

## **Articulación artesanal-computacional para el estudio interdisciplinario de posteos en cuentas de Facebook. Relato de una experiencia asequible**

Natalia Raimondo Anselmino<sup>1,2</sup>, Ana Laura Cardoso<sup>3</sup> & José Rostagno<sup>4</sup>

<sup>1</sup> Consejo Nacional de Investigaciones Científicas y Técnicas (Conicet)

<sup>2</sup> Centro de Investigaciones en Mediatizaciones de la Universidad Nacional de Rosario, Argentina

`nraimondo@conicet.gov.ar`

<sup>3</sup> Universidad Tecnológica Nacional, Facultad Regional Rosario, Departamento de Ingeniería en Sistemas de Información, Argentina

`cardosoalaura@gmail.com`

<sup>4</sup> Universidad Tecnológica Nacional, Facultad Regional Rosario, Departamento de Ingeniería en Sistemas de Información, Argentina

`joserostagno@hotmail.com`

**Abstract.** Esta ponencia relata una experiencia asequible (y, en cierto modo, replicable) de articulación de procedimientos artesanales y automatizados surgida como resultado del inmenso desafío metodológico que supone producir conocimiento en torno a los fenómenos socio-tecnológicos contemporáneos. Se comparte, así, el derrotero y los logros metodológicos de una investigación interdisciplinaria que estudia los posteos que publicaron en Facebook los diarios argentinos *Clarín* y *La Nación* durante el período 2010-2015, a partir de hacer colaborar sinérgicamente las labores propias del análisis sociosemiótico con el análisis de datos cualitativo y cuantitativo provisto gracias a métodos computacionales y un proceso de minería de datos que se encuentra en su último tramo de ejecución. Pesquisa que supuso, incluso, la elaboración de herramientas computacionales que optimizan el acceso a los materiales de análisis. De este modo, la presentación que aquí se expone procura favorecer el efectivo desarrollo de futuras producciones académico-científicas que presenten objetivos, limitaciones y potencialidades similares.

**Keywords:** diseño metodológico, labor artesanal, métodos computacionales, interdisciplina, Facebook

### **1 Introducción**

Abocado a caracterizar el desarrollo del proceso histórico de mediatización, Verón [1] advierte que la novedad específica introducida por Internet consiste en haber producido

una *revolución del acceso*, es decir, del acceso de los actores socio-individuales a los discursos que circulan en la red: “nunca antes el surgimiento de un dispositivo técnico de comunicación había provocado en tan poco tiempo movimientos que atraviesan a la vez los campos económico, tecnológico, político, social y cultural de nuestros viejos Estados-naciones”. Como también se afirma en Verón [2], “los emergentes decisivos son entonces el *alcance* y la *velocidad*, resultados de la *automatización*” [el resaltado es nuestro]. Esto nos ubica ante una de las consecuencias que dicha transformación supone para el campo de investigación sobre fenómenos socio-tecnológicos contemporáneos: “El ámbito del que debemos tener un panorama se ha vuelto inabarcable debido a su digitalización” [3]. De modo que, el grado de sofisticación y complejidad de los fenómenos a estudiar, así como la multiplicación del volumen de información acumulada, requiere, ineludiblemente, de abordajes que articulen saberes en pos de una comprensión más cabal de los objetos de estudio. Todo esto aceptando, claro está, que dichos objetos son producto de una determinada construcción [4] que, en el caso de estudios como el que aquí se problematiza, se hace cargo de los interrogantes, las potencialidades y las limitaciones propias de aquellas pesquisas que se abocan al análisis de grandes cantidades de datos provenientes de plataformas de redes sociales en entornos digitales [5].

Situado en dicha coyuntura, este trabajo no tiene como finalidad exponer resultados<sup>1</sup> sino dar cuenta de las experiencias, dificultades y aprendizajes acaecidos en el marco de un proyecto de investigación que vincula a la Universidad Nacional de Rosario con la Universidad Tecnológica Nacional. El mismo procuró analizar los posteos publicados por los diarios argentinos *Clarín* y *La Nación* en sus respectivas cuentas oficiales en Facebook<sup>2</sup>, durante un período de seis años que va de 2010 a 2015, atendiendo a dos propósitos: por un lado, indagar el tipo de contenido difundido, la frecuencia de publicación y las interacciones obtenidas en ambas *fanpages* y, por otro lado, analizar ciertas regularidades en términos de estrategias discursivas teniendo en cuenta factores o rasgos retóricos, temáticos y enunciativos.

Para ello se propuso un abordaje metodológico que combinó las labores artesanales propias del análisis sociosemiótico —que tradicionalmente solían aplicarse a paquetes discursivos de menor tamaño, por decirlo escuetamente— orientado al estudio empírico de la puesta en discurso<sup>3</sup>, con un análisis de datos cualitativo y cuantitativo proveniente

<sup>1</sup> Sus principales hallazgos fueron en parte presentados en [6, 7].

<sup>2</sup> *Clarín* y *La Nación* son los únicos dos diarios argentinos online de información general y alcance nacional que contaban también con una edición en papel y una cuenta oficial en Facebook durante el período bajo estudio.

<sup>3</sup> Desde esta perspectiva, el análisis discursivo recae en el nivel de la enunciación —no del *dictum* sino del *modus*— identificando aquellas operaciones enunciativas regulares, es decir, modalidades del discurso cuyo funcionamiento es relativamente constante. Nuevamente en palabras de Verón [8]: “El análisis semiológico tiene por objetivo identificar y describir todas las operaciones que (...) determinan la posición de enunciador y, en consecuencia, la del destinatario”. Se considera que el estudio es empírico porque parte de una manifestación material: “Siempre partimos de 'paquetes' de materias sensibles investidas de sentido que son productos” [9], es decir, de fragmentos de la más extensa red de la semiosis social. De modo que, la posibilidad de este tipo de análisis descansa sobre la hipótesis según la cual todo proceso

de las ciencias computacionales. Dicha articulación deviene necesaria en tanto, como observan Lewis, Zamith & Hermida [10], los grandes volúmenes de datos que circulan desafían los métodos tradicionales de análisis propios de las ciencias sociales, mientras que, paralelamente, el análisis algorítmico<sup>4</sup> mantiene una limitada capacidad para entender los significados latentes del lenguaje humano.

Por ello, a la hora de diseñar un andamiaje metodológico fértil para alcanzar los objetivos planteados fue ineludible la constitución de un grupo de trabajo interdisciplinario, entendiendo que la interdisciplinariedad supone diferentes niveles de integración teórico-metodológicos entre disciplinas [11]. En este proceso de conformación de equipo no fue suficiente que la ingeniería en sistemas de información pusiera a disposición sus técnicas y sus algoritmos para el cumplimiento de los objetivos planteados inicialmente, sino que fue necesario un mayor involucramiento en el proceso de definición de los mismos; una tarea en común. Al mismo tiempo, las limitaciones y carencias de sendas disciplinas junto con la complejidad del problema de investigación planteado provocan la necesidad de pensar nuevos horizontes que permitan vislumbrar la posibilidad futura de una labor que se asuma como transdisciplinar<sup>5</sup>.

## 2 Finalidad y objetivos de la investigación

En el ámbito de la semiótica de los medios se evidencia, desde hace tiempo, al interior del sistema de medios masivos de comunicación —recuperando “la hipótesis de que las transformaciones de los diferentes soportes mediales no son autónomas, sino que se derivan fundamentalmente de los cambios dominantes en el sistema entendido como totalidad” [12]—, una serie de modificaciones a las modalidades clásicas de la prensa que atañen, sobre todo, al modo en que se construye el vínculo diario-lector [13,14]. Durante el último decenio del siglo XX y comienzos del XXI se pudo observar, en la prensa, la emergencia de un conjunto de estrategias discursivas del contacto [15], en tanto modo peculiar de configuración del vínculo enunciativo fuertemente anclado en la tecnología de la transmisión en directo de imagen-sonido que impuso la televisión. Transformación que, desde el punto de vista de los efectos sobre el lazo social, puso en cuestión “el clásico ideal utópico de democracia crítica y argumentativa” [12], de la opinión pública moderna sujeta a razón de la cual la prensa se construía como garante, para dar paso a “un vínculo preeminentemente de contacto (...) de carácter pasional y somático, de presente absoluto, inmediatez y celeridad”. Con el paso del tiempo, a las modificaciones producidas por el dispositivo del directo televisivo [32] le siguieron

---

productivo deja huellas en los productos; proceso que es pasible de reconstrucción gracias al reconocimiento y estudio de dichas huellas.

<sup>4</sup> En van Dijk [16] se define al algoritmo como “una lista finita de instrucciones definidas para calcular una función, una directiva paso a paso que permite un procesamiento o razonamiento automático que ordena a la máquina producir cierto output a partir de determinados *inputs*”.

<sup>5</sup> Según Rosenfield [17]: “los proyectos transdisciplinarios son aquellos en los cuales investigadores de diferentes campos no solo trabajan juntos en un problema en común en una considerable cantidad de tiempo, sino que también crean un modelo compartido conceptual del problema que integra y trasciende cada una de sus perspectivas disciplinares separadas”.

aquellas asentadas en las nuevas lógicas propias del ambiente-Internet [14]. Desde que la prensa diaria arribó al ciberespacio, los diarios han multiplicado sus esfuerzos por construir un contrato de lectura que les permita sostener en el tiempo el vínculo con sus lectores y escapar así de los pesimistas augurios sobre su futuro [18], viéndose ante la necesidad de seguir los desplazamientos de un público cada vez más inasible. Por ello, a medida que los usuarios van adquiriendo nuevas habilidades o practicando otros espacios —la blogósfera, Facebook, Twitter, YouTube, Instagram, entre otros— los diarios digitales han ido incorporando dichas modificaciones de alguna u otra manera.

Es precisamente en esta encrucijada, que la investigación que aquí se relata se propuso estudiar el modo en que entran en contacto el mundo de la prensa tradicional argentina (en sus versiones *online*) y el novedoso universo de las redes sociales en Internet, específicamente Facebook, para poder conocer cómo enuncian los diarios en esta plataforma y qué tipo de vínculo le proponen a sus usuarios-lectores. En función de ese horizonte y apoyados en la perspectiva sociosemiótica es que, como se señaló en la introducción, se identificó como objetivo central el análisis de los posteos publicados por los diarios argentinos *Clarín* y *La Nación* en sus respectivas cuentas oficiales en Facebook, durante el período 2010-2015. En tren de arribar a la meta propuesta nos interrogamos acerca de:

- el tipo de contenido que los diarios publican en sus *fanpages*, teniendo en cuenta diversos aspectos como: tipo de posteo (status, link, video, foto), localización geográfica de la información publicada (local, nacional o internacional), género periodístico (noticia, crónica, opinión, entrevista, etc.), temática de referencia (política, economía, deportes, espectáculos, etc.) y temporalidad de los acontecimientos presentados (pasado, presente, instante o futuro).
- la frecuencia de publicación de ambas *fanpages*, reconociendo, a su vez, los comportamientos de publicación en relación con las diferentes franjas horarias (madrugada, mañana, tarde y noche);
- los modos de composición de los posteos, es decir, se identificaron los componentes elementales (texto del post, enlace/s a sitio web, imagen o imágenes y, video) y las diferentes relaciones que, a lo largo del período estudiado, se establecieron entre los mismos;
- las modalidades discursivas prevalentes, a partir del empleo de la clasificación de los cuatro tipos de modalidades inter-sujetos que recuperan Fisher y Verón [19] del lingüista Antoine Culioli;
- las interacciones obtenidas por las publicaciones de las cuentas (en términos de *likes*, compartir y comentar) y su relación con todos los factores antes listados.

Esta investigación se encuentra ya en su última etapa de ejecución y se han producido significativos resultados, algunos de los cuales han sido compartidos en [6,7], mientras que otros están siendo sistematizados para ser publicados en breve. Dichos hallazgos no serán aquí desarrollados dado que, como se ha advertido, no son ellos sino la articulación interdisciplinaria que los hizo posibles el eje de la presente exposición.

### 3 Disquisiciones sobre la configuración del corpus de análisis

Un aspecto crucial en todo análisis de los discursos consiste en discernir cómo se conformará el *corpus* sobre el cual trabajar, es decir, en palabras de Barthes [20], esa "colección finita de materiales, determinada previamente por el analista, con cierta (inevitable) arbitrariedad" cuya construcción, como advierten Gindin & Busso [5] "se erige como resultado y condición de una serie de interrogantes que guían nuestros trabajos, como ese conjunto significativo sobre el que ponemos a trabajar nuestras hipótesis y marcos teóricos". Es así que, en primer lugar, se construyó un corpus —que denominamos, *corpus de base*— teniendo en cuenta los siguientes criterios: se relevó, por cada diario, todos los posteos publicados durante una semana consecutiva por año<sup>6</sup>, eligiendo distintos meses en forma alternada (esto es, un mes sí y un mes no) y variando la semana escogida (por ejemplo, la primera semana completa de diciembre de 2011, la segunda de febrero de 2012 y así sucesivamente), para propiciar cierta cuota de aleatoriedad. De este modo, empleando la herramienta digital *Netvizz*<sup>7</sup>, se constituyó una colección que comprende un total de 1129 posteos —534 de *Clarín* y 595 de *La Nación*.

Para sistematizar toda la información necesaria y clasificar los distintos atributos de interés, se enriqueció la planilla de datos obtenida automáticamente mediante *Netvizz* con la inclusión de seis variables cualitativas *ad hoc* [6] que se completaron manualmente, luego de la observación directa y pormenorizada de cada uno de los posteos y el cotejo de los mismos con las notas publicadas en las versiones *online* de los diarios a las que dichos post reenvían, en caso de que hubiera enlace.

Ahora bien, teniendo en cuenta que, como también lo señalan Gindin & Busso [5], todo corpus se construye luego del acceso a un conjunto mayor del que los materiales recopilados son parte, es evidente que en tiempos de redes sociales en Internet (en adelante, RSI) esa correlación entre la parte y el todo supone una relación de proporción, una escala, diferente. Por ejemplo, si a la hora de validar como satisfactoria una determinada *colección de materiales* uno de los principios a emplear es el de la *saturación*<sup>8</sup>, es evidente que dicho procedimiento adquiere otras dimensiones cuando no trabajamos, pongamos por caso, sobre periódicos impresos compuestos a lo sumo por una limitada cantidad de páginas sino con universos de análisis digitalizados y en línea que comprenden decenas de miles de posteos. En otras palabras, a la hora de trabajar sobre universos discursivos voluminosos, como el que aquí nos atañe, la tarea de elaboración del corpus se complejiza.

<sup>6</sup> Dicho criterio tuvo como finalidad poder observar la producción *rutinizada* de la información que los diarios comparten en las redes sociales en Internet.

<sup>7</sup> Se trata de una aplicación que permite extraer datos de diferentes secciones de la plataforma de Facebook, a través de su API (Application Programming Interface), disponible en: <https://apps.facebook.com/netvizz/>, accedida el 16/04/2018.

<sup>8</sup> Dice Barthes [20]: "el corpus tiene que ser suficientemente amplio como para que se pueda suponer razonablemente que sus elementos saturan un sistema completo de semejanzas y de diferencias; es seguro que si se entresaca un conjunto de materiales se llega, al cabo de un cierto tiempo, a encontrar nuevamente hechos y relaciones ya aislados anteriormente (...); estas 'vueltas atrás' se hacen cada vez más frecuentes, hasta que se llega a un punto en que no se descubre ya ningún material nuevo: el corpus está entonces saturado".

Es por ello que, en segundo lugar y para ampliar el alcance de los resultados obtenidos mediante el análisis del *corpus de base*, se decidió recuperar —nuevamente a través de *Netvizz*— un segundo conjunto de posteos que denominaremos *corpus total*. El mismo, comprende el total de los 54.742 posteos —29.341 de *Clarín* y 25.401 de *La Nación*— realizados por ambos diarios entre 2010 y 2015. Esto permitió comparar, gracias al empleo de métodos computacionales, parte de los hallazgos realizados a partir del análisis del *corpus de base*<sup>9</sup>, así como evaluar el grado de generalidad de ciertas operaciones allí identificadas con aquello que se deriva del estudio del universo completo recopilado en el *corpus total*. Entre uno y otro corpus, podría decirse, hay un cambio de escala que evita, como propone Manovich [21], tener que escoger entre *tamaño* y *profundidad*.

Es así que, gracias a la conformación interdisciplinaria del equipo, sobre ambos conjuntos significantes se ha llevado a cabo un análisis exploratorio y descriptivo<sup>10</sup> sobre una base de datos *MySQL*. En dicha base y con la ayuda de un módulo para extracción, transformación y carga de datos de *Pentaho*<sup>11</sup>, se sistematizaron las planillas de cálculo de ambos corpus. El empleo de herramientas de software libre como *MySQL* y *Pentaho* ofrece una manera más sencilla de administrar grandes volúmenes de datos. Una vez terminado ese proceso, se realizó un análisis univariante y multivariante, con el fin de reconocer ciertas tendencias en el comportamiento de los datos. Para esto último, se recurrió a dos herramientas de minería de datos<sup>12</sup> (también de software libre), *Rapid-Miner*<sup>13</sup> y *R*<sup>14</sup>, en tanto cada uno de los corpus se encuentra estructurado como una *vista minable*<sup>15</sup> y, como sostiene Bunz [3] “al juntarse cantidades gigantescas de datos surge la demanda de algoritmos que preserven la visión de conjunto”.

De este modo, y en concordancia con el proceso emergente que Berry [22] denomina como *giro computacional* o que Lazer et al. [23] distinguen como *ciencias sociales computacionales*, esta investigación supone el ensayo de una manera complementaria de abordar el estudio de la configuración discursiva de las publicaciones que los diarios realizan en sus *fanpages*, articulada con un proceso de minería de datos que se encuentra aún en su última etapa de ejecución. Como se ha dicho antes, se trata de una estra-

<sup>9</sup> En términos de proporción, el *corpus de base* es el 2,06% del *corpus total*.

<sup>10</sup> El análisis descriptivo proporciona información sobre las relaciones entre los datos y sus características.

<sup>11</sup> Software libre de inteligencia de negocios y big data, disponible en: <https://community.hds.com/docs/DOC-1009855-data-integration-kettle>, accedido el 16/04/2018.

<sup>12</sup> Tal como se explica en Arcila, Barbosa & Cabezuelo [24] la mencionada minería de datos “implica la extracción de conocimiento a partir de datos masivos y las relaciones subyacentes que pueden existir entre ellos”.

<sup>13</sup> Software libre que permite la aplicación de algoritmos de minería de datos de forma visual e intuitiva, disponible en: <https://rapidminer.com/>, visitado el 16/04/2018.

<sup>14</sup> Proyecto de software libre compuesto por un entorno y un lenguaje de programación enfocado en el análisis estadístico y gráficos, disponible en: <https://www.r-project.org/>, accedido el 16/04/2018.

<sup>15</sup> “Una Vista Minable es la consolidación en una única tabla de todas las observaciones y los atributos sobre los que se aplicarán los algoritmos de minería de datos” [25].

tegia metodológica cuyo diseño combina las labores artesanales provenientes del análisis sociosemiótico con el empleo de herramientas digitales y métodos computacionales que permiten la recopilación, el pre-procesamiento, el procesamiento y la visualización de cantidades masivas de datos y de metadatos que son, al menos inicialmente, no estructurados, y que pueden ser tanto textuales como no textuales.

#### **4 Pasos y traspies de la labor interdisciplinar**

Más allá de las dificultades ya observadas acerca de la conformación de los corpus, vale mencionar aquí también aquellas surgidas a raíz de la colaboración interdisciplinar, algunas de las cuales fueron más sencillas de reconocer y resolver y otras no tanto, aunque el resultado ha sido en general sumamente provechoso. Por eso este trabajo se propone exponerlas, en todo el sentido de esa palabra, en función de poder analizarlas y facilitar futuras implementaciones de este tipo.

El primer escollo no trivial a sortear fue la construcción de un vocabulario que funcionara como meta-lenguaje en común entre ambas disciplinas. Esto aparentó ser un problema fútil en un principio, pero resultó ser fundamental, ya que es la base sobre la que se cimenta toda la investigación realizada hasta el momento y subyace en las tareas que se encuentran aún en proceso. Este nuevo lenguaje híbrido, que está en constante expansión, permite exponer ideas, inquietudes e interrogantes entre los distintos miembros del equipo, haciendo asequibles aspectos del fenómeno a estudiar que antes quedaban invisibilizados.

Como segunda dificultad puede mencionarse la necesidad de reflexionar sobre los niveles de observación, en función de ensamblar productivamente el análisis detallado —generalmente de nivel *micro* [26]— por medio del cual inferir procesos a partir de reconocer operaciones, proveniente de la mirada sociosemiótica, con la búsqueda de patrones e información en el conjunto de los datos<sup>16</sup> y el enfoque orientado a la obtención de modelos matemáticos propios de las ciencias de la computación, en particular de la minería de datos. Todo esto sin perder de vista que el resultado obtenido por medio de algoritmos requiere *a posteriori*, también y necesariamente, de un proceso inferencial en cierto modo interpretativo, dado que, como sentencia Magnani [27]: “los algoritmos no buscan hipótesis: nadan en la superficie de la empiria”.

#### **5 Pasos y traspies de cara a la automatización de procedimientos**

Tal como hicimos con las dificultades que surgen a raíz de la conformación del corpus y como consecuencia de la colaboración interdisciplinar, ahora llegó el momento de

---

<sup>16</sup> “La tecnología de la codificación emplea como recursos básicos los *datos* y los *metadatos*. Por ‘dato’ se entiende cualquier tipo de información en un formato apto para su utilización por parte de una computadora (...) Los metadatos, por su parte, contienen información estructurada que describe, explica y localiza recursos de información o bien simplifica la tarea de recuperarlos, emplearlos o gestionarlos” [16].

aquellas que aparecen en el camino por automatizar al menos una parte de los procedimientos de análisis de las RSI gracias a la mediación de tecnologías que no requieren la intervención constante de operadores humanos<sup>17</sup>. En este sentido, se alude aquí a las limitaciones relativas a la disponibilidad de los datos y a aquellas derivadas de las modificaciones acaecidas en las plataformas, así como a la consecuente elaboración de tres herramientas computacionales en lenguaje Python y la aplicación de algoritmos de clustering que han permitido superar dichas restricciones.

Por un lado, debido a que las RSI son un entorno muy rico en datos y metadatos se tiende a pensar en los posibles usos de los mismos sin recalcar en la verdadera disponibilidad que suponen para terceros. Tal cual se plantea en Gindin y Busso [5], “el primer presupuesto a desmontarse es aquel que sostiene que los datos presentes en Internet, debido al supuesto libre y fácil acceso, se encuentran disponibles para todo aquel que se interese en su compilación y abordaje”.

Usualmente los datos pasan a través de varios filtros antes de poder ser recuperados por las APIs públicas disponibles y son muchos los que están restringidos por las configuraciones de privacidad de los perfiles o están limitados por las plataformas mismas que suelen monetizar los datos bajo el modelo de suscripciones “premium”. Dicho modelo es conocido como *software como servicio* y consiste en ofrecer una suscripción que permite acceder a datos más detallados o a una expansión de límites de la cantidad de peticiones o de los rangos históricos entre otras ventajas. Un tercer filtro es consecuencia de la preselección de datos históricos que realiza cada RSI y que pone a disposición a través de su respectiva API. Si bien en la mayoría de los casos se pueden recuperar datos históricos no siempre están completos en la medida que el investigador lo requiere. Por ejemplo, la plataforma de Facebook permite buscar posteos históricos en las cuentas de los periódicos bajo estudio, pero no brinda información sobre la cantidad de seguidores que cada *fanpage* tiene en una fecha determinada ni de la evolución del *engagement*, lo cual, en nuestro caso, hizo imposible evaluar la correlación entre los niveles de reacciones de un posteo y la cantidad de seguidores del diario en cada momento de los seis años analizados. En la investigación que aquí se narra estas dificultades conllevaron el replanteo de algunos objetivos que no eran factibles con la información disponible *online*.

Por otro lado, están los escollos padecidos en ocasión de la construcción y utilización de las herramientas con las que se recolecta la información deseada. Como es sabido, la web está en continua evolución y esto provoca cambios que impactan en los proyectos cuando los mismos se extienden en el tiempo o cuando se quiere replicar la metodología utilizada para nuevos casos de estudio —sobre esto se ha reflexionado ya en Raimondo Anselmino [29].

En nuestra búsqueda por analizar las estrategias enunciativas, debimos contemplar la forma en que evolucionó la composición de los posteos. Durante el período de los seis años comprendido en el estudio de las *fanpages* de *Clarín* y *La Nación* se ha podido

---

<sup>17</sup> Según DiFrank (en Maya Palacios, González Hernández & Ocampo Casados [28]) la automatización se logra mediante operaciones “realizadas por un aparato, proceso o sistema que están controladas por elementos mecánicos o electrónicos que actúan como los órganos del ser humano”.

observar, por ejemplo, la incorporación de los *hashtags*, de los emoticones y *emojis*, de las encuestas, *gifs* animados, juegos e, incluso, la multiplicación de las reacciones por medio de las cuales los usuarios interactúan con el contenido de los posteos. La dificultad aquí viene dada por las herramientas de captura de datos que no siguen el mismo ritmo de actualización que las RSI. Fue por esto que, para subsanar las carencias encontradas, se desarrolló una herramienta en lenguaje Python que denominamos *buscar-TitulosFacebook*<sup>18</sup>. La misma permite, dado un archivo .CSV compuesto por un conjunto de enlaces a posteos de Facebook, acceder automáticamente a estos para extraer los *hashtags*, menciones y los títulos de enlace que los componen. Se trata, por tanto, de una herramienta que simula el comportamiento de un operador humano dentro de la red social y, así, sorteas las restricciones y/o limitaciones de *Netvizz* y de la API de Facebook, para luego volcar la información recolectada en un archivo de texto.

En una segunda instancia, y en función de acompañar el análisis (artesanal) de las estrategias discursivas para comprender cómo enuncian los diarios en Facebook, se procedió a comparar el contenido de los posteos con el publicado en los portales de los medios en cuestión. Para acceder a las publicaciones realizadas por los diarios en sus respectivos sitios web, se recurrió a procesos de ingeniería inversa<sup>19</sup> para poder, así, construir el programa denominado *buscarEnPortalesDiarios*<sup>20</sup> en lenguaje Python que dado un archivo .CSV compuesto por un conjunto de enlaces que redireccionan a las publicaciones de cada diario en su respectivo portal extraiga de ellas los datos de interés, tales como: fecha y hora de publicación, título, volanta, bajada, copete, sección del diario donde se publicó y cuerpo. Entre las ventajas de la herramienta desarrollada puede mencionarse que además de ofrecer una forma relativamente sencilla de recolectar información de manera automática, la misma posibilita realizar un gran número de peticiones de información a los portales saltando las limitaciones con las que nos topáramos si ingresáramos artesanalmente. Estas restricciones de acceso fueron incorporadas en abril de 2017 en el caso de *Clarín* y en agosto de ese mismo año en el caso de *La Nación*, y consisten en un máximo de 10 notas libres por mes para lectores no suscriptos.

Esta segunda herramienta permitió detectar la existencia de *links* rotos en la vista minable inicialmente construida con *Netvizz* (pertenecientes, en su mayoría, a *La Nación*); *links*, usualmente elaborados a partir de softwares acortadores de enlaces, que estaban rotos porque habían sido discontinuados por el medio y cuyo contenido sólo podía ser hallado, entonces, manualmente (por ejemplo, mediante una búsqueda en Google). En función de resolver este inconveniente, el equipo que lleva a cabo esta

---

<sup>18</sup> Disponible como software libre en: <https://github.com/Departamento-Sistemas-UTNFRRO/buscarTitulosFacebook>.

<sup>19</sup> En Juárez Ramírez, Licea y Salas [30], se considera ingeniería inversa al “proceso de descubrimiento de los principios tecnológicos de un dispositivo, un objeto o un sistema, mediante el análisis de su estructura, funcionamiento u operación. Este proceso consiste en tomar una entidad por separado y analizar a detalle su funcionamiento, usualmente con el fin de construir un dispositivo o programa nuevo”.

<sup>20</sup> Disponible como software libre en: <https://github.com/Departamento-Sistemas-UTNFRRO/buscarEnPortalesDiarios>.

investigación desarrolló una tercera herramienta denominada *buscarLinksRotos*<sup>21</sup> que permite automatizar la búsqueda del contenido faltante a través de Google. La herramienta realiza dicha búsqueda utilizando otros datos provenientes de la vista minable (tales como: títulos de link al portal, dominio del link y fecha de publicación del posteo) y seleccionando en Google el de mayor similitud, entendiendo por similitud la medida de correspondencia o parecido entre objetos.

Finalmente, también se está trabajando en la aplicación de *algoritmos de clustering*<sup>22</sup> sobre el conjunto de aquellos posteos que fueron clasificados como “Otros” en la variable “género periodístico” del *corpus de base*, una de las variables incorporadas *ad hoc* y categorizadas de modo artesanal para enriquecer la planilla obtenida mediante *Netvizz*. Al clasificar los posteos del *corpus de base* según la variable “género periodístico”, se detectó una sorprendente presencia de “otros”<sup>23</sup>, es decir, de posteos que no pueden ser considerados ni como noticia, crónica, opinión, anuncio o posteo de saludo a usuarios. Se espera, por lo tanto, que el agrupamiento producido a partir de un algoritmo de clustering pueda ayudar a agrupar dichos posteos clasificados como “Otros” en clases de objetos similares a partir de la identificación de ciertos patrones encontrados en las características intrínsecas de las variables de entrada. De modo tal que, una vez identificados los grupos, sea posible inferir (interpretativamente) qué hay dentro de esos grupos y, así, poder caracterizarlos.

## 6 Epílogo

En los apartados precedentes se ha relatado una experiencia que ha hecho asequible el estudio de posteos de Facebook gracias a la articulación (surgida al amparo de una ardua labor interdisciplinaria) de procedimientos artesanales y automatizados.

Fruto de ese trabajo son las inferencias y reflexiones publicadas en [6,7], a partir de las cuales se caracteriza el contenido difundido, la frecuencia de publicación y las interacciones obtenidas por las *fanpages* de *Clarín* y *La Nación* durante sus primeros seis años de gestión en la plataforma de Facebook. Gracias a la mirada diacrónica y al cotejo de los dos *corpus* recopilados pudo observarse el desarrollo del comportamiento de las cuentas y reconocerse que la configuración discursiva de los posteos ha ido variando significativamente a lo largo del tiempo. Esto último no sólo en términos de cantidad de publicaciones sino, también, en relación a las franjas horarias en las cuales se realizan, las temáticas que abordan, la temporalidad de los acontecimientos relatados, los géneros periodísticos de las notas compartidas y, asimismo, los elementos y modos de composición de los *posts*.

<sup>21</sup> Disponible como software libre en: <https://github.com/Departamento-Sistemas-UTNFRRO/buscarLinksRotos>.

<sup>22</sup> Se denomina clustering al “proceso de agrupar un conjunto de objetos físicos o abstractos en clases de objetos similares” (Kaufmann, Han y Kamber, en [31]), a partir de la identificación de ciertos patrones.

<sup>23</sup> Que, en proporción, es el 15,5% de los 1129 posteos del *corpus de base* y ocupa el segundo y tercer lugar en el *ranking* por géneros en las *fanpages* de *La Nación* y *Clarín*, respectivamente.

Por otra parte, al momento de presentación de esta colaboración enviada al Simposio Argentino sobre Tecnología y Sociedad, se está concluyendo el último tramo de la investigación, procurando sistematizar algunos hallazgos que quedaron sin exponerse — como, por ejemplo, una tipología de posts según su composición; una clasificación de los tipos de enlaces, tanto internos como externos, compartidos; el estudio sobre el empleo de etiquetas y hashtags en posts o; la identificación de las modalidades discursivas prevalentes— así como el resultado de la aplicación de algoritmos de clustering mencionada en el apartado anterior.

Se espera, como corolario, que esta narración favorezca el desarrollo efectivo de futuras investigaciones que se enfrenten, como este estudio, a la necesidad de construir conocimiento sobre las sociedades actuales. Todo esto sin desconocer, claro está, las continuas transformaciones que tensionan la producción de conocimiento científico sobre los fenómenos sociales, en épocas en que la tecnología interviene de un modo cada vez más decisivo.

## 6. Referencias

1. Verón E.: *La semiología social*, 2. Ideas, momentos, interpretantes. Paidós, Buenos Aires (2013) 277
2. Verón E.: Prólogo. En Carlón, M. y Fausto Neto, A. *Las políticas de los internautas. Nuevas formas de participación*. La Crujía, Buenos Aires (2012) 12
3. Bunz, M.: *La revolución silenciosa. Cómo los algoritmos transforman el conocimiento, el trabajo, la opinión pública y la política sin hacer mucho ruido*. Cruce, Buenos Aires (2017)
4. Bourdieu, P., Chamboredon, J-C. & Passeron, J-C.: *La construcción del objeto*. En: *El oficio del sociólogo*. Siglo XXI, Buenos Aires (1987)
5. Gindin, I. & Busso, M.: *Investigaciones en comunicación en tiempos de big data: sobre metodologías y temporalidades en el abordaje de redes sociales* En: *adComunica. Revista Científica de Estrategias, Tendencias e Innovación en Comunicación*, n°15. DOI: <http://dx.doi.org/10.6035/2174-0992.2018.15.3> (2018)
6. Rainmondo Anselmino, N., Sambrana, A. & Cardoso, A. L.: *Medios tradicionales y redes sociales en internet: un análisis de los posts compartidos por los diarios argentinos Clarín y La Nación en Facebook (2010-2015)*. *Revista Astrolabio Nueva Época*, número 19 (2017)
7. Raimondo Anselmino, N.: *Prensa online y redes sociales en Internet: notas sobre la circulación de los discursos mediáticos contemporáneos*, conferencia en Pentálogo VIII de Ciseco (Centro Internacional de Semiótica e de Comunicação) “A circulação discursiva e transformação da sociedade”, Japaratinga, Brasil, 25 al 29 de septiembre (2017)
8. Verón, E.: *Fragmentos de un tejido*. Gedisa, Barcelona (2004) 179
9. Verón, E.: *La semiología social. Fragmentos de una teoría de la discursividad*. Gedisa, Buenos Aires (1987) 126
10. Lewis, S. C., Zamith, R., & Hermida, A.: *Content Analysis in an Era of Big Data: A Hybrid Approach to Computational and Manual Methods*. *Journal of Broadcasting & Electronic Media*, 57(1), 34–52 (2013)
11. Vázquez, H.: *La investigación sociocultural: crítica de la razón teórica y de la razón instrumental*. Biblos, Buenos Aires (1994)

12. Valdetaro, S.: Algunas consideraciones acerca de las estrategias del contacto: del papel a la intermediación de las interfaces. Revista LIS, Letra, Imagen, Sonido. Ciudad Mediatizada, N° 1, Buenos Aires (2008) 40-42
13. Valdetaro, S.: Prensa y temporalidad. La Trama de la Comunicación. Vol. 8. UNR Editora. Rosario (2005)
14. Raimondo Anselmino, N.: La prensa online y su público. Un estudio de los espacios de intervención y participación del lector en Clarín y La Nación. Teseo (Colección UAI-Investigación). Buenos Aires (2012)
15. Biselli, R. y Valdetaro, S.: Las estrategias discursivas del contacto en la prensa escrita. Revista La Trama de la Comunicación. Vol. 9. UNR Editora. Rosario (2004)
16. Van Dijk, J.: La cultura de la conectividad. Una historia crítica de las redes sociales. Siglo XXI. Buenos Aires (2016) 56-57
17. Rosenfield, P.: The potential of transdisciplinary research for sustaining and extending linkages between the health and social sciences. Soc Sci Med, 35, (1992) 1343-1357
18. Raimondo Anselmino, N.: Crisis, cambio y transformación de la prensa diaria. Algunos datos para pensar el futuro de un medio. Revista Razón y Palabra. N° 80 México (2012b)
19. Fisher, S. y Verón, E.: Théorie de l'énonciation et discours sociaux. Etudes de lettres.. (Teoría de la enunciación y discursos sociales. Traducción al español de Sergio Mollinedo para la cátedra Teorías y Medios de Comunicación, Universidad de Buenos Aires.) (1986). 211, 71-92
20. Barthes, R.: *La aventura semiológica*. Paidós, Barcelona (1993) 80-81
21. Manovich, L.: Trending: The promises and the challenges of big social data. Gold, M. K. (ed.), *Debates in the Digital Humanities*. Minneapolis: University of Minnesota Press(2012)
22. Berry, D.: The computational turn: thinking about the digital humanities. Culture Machine(2011) 12
23. Lazer, D. et al.: Computational Social Science. Science, 323, issue 5915. (2009). 721-723
24. Arcila, C., Barbosa, E. & Cabezuelo, F.: Técnicas big data: análisis de textos a gran escala para la investigación científica y periodística. El profesional de la información, v. 25, n. 4. (2016) 627
25. Quinteros, O., Funes, A. & Ahumada, H.: Construcción de una Vista Minable para aplicar Minería de Datos Secuenciales Temporales. XXII Congreso Argentino de Ciencias de la Computación.(2016) 672
26. Fernández, J.L.: Las mediatizaciones y su materialidad: revisiones. Busso, M. y Camusso, M. Mediatizaciones en tensión: el atravesamiento de lo público. UNR Editora, Rosario (2017)
27. Magnani, E.: Hacete amigo del big data. En Revista Anfibia. Buenos Aires (2017)
28. Maya Palacios, E., González Hernández, J. & Ocampo Casados, J.L.: Aprendizaje basado en problemas para la enseñanza de los PLC en la Universidad Tecnológica de Altamira, En Revista Iberoamericana para la Investigación y el Desarrollo Educativo, Vol. 8, Núm. 15 (2017) 8
29. Raimondo Anselmino, N.: Un análisis sociosemiótico de la prensa online: investigar el presente en transición. Raimondo Anselmino, N. & Reviglio, M.C Territorios de comunicación. Recorridos de investigación para abordar un campo heterogéneo. Ciespal, Quito (2013)
30. Juárez-Ramírez, R., Licea, G. & Cristóbal-Salas, A.: Ingeniería Inversa y Reingeniería Aplicadas a Proyectos de Software Desarrollados por Alumnos de Nivel Licenciatura. Revista sistemas, cibernética e informática volumen 4 - número 2 - (2007) 49
31. Farias, R., Durán, E. y Figueroa, S.: Las Técnicas de Clustering en la Personalización de Sistemas de e-Learning. XIV Congreso Argentino de Ciencias de la Computación. 2008.

32. Carlón, M.: Sobre lo televisivo: dispositivos, discursos y sujetos, Buenos Aires: La Crujía (2004)