformas.

Resumo:

Este trabalho propõe discutir o problema da construção do corpus e da captação dos dados de análise no estudo de fenômenos relacionados às plataformas, em particular aqueles vinculados a aspectos da vida musical. Com esse objetivo, propõe-se percorrer os limites e as interseções entre a semiótica das mediações, a etnografia digital e a ciência de dados.

Abstract:

This paper aims to discuss the problem of corpus construction and data collection for analysis in the study of platform-related phenomena, particularly those connected to aspects of musical life. To this end, it proposes to explore the boundaries and intersections between the semiotics of mediatization, digital ethnography, and data science.

Keywords: Mediatization; Sociosemiotics; Methodology; Platforms.

[1] Trabalho apresentado ao VII Seminário Internacional de Pesquisas em Midiatização e Processos Sociais. POSCOM-UFSM. Santa Maria, RS. ECA-USP. São Paulo/SP.

ISSN 2675-4169 Vol. 1, N. 7 (2025)

Este trabajo se propone dar cuenta de los desafíos de construir un corpus, que funcione como una muestra representativa, cuando nos enfrentamos a un universo como es el de las plataformas mediáticas, que es difícil de definir como tal por su persistente variablidad. En el mismo sentido trataremos de establecer los desafíos que supone establecer los aspectos/propiedades 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. Se trata pues de ajustar un proceso de diálogo alrededor de un conjunto de objetos que a la fecha se presentan con dificultades frente al analista.

Desde nuestra perspectiva eso que pasa en las plataformas es, en principio, una mediatización. Es decir, un tipo de sistema de intercambio que pone en juego cuestiones tecnológicas, restricciones discursivas y prácticas sociales asociadas (Fernández 2023)

Así, por caso, lo que pasa en un comentario de un posteo en Instagram de la cantante argentina Lali Espósito será en parte las restricciones y posibilidades técnicas que Instagram determine (allí se pueden usar GIFs, pero en Youtube no, por ejemplo), en parte los géneros y estilos presentes, y en parte las decisiones de los nativos en esos contextos. En Videla (2023) mostramos como un posteo de colaboración publicitaria de una artista musical rumana (Inna) es respondido tanto con mensajes esperables como frases de afecto o emojis (como fueguitos), como también con otro mensaje hecho de combinación de estos, más títulos de sus canciones y frases de sus letras, en una estructura sofisticada que bien pude pensarse, equivalente a los *caligramas* de Appolinaire. También mostramos allí un posteo de una *selfie* frente al espejo de cantante argentina Lali Espósito que recibía respuestas ligadas a él, y otras como como *"comele la boca a Rosalía"* (un tercero no presente en la foto, pero en referencia a un

ISSN 2675-4169 Vol. 1, N. 7 (2025)

cercano paso de la artista por España, lugar de residencia de la mencionada) o *"Es más fácil sacarle el puesto a Roncaglia que responda lali"* (en referencia a la persistencia en el puesto de un jugador de futbol del Club Atlético Boca Juniors).

Es cierto, que eso que pasa tanto en el posteo como en el comentario condensa, en alguna medida, prácticas previas y hasta pre-mediatizadas del vínculo músico-fan. La declaración de amor siempre existió. Antes de las mediatizaciones ocurría al borde del escenario, y ahora se esboza en una app con un fueguito o un corazón. Lo mismo, respecto de la aparición en el comentario del refranero popular entre amigos (el comentario futbolero), que ahora en el pasaje al espacio público masivo, pareciera construir efectos de complicidad con terceros no necesariamente conocidos por su autor. También sabemos que el desarrollo de algún tipo de práctica artística visual para ofrendar al ídolo ya estaba en los inicios de la vida de los músicos populares (Videla 2009). Por eso, el desafío primero es entender el nivel de novedad que presentan, en función de la serie social que invocan estas prácticas en el sistema de intercambio. Luego acercarse a determinar que peso tienen en el total de las interacciones del músico y de la plataforma. De allí que cuando encontremos comentarios que aparentemente parezcan despegados de la propuesta del posteo, no deban tratarse, en principio, como un caso aislado.

En Videla (2025) insistíamos también que estas apariciones, tal vez *raras* o *infrecuentes* en las redes de otros artistas debían ser parte de las restricciones del sistema de intercambio. Es que por definición son parte de él y deben poder ser recuperados como tales, entendiendo además su incidencia en el conjunto de las interacciones. Así, el primer desafío es entender entonces, el volumen y la frecuencia de estos casos.

La mirada de la ciencia de datos.

La ciencia de datos persigue la búsqueda de modelos "que describan patrones y comportamientos a partir de los datos con el fin de tomar decisiones o hacer predicciones" (García y otros 2018, p. 5) Se trata de un campo de enorme crecimiento,

ISSN 2675-4169 Vol. 1, N. 7 (2025)

como resultado del gigantesco volumen de información que se produce por el hecho de que vivimos en plataformas desde el inicio de este siglo, en la etapa que los textos denominan de Big Data. Su injerencia abarca "a numerosos grupos de investigación de diferentes áreas (computación, estadística, matemáticas, ingeniería, etc.) que trabajan en la propuesta de nuevos algoritmos, técnicas de computación e infraestructuras para la captura, almacenamiento y procesado de datos, etc." (García y otros, p. 6)

Tradicionalmente, en los manuales de metodología, se asigna a los estudios cualitativos el análisis de pocos casos, *pero en profundidad*, y el análisis de muchos casos o universos más grandes a lo cuantitativo (Sabino 1992; Sautu 2005; Marradi y otros 2018; Hernandez Sampieri y otros 2005; Corbetta 2007) Para Sabino los métodos cualitativos "intentan recuperar para el análisis parte de esta complejidad del sujeto y de sus modos de ser y de hacer en el medio que lo rodea. Lo íntimo, lo subjetivo, por definición difícilmente cuantificables, son el terreno donde se mueven por lo tanto los métodos cualitativos" (1992: 65). Sin embargo, esa búsqueda de comprensión subjetiva en los textos, está en tensión con la escala intrínseca del fenómeno que nos ocupa en las plataformas. El comentario aparece entre miles de comentarios y el posteo entre miles de posteos[2]. Necesariamente estamos frente a algo que tensiona lo propio del estilo individual, con el estilo esperable en plataformas (Videla 2023) Por eso nos resulta complementario recurrir al auxilio de la propuesta de distinción entre análisis estadísticos y mecánicos de Levi-Strauss. Para él, "un modelo cuyos elementos

[2] No está de más aclarar que para la mayoría de los manuales de metodología (Sabino 1992; Sautu 2005; Marradi y otros 2018; Hernandez Sampieri y otros 2005; Corbetta 2007) el análisis discursivo es una parte menor de la investigación social y suele asociarse con perspectivas lingüísticas o desarrollos de la semiólogía de mediados del siglo XX. Este trabajo presupone que la sociosemiótica, a partir de los trabajos de Eliseo Verón, Oscar Steimberg, Oscar Traversa, Fausto Neto entre otros (como se recupera en Fernández (2021) es un campo validado y con amplia bibliografía que lo respalda.

ISSN 2675-4169 Vol. 1, N. 7 (2025)

constitutivos se encuentran a la misma escala que los fenómenos será llamado ´modelo mecánico´, y ´modelo estadístico´ aquel cuyos elementos se encuentran en una escala diferente" (1987:325) El objeto que nos ocupa implica una escala que excede la posición del observador y por lo tanto el acercamiento retomará aspectos estadísticos. Desde allí, no hay que perder de vista que la naturaleza de eso que se analizará será un conjunto de palabras, imágenes y sonidos. Sólo la observación de palabras ya ha sido un problema para la tradición del *content análysis,* que desde hace más de 5 décadas ha trabajado el uso de herramientas computacionales para el procesamiento de la información, poniendo el acento en que, si no se presta atención a lo que nosotros llamamos el sistema de intercambio "a menudo, tales enfoques no son más que contar palabras sin considerar sus significados" Krippendorff (1980, p.87).

Para García y Otros el Big data es "el conjunto de datos cuyo tamaño supera considerablemente la capacidad de captura, almacenado, gestión y análisis del software convencional de bases de datos" No obstante, la definición presupone "también a la variedad del contenido y a la velocidad con la que los datos se generan, almacenan y analizan. Estas dimensiones son las ´3V´ con las que la empresa Gartner describió big data, es decir volumen, velocidad y variedad de los datos" (2018, p.14)

En la actualidad, buena parte de los avances en esta área tienen en común que se problematiza la búsqueda de eficiencia del procesamiento de ese volumen creciente de datos. Afirman que son generados por los que entienden como *la vida fuertemente digitalizada de la sociedad*. En esa línea, se plantea el pasaje del modelo de Big Data al de Fast Data (Sánchez Piccardi y Palomo, 2021). Ya no alcanza con procesar mucho, ahora se requiere que sea veloz. 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 de bases como *Apache Stark* y *Aáche Flink*. Ambas opciones permiten al analista acceder a volúmenes incalculables de información.

ISSN 2675-4169 Vol. 1, N. 7 (2025)

A esa información, en esta corriente, la llaman *dato*. Para ellos un dato es la recolección de "un conjunto de hechos (una base de datos, BD) y los patrones son expresiones que describen un subconjunto de los datos (un modelo aplicable a ese subconjunto)" (García y otros, p. 30) Sin embargo, la naturaleza de la captura de los datos, su ponderación y representabilidad no son parte de las problematizaciones. Araoz y Cellone (2025) recorren la discusión sobre la noción de dato en las diversas ciencias sociales enfatizando la idea de que los datos son construcciones del analista y no *algo dado*. En ese sentido, la noción de dato a la que adscribimos está más cerca de la noción de *patrones* de García y otros.

Recordemos que, como adelantáramos anteriormente, nuestro objeto de estudio (el posteo en una red, la interfaz del homebanking, una aplicación mobile, un sitio web) es una mediatización. Cómo tal tiene *propiedades significantes en su superficie discursiva*. Eso es lo que Verón (1987) define como *marca*. Cuando en las ciencias de datos, hablan de *datos*, para nosotros hablan de *marcas*. Es el objetivo del analista convertir esas marcas en *huellas* del proceso productivo. Así, siguiendo a Fernández (2021) *un dato es una marca puesta relación con sus condiciones productivas*, cuando se lo inscribe en el análisis en relación con el proceso productivo de producción de sentido. Así, un dato es *algo puesto en relación con algo*. Esto Fernández también lo llama el pasaje de grillado a matriz.

Esto no supone bajo ningún concepto una crítica a las ciencias de datos y sus resultados. Todo lo contrario, se trata del primer paso de ajuste terminológico que permita establecer puentes de trabajo.

Las etnografías digitales

La definición de mediatización que trabajamos entiende que estamos frente a un sistema de intercambio de sentido que pone en funcionamiento tres series de los social: la de los dispositivos técnicos, la de los discursos y la de las prácticas sociales. Hace más de 50 años, Christian Metz en ponía el acento, en su definición del gran régimen del

ISSN 2675-4169 Vol. 1, N. 7 (2025)

significante cinematográfico, en la incidencia de estos tres niveles (como lo rescata Fernández, 1994). Lo que hace al cine es pues, el cinematógrafo (lo tecnológico), la selección genérico-estilística y una arquitectura particular de las salas y un modo de acercarse a ella. Además, para el caso del cine, la televisión o la radio misma, contamos con el auxilio de más de 60 años de literatura describiendo el fenómeno. Allí las referencias a lo que hacen los nativos son muchas y con relativo acuerdo entre investigadores. En cambio, con las plataformas, en parte por la novedad y en parte por las limitaciones materiales de los análisis, sabemos poco de lo que la gente hace con sus aplicaciones. Se escribe bastante sobre patrones de comportamientos (% de uso del like, % de compartidos, % visualizaciones) pero poco sobre el significado de eso y sus efectos discursivos.

Por caso, buena parte de la literatura sobre *narrativas transmedia* presupone la existencia de un usuario típico de carácter universal basado asunciones aceptadas por el conjunto de autores, que consume ordenadamente relatos y vive una vida digital, sin un aparente contraste empírico que lo demuestre (como puede verse en los trabajos de Irigray 2021, 2022; Alabadejo 2017; Lovato 2018; Pratten 2011) Sin que se entienda como una crítica a esta posición (de la que valoramos su productividad) destacamos que lo propio de estos espacios dedicados al estudio de *lo comunicacional* es la presuposición de un conjunto de acuerdos sobre las prácticas de los nativos que en parte se basa en la propia experiencia de los autores y en parte en repetir cadenas de citas asignándole valor de dato.

Eso nos obliga a revalorizar áreas de trabajo como la de los textos de Pink y otros (2019) que se ubican en un extremo opuesto. Enmarcados en lo que denomina etnografía digital, explican que, se trata de una disciplina que “esboza una aproximación al trabajo etnográfico en el mundo actual. Invita a los investigadores a considerar cómo vivimos e investigamos en un entorno digital, material y sensorial” (2019, p.17) Es decir, que el foco está en la comprensión de que hace *ese* nativo con *esa* plataforma. No se trata por cierto de un campo homogéneo. Aronica nos muestra que “A la etnografía

ISSN 2675-4169 Vol. 1, N. 7 (2025)

que se realiza en Internet se la conoce con diversos términos: ciberetnografía (Escobar, 1994); etnografía del ciberespacio (Hakken,1999); etnografía virtual (Hine, 2000); antropología de los medios (Ardévol y Vayreda, 2002); etnografía mediada o etnografía de/en/a través de los medios (Beaullieu, 2004); etnografía de lo digital (Estalella, 2006); autoetnografía (Espinosa, 2007); netnografía (Turpo, 2008); etnografía de internet (Ardévol, Estalella y Domínguez, 2008) y etnografía digital" (2019, p.30) No obstante encuentra cierto acuerdo en que "Consiste en descripciones detalladas de entornos, sucesos, personas, interacciones y conductas que son observables. Incorpora lo que los participantes comentan, sus experiencias, actitudes, creencias, pensamientos y reflexiones tal como son expresadas por ellos y no como las percibe el investigador" (Aronica 2019, p. 29)

Las interacciones en plataformas suelen ser descritas como parte de un conjunto de conductas. Para Pink y otros

> "la experiencia de todos los días apunta a que dicha idea podría arrojar luz sobre la dinámica social presente en plataformas sociales tan distintas como las listas de correo, Twitter, Weibo, Facebook y WhatsApp. La terminología y la sintaxis varían de una plataforma a otra ("hilo", "hashtag", "tendencias", "comentarios", "chat", etc.), pero todos estos sitios organizan las conversaciones mediante una serie específica de entradas unidas, es decir, mediante hilos." (2019, p 138)

Este enfoque no se pregunta (ni tiene por qué hacerlo) sobre la significación de los *likes* o si es lo mismo comentar en una burbuja de comentarios de Facebook a hacerlo luego de un hilo de X (ex Twitter). En un trabajo previo, Theviot (2014) analiza una serie de conversaciones en Facebook alrededor de la militancia política de una serie de usuarios. El trabajo es hiperdetallado y reconstruye aspectos del sistema de intercambio entre el muro y los comentarios, aunque no se detiene en la descripción de la superficie discursiva. Es decir, los fenómenos de significación suelen ser relegados en el análisis y propuestas como la de *semiodata* de Raimodo Anselmino (2024) cobra un valor especial al intentar su recuperación.

ISSN 2675-4169 Vol. 1, N. 7 (2025)

Siguiendo a Fernández, no es menor prestar atención a lo discursivo. En particular, a eso que denomina *megusteos diferentes*, que son "una especie de efecto lateral, micro dentro de lo micro y que, sin embargo, forma parte de lo más profundo de la comprensión interaccional de Facebook como red de contactos" (2018, p.108). La misma acción de dar me gusta a una selfie en un lugar público ingresa en diferentes sistemas de intercambio, si entre *megusteante* y *megusteado* existe vínculo de jefe-empleado, de amigos o de conquistante-conquistado. También sabemos que, en última instancia un *hilo*, no es lo mismo que un *hashtag* o una *tendencia*. Estos presuponen fenómenos de significación diferentes derivados de los dispositivos en los que se instalan.

También aparece en estos trabajos el problema de la escala de observación. Suele ser de decenas de casos y a veces, cientos de posteos o comentarios en redes. Aún con la riqueza inherente de la capacidad del investigador, no es lo mismo elegir como objeto un artista que postea 3 veces por día y recibe de a 3000 comentarios que un artista que postea una vez cada 15 días y recibe de a 50 comentarios. El valor de los hallazgos para avanzar a conclusiones de nivel *meso* es muy diferente y riesgoso (En la línea del efecto túnel que advierte Paolo Fabri, 1999).

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 (como pueden verse en 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 importancia central de las observaciones de nivel *micro* no deje de atender a su inevitable escalabilidad. Pero insistimos, estos aportes y miradas que venimos describiendo son claves en la construcción del objeto y nos obligan a un diálogo articulador.

La propuesta sociosemiótica

ISSN 2675-4169 Vol. 1, N. 7 (2025)

Como ya dijimos, nuestro objeto de investigación es el sistema de intercambio de sentido mediatizado. Al analizarlo relevaremos marcas de textura del dispositivo técnico, aspectos de formatos discursivos y trazamos hipótesis de prácticas de uso a partir de las posibilidades y restricciones materiales. Siempre teniendo en cuenta la seie histórica que hace posible cada uno de los elementos que nos ocupa.

Respecto del analista de medios tradicionales del siglo pasado estamos en desventaja. Ya no nos enfrentamos a un objeto *más o menos estable* como el de ellos. 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 16 y los 18 años, entre los 19 y los 21, o entre los 26 y los 28. Se enamora, se pelea, se entristece, cambia a su equipo de trabajo, fracasa, tiene éxito… Y eso afecta en parte el modo en el que construye la comunicación. 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 estableciendo nuevos posibles, restricciones y negociaciones de los usuarios a la hora de comunicar.

Sin que esto implique negar que existen rasgos que permanecen en el tiempo, la variabilidad individual en todo el sistema es 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 *Una mecánica metodológica*, José Luis Fernández afirma que

> "No puede seleccionarse un corpus útil y representativo para los objetivos del análisis que se propone sin tener en cuenta el contexto sociocultural en el que se desenvuelven los intercambios y el estado del arte existente sobre el tema. Si ya se conocen esos planos, deberán ser sintetizados, de alguna manera más o menos exhaustiva, en el momento de decidir qué materiales se analizarán efectivamente y cuáles, lo que es muy importante, quedan afuera" (2023, p.154)

ISSN 2675-4169 Vol. 1, N. 7 (2025)

Se trata de una selección que pueda dar cuenta de que la unidad de análisis es el *sistema de intercambio mediatizado*, ya que este enfoque "permite dejar de lado (…) tanto las grandes configuraciones macro de las mediatizaciones y sus relaciones con lo social, como la esperanza parroquial de la especialización" (Fernández 2021, p. 196)

La deuda de esta propuesta, si es que se le pudiera adjudicar al autor, está en la resolución procedimental. Nos se muestra un caso de ejemplo de corpus que ordene al lector. Por eso, antes de avanzar deben quedar claros los márgenes dentro de los que perseguimos el objetivo de investigación:

- La mirada analítica debe ser micro, con esperanzas de formalizaciones de nivel meso (Fernández 2021). Esto significa necesariamente realizar una descripción exhaustiva de casos.
- El objeto tensiona las categorías tradicionales de lo cualitativo y lo cuantitativo. Es decir, la mirada micro no obliga, sin desconocer la escala del fenómeno, a mantener la mirada cuali.

En la etnografía digital, el objeto se describe en relación con lo que los nativos hacen. Nosotros entendemos, que una mirada sociosemiótica parte del sentido producido (Verón 1987). Es la materialidad que debe ser puesta en relación con sus condiciones productivas. Borelli y otros retoman a Verón para afirmar que "El principio de la estructura interna de un corpus parte de la elección en función de cierta homogeneidad" recordando que para él "todo texto 'es un objeto heterogéneo, susceptible de múltiples lecturas, situado en el cruce de una pluralidad de 'causalidades' diferentes" (2024, p. 249)

En Videla (2023) habíamos propuesto que lo más cercano a una muestra representativa es primero asumir una limitación: solo se tendrá el equivalente a una *foto* de un momento determinado. Cualquier generalización que presuponga validez en el tiempo es necesariamente cuestionable.

Para avanzar, se habrá de elegir ya sea un conjunto de hashtags, de posteos o de cuentas (a modo de conglomerados) lo suficientemente amplios y representativos del segmento

ISSN 2675-4169 Vol. 1, N. 7 (2025)

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. Para varios autores, este tipo de selección por conglomerados no garantiza a todos los elementos del universo la misma probabilidad de aparecer (Marradi y otros 2018). Pero el acceso a la totalidad sin una restricción de este tipo es aún peor. Más cuando no existe, al momento de publicarse este artículo, una *api* que capture todo de una red, ni un todo que pueda considerarse representativo más allá del momento de la captura.

Dado que la clave es evitar el sesgo fundacional en cualquier avance científico, nos proponemos dialogar con otros equipos, que llevan adelante esfuerzos en esta línea. Uno de ellos es el que mencionamos de Raimondo Anselmino, que propone una articulación de la semiótica con los análisis computacionales, denominándolo *semiodata*. Allí "se integra el estudio empírico de la puesta en discurso desde un punto de vista sociosemiótico (Verón, 1998) con el empleo de herramientas y métodos computacionales diversos y el análisis univariante y multivariante de datos y metadatos" (Raimondo Anselmino y otros, 2024: 194).

Su esfuerzo se concentra en conseguir mediante procedimientos computacionales, el acceso a la mayor cantidad de unidades de análisis posibles, más allá de las restricciones de las plataformas. El material se recopila en bases de datos, y a "su vez, el módulo para extracción, transformación y carga de datos de Pentaho permitió sistematizar las planillas de cálculo de ambos corpus en una base de datos relacional. Sobre la base de datos MySQL construida con el gestor MySQl WorkBench se efectuó un análisis exploratorio y descriptivo que acompañó a la observación semiótica convencional" (Raimondo Anselmino y otros, 2024: 211). Si bien los primeros resultados que se exhiben pueden parecer, al ojo distraído, iguales a los que ofrece cualquier solución existente desde 2010 tipo Hootsuite (cantidad de clicks, cantidad de visualizaciones, cantidad de posteos con links, entre otros) hay en la propuesta un esfuerzo por avanzar

ISSN 2675-4169 Vol. 1, N. 7 (2025)

en la complejidad estilística de los usuarios y las publicaciones que analizan que debe ser valorado y atendido.

El enfoque se enfrenta con dificultades equivalentes a las que aquí pretendemos superar. La gran calidad del análisis de la información a veces entra en tensión con la masividad de los datos y el origen no semiótico de las api y sus desarrolladores. La información que brindan obligan a leerla por fuera de contexto y está siempre frente al riesgo del *content analysis* que mencionábamos citando a Krippendorff. O, en la misma línea, siguiendo la propuesta sobre los *megusteos* de Fernández, tampoco podríamos saber lo que significa ese número, sea el que fuere, de interacciones. Un punto de partida equivalente tiene el trabajo de Borelli y otros (2024) que para el estudio de fenómenos de circulación de sentido explica que primero analizan "las métricas relacionadas con el árbol máximo de los títulos y, posteriormente, las de los textos, así como aplicamos la clasificación jerárquica descendente únicamente en los textos" (p. 249) lo que les permite luego trabajar con grafos arborecentes de relaciones semánticas. Ellos logran así hacer evidentes ejes de circulación discursiva que iluminan el peso de ciertas configuraciones de sentido, que en modo alguno habrían sido comprendidos o atendidos sin su intervención.

Raimondo y otros (2024) hacen también el esfuerzo de vincular en *engaging* con el uso de emojis en los posteos dando cuenta de estilos de consumo de usuarios. Ellos afirman que "Como sucede con el resto de los recursos de este tipo, los emojis funcionan como huellas de ciertos mecanismos significantes actuantes en las propiedades específicas del discurso de información plataformizado" (Raimondo y otros 2024, 212). Esta afirmación nos lleva a complejas preguntas sobre su funcionamiento en la instancia de reconocimiento. Por caso, tomemos un simple emoji de una carita con una lágrima en la caja de comentarios. Sin entrar en la distinción de uso por segmento etario (para los menores de 25 está del lado de la empatía con el otro, para los mayores está del lado de la expresión del sentimiento propio y la repetición indica mayor sentimiento) ¿Ese emoji da cuenta de que el posteo lo apena al comentador? (algo como *yo vi esto y me*

ISSN 2675-4169 Vol. 1, N. 7 (2025)

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*) Este problema nos lo enfrentamos en Videla (2023) y al presente no tenemos otra propuesta que marcarlo como algo a considerar, e insistir en que uno de los caminos de solución es avanzar en la importancia de los estudios estilísticos entre los usuarios.

En la propuesta de *semiodata* se ve todo el potencial y los desafíos en la generación de descriptores. Por caso, en Videla (2023) en buena parte del trabajo surgió la dificultad de no saber cómo denominar a muchos de los videos de Tik Tok en términos genérico-estilísticos y en términos temáticos. Allí presentamos una clasificación que, aunque resultó relativamente ordenadora de lo que pasaba en Tik Tok, era deficiente por el alto nivel de matices y ablandamientos conceptuales puestos en juego[3]. En etapas posteriores preferimos hablar, en lugar de géneros, de formatos como lo propone Fernández (2023). Esto se debe a que la noción de género invoca a la historia de la literatura de los últimos 2000 años en occidente y la idea de formato tiene metadiscursos menos rígidos, pero más o menos estables, aunque no termine de resolver la variedad y diversidad de lo que aparece en plataformas. Así como nosotros tuvimos dificultad de diferenciar la presencia del humor como formato genérico del humor como matiz estilístico, Raimondo y otros terminan encontrando la categoría *contenido multimedia* para denominar a algo que "comprende publicaciones con contenido visual o audiovisual que suelen versar sobre temáticas consideradas ´blandas´ (espectáculos, entretenimiento, interés humano)" (2024, p. 211) El objeto obliga al investigador a tomar soluciones que anulan las diferencias entre una temática (espectáculos), un tipo discursivo

[3] En ese trabajo mostramos que usamos definiciones duras de a que llamábamos cada cosa para evitar sesgos en el grillado. Pero reconocemos la incomodidad de la blandura de los nombres usados.

ISSN 2675-4169 Vol. 1, N. 7 (2025)

(entretenimiento), una clasificación blanda (interés humano) y un aspecto del dispositivo técnico (contenido audiovisual)

Esfuerzos como los que describimos nos llaman la atención a lo que debe ser el centro de nuestras discusiones: la definición de acuerdos en la búsqueda de aspectos a analizar del objeto y acuerdos en el pasaje de marca a huella. Acuerdos sobre como los convertimos en atributos/variables.

Sobre grillas y matrices

Siguiendo a Fernández, entonces, hay que atender a las propiedades del sistema de intercambio y a las del objeto de estudio. Si este se compone de posteos puedo empezar la investigación sin atender a sus comentarios. Su vida discursiva es, en principio, independiente y está más en relación con la plataforma y con otros posteos, que si se deben atender y relevar. En cambio, si mi pregunta a resolver se orienta a los comentarios es imposible estudiarlos sin tener un análisis del posteo y de la plataforma. Del mismo modo que no puedo dejar de incluir en la muestra posteos equivalente y sus comentarios. Si no se comparara, se corriera el riesgo de afirmar que los juegos visuales con emojis y palabras o los comentarios del acervo popular están en todos los posteos de músicos y fans, algo que es en principio falso. Hay músicos a los que no les comentan así. Hay variaciones entre artistas, entre plataformas, y entre posteos de artistas.

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.

ISSN 2675-4169 Vol. 1, N. 7 (2025)

La matriz de datos es algo que se suele considerar de base en las ciencias sociales. Es una herramienta "formada por un cruce de un haz de vectores paralelos verticales y un haz de vectores paralelos horizontales" (Marradi 2018). Dependiendo de la tradición de investigación los casos estarán en las filas o en las columnas. La matriz es una construcción del analista.

El grillado que propone Fernández es el paso previo. Eso que tiene un posteo debe ser traducido a un conjunto de aspectos que permitan su descripción. Para Lazarsfeld estamos frente a *aspectos* o *dimensiones* cuya nominación (el nombre en la fila o columna) es un proceso literario (1985) y que debe ser descompuesto en indicadores.

"Recién a esta altura es dable comenzar a discutir acerca del tratamiento de datos desde diferentes puntos de vista, y el grillado comienza a convertirse en una matriz de datos" (Fernández 2023, p211). Estamos frente al pasaje de un aspecto, que para nosotros es *marca* a huella del proceso productivo, en el que se lo convierte al nominarlo por su presencia en la grilla. Un paso inevitable del trabajo de análisis que debe hacerse *a mano*. Para eso, y para facilitar la visualización humana del análisis es más favorable que los casos estén en columnas. Las bases de datos en general se organizan con los casos en las filas. Por eso con el solo uso de la función trasponer se puede generar una alternativa para el investigador humano, nuevamente reversible con el uso de la misma función.

Las dificultades de los indicadores.

Si la mirada del analista ingresa por el posteo, por caso un video musical, eso que pasa en la pantalla serán aspectos de textura. El ritmo podrá, por ejemplo, pre-codificarse en la grilla como *agitado, calmo o intenso.*

Esto que para el investigador humano será resuelto en un acuerdo sencillo y con alto grado de consenso, medible en estudios de validación, no es tan sencillo al interactuar con algoritmos sencillos o de *deep learning*. ¿El software que determinará *calmo* debe tomar el BPM? ¿La cantidad de cortes de planos por minuto? Si trabajásemos con un

ISSN 2675-4169 Vol. 1, N. 7 (2025)

transformer que convierta video en texto, tal vez la cantidad de escenas descritas por unidad de tiempo fuera una unidad de medida del ritmo. Eso que es *obvio* para el observador humano, es difícil de traducir y estandarizar.

Más problemático se vuelve aún a la hora de procesar categorías más elaboradas como la de *motivo* (Fernández, 2023). Es que recordemos que estamos frente a procesos de metaforización (Lazarsfeld, 1985) de aspectos del objeto. Para un proyecto con colegas de Brasil, Mar del Plata, Torino y Polonia estamos trabajando el análisis de párrafos producidos por una IA. Tomemos unos de los párrafos que analizamos de ejemplo:

> *"Tras muchas discusiones —y un par de amenazas cruzadas— decidieron dividir el premio y usar una parte para montar una consultora juntos. Contra todo pronóstico, descubrieron que sus diferencias, bien canalizadas, se complementaban. Martina se encargaba de la estrategia y los procesos, mientras que Diego se ocupaba de las relaciones con los clientes. No se hicieron amigos, pero aprendieron a respetarse, y la empresa prosperó más de lo que cualquiera de los dos habría imaginado"*

Por nuestras experiencias personales, rápidamente entendemos que allí está presenta la figura/motivo del emprendedor en la frase *Tomaron un monto de dinero y crearon una consultora que ellos mismo atendían.*

Acordaremos con el lector, que empresario se diferencia de emprendedor por los montos de inversión. Sin embargo, los creadores de aplicaciones que llegan a valuaciones millonarias y parten de aportes de capital familiares de varios ceros son considerados también emprendedores. Otra categoría es la cantidad de empleados. Pocos empleados es un emprendedor. Pero es difícil diferenciar a un emprendedor de una empresa familiar, que bien podría ser el caso de una consultora de dos amigos. Mismo problema tendríamos si insistimos en el nivel de facturación. ¿Cuánto es *"prosperó más de lo que se imaginaron"*? Aun cuando la definición de indicadores sea una tarea propia de la investigación social desde casi sus orígenes, no es menor que

frente a este tipo de objetos (asumimos que ese párrafo de IA propone los mismo desafíos que cualquier comentario humano en plataformas) la dificultad se expande.

La solución, entendemos no es individual. Requiere además de articularse con las ingenierías de datos, generar consensos de investigación. Es además de ineficiente económicamente, muy poco productivo el trabajo aislado de diversos equipos hacia objetivos equivalentes sin acuerdos terminológicos previos. No se trata solo de negociar con un informático o un etnógrafo una fila de una grilla. Tampoco de declamar la importancia de un campo de estudio.

Esfuerzos como los de la red de investigadores en plataformas (REDINPLA) o de la red de investigadores en Mediatizaciones (Midiaticom) van en la dirección correcta y este trabajo se propone en ese recorrido de discusión.

Referências

Albaladejo Ortega. LEGO BRICK LEARNING: Hacia un modelo de alfabetización transmediática a través del storytelling. Tesis doctoral. 2017

Araoz V., y Cellone F. Restricciones y posibilidades de la investigación con datos digitales En: Porto Lopez, P. Intervenciones semióticas: focalizar: transformar: expandir: Actas 11° Congreso Argentino de Semiótica. Libros de Crítica. Área Transdepartamental de Crítica de Artes, 2025.

Aronica, S. La etnografía digital: descripción de un caso de aplicación para el análisis de interacciones virtuales. En: VI Simposio Argentino sobre Tecnología y Sociedad (STS 2019) - JAIIO 48 (Salta). PP 28-39. 2019

Borelli, V., Frigo, D., & Romero, L. M. (2024). Circulação de sentidos em textos noticiosos sobre mortes pela pandemia no Brasil. *MATRIZes*, *18*(1), 239-263. https://doi.org/10.11606/issn.1982-8160.v18i1p239-263

AAVV Mediatizaciones en Pandemia. Editorial Carrera Comunicación UBA. 2023

Corbetta, P. Metodología y técnicas de investigación social. Mg Graw Hill. 2007

Fabbri, P. La caja de los eslabones que faltan. En: El giro semiótico. Gedisa.1999

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

García y otros. Ciencia de datos. Técnicas analíticas y aprendizaje estadístico. Altaria. 2018

Hernandez Sampieri y otros. Metodología de la investigación. MCGraw Hill. 2005

Sánchez Piccardi, M. L., e Palomo, L. E. 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

Krippendorff, K. Validity in content analysis. In E. Mochmann (Ed.), Computerstrategien fÃ¼r die kommunikationsanalyse (pp. 69-112). Frankfurt, Germany: Campus. Retrieved from http://repository.upenn.edu/asc_papers/291 1980

Pratten, H. Getting Started with Transmedia Storytelling. Robert Pratten. 2011

Raimondo Anselmino, N., Cardoso, A., Rostagno, J. y Sambrana, A. El discurso de la prensa argentina en tiempos de algoritmos: una mirada diacrónica sobre la composición de posteos en las fanpages de Clarín y La Nación. En: Âncora, vol. 5, núm. 1, Universidade Federal da Paraíba, Programa de Pós-graduação em Jornalismo, 10-2018, pp. 31-59, en <dc.identifier.urihttp://hdl.handle.net/11336/114457>. 2018

Raimondo Anselmino, Natalia; Cardoso, Ana Laura; Rostagno, José (2024). «El engagement de La Nación y Clarín en Facebook. Medida y significado de las interacciones materializadas del público en el período 2010-2017», Zer, 29(57), 191-220. (https://doi.org/10.1387/zer.26698).Recibido: 28 junio, 2024; aceptado: 03

noviembre, 2024.ISSN 1137-1102 - eISSN 1989-631X / © UPV/EHU Press Esta obra está bajo una Licencia Creative Commons Atribución-NoComercial-SinDerivadas 4.0 Internacional. 2024

Theviot, A Devenir « ami » avec 4500 enquêtés. Les enjeux éthiques de l'analyse d'interfaces semi-privées. Tic&société [Online], Vol. 7, N° 2 | 2ème semestre 2013, Online since 01 June 2014, http://journals.openedition.org/ticetsociete/1608 ; DOI : 10.4000/ ticetsociete.1608. 2014

Videla, S. *TIK TOK en confinamiento: análisis de los efectos de la pandemia de COVID-19 en las formas de producción de contenido.* 2023. En: AAVV Mediatizaciones en Pandemia. Editorial Carrera Comunicación UBA. 2023

Videla, S. *Sistemas de intercambio en plataformas entre músicos y fans*. En: En: Videla, S., Roes Dalmolin, A. (comp) Investigar en plataformas mediáticas desde américa Latina. Um estado de situação e uma ferramenta para o diálogo. pp 39-70. Editorial Universidad Federal de Santa María. 2025 (En prensa)

Este es el último de automatización:
https://library.oapen.org/bitstream/handle/20.500.12657/53680/9781000583335.pdf?sequence=1