

Investigaciones en comunicación en tiempos de big data: sobre metodologías y temporalidades en el abordaje de redes sociales

Communication research in times of big data: methodologies and temporalities in the approach of social networks

Irene Lis Gindin

CONICET (Consejo Nacional de Investigaciones Científicas y Técnicas). UNR-CIM (Universidad Nacional de Rosario-Centro de Investigaciones en Mediatizaciones)

Mariana Patricia Busso

CONICET (Consejo Nacional de Investigaciones Científicas y Técnicas). UNR-CIM (Universidad Nacional de Rosario-Centro de Investigaciones en Mediatizaciones)

Referencia de este artículo

Gindin, Irene Lis y Busso, Mariana Patricia (2018). 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. Castellón: Asociación para el Desarrollo de la Comunicación adComunica y Universitat Jaume I, 25-43. DOI: <http://dx.doi.org/10.6035/2174-0992.2018.15.3>

Palabras clave

Sitios de redes sociales; big data; metodología; técnicas de recolección de datos; archivo; discurso.

Keywords

Social networking sites; big data; methodology; data collection techniques; archive; discourse.

Resumen

Este artículo, de corte ensayístico, expone una serie de problemáticas relativas a investigaciones en comunicación social, particularmente aquéllas que proponen como objeto de estudio materiales de análisis recabados de sitios de redes sociales en Internet. De aquí se propone un recorrido en torno a tres ejes principales. Por un lado y tomando como disparador dos investigaciones postdoctorales en curso, se plantean las dificultades principales que supone el trabajo de construcción de corpus basado en grandes cantidades de datos. Por el otro, se especifica el modo en que es entendido el *big data* en tanto fenómeno no circunscripto exclusivamente a cuestiones de cantidad y en donde se vuelve necesaria la hibridación entre métodos de análisis computacionales y estrategias más bien artesanales de recolección e indagación. Por último, se reflexiona en torno a la dimensión temporal propia de estos conjuntos de datos, derivada de las modalidades de vinculación online y de las técnicas de recolección empleadas. Lo aquí trabajado propone una mirada crítica ante el optimismo que vincula la disponibilidad del material con el resultado exitoso de una investigación. Además, se resalta que estas dificultades relativas a la construcción de corpus de análisis no refieren solamente a particularidades técnicas sino que hacen mella, también, en los principios epistemológicos y metodológicos que guían las investigaciones.

Abstract

This article, of essayistic tone approaches a series of central issues related to social communication research, particularly those that propose study materials collected from social networking. It proposes a reflection around three main axes. On the one hand and taking two postdoctoral research in progress, the main difficulties involved in corpus construction work based on large amounts of data are presented. On the other, it specifies the way in which *big data* is understood as a phenomenon not circumscribed exclusively to quantity and where it becomes necessary the hybridization between computational analysis methods and artisanal strategies of collection and inquiry. Finally, we reflect on the temporal dimension of these data sets, derived from the online linking modalities and the collection techniques used. What has been worked on here proposes a critical look at the optimism that links the availability of material with the successful outcome of an investigation. In addition, it is emphasized that these difficulties related to the construction of corpus of analysis do not refer only to technical particularities but also influence the epistemological and methodological principles that guide the investigations.

Autoras

Irene Lis Gindin [iregindin@hotmail.com] es doctora y licenciada en Comunicación Social por la Universidad Nacional de Rosario. Es docente de la Facultad de Ciencia Política y RRII y miembro del Centro de Investigaciones en Mediatizaciones de la

mencionada casa de estudios. Su área de interés es el análisis del discurso, particularmente del discurso político.

Mariana Patricia Busso [mar_busso@hotmail.com] es becaria post-doctoral de CONICET. Es miembro del Comité Académico del Centro de Investigaciones en Mediatizaciones (CIM) de la Universidad Nacional de Rosario (UNR). Es doctora en Comunicación Social por la Universidad Nacional de Rosario, con beca de posgrado de CONICET y en cotutela con el Dottorato in Sociologia della Comunicazione e Scienze dello Spettacolo de la Universidad de Urbino (Italia).

1. Introducción¹

El objetivo de este artículo es exponer una serie de reflexiones en torno a una problemática creciente en la investigación en comunicación social y que tiene que ver con aquellas investigaciones que se proponen el abordaje de grandes cantidades de datos extraídos de espacios digitales. Las preguntas e hipótesis aquí planteadas tienen su raíz en dos investigaciones postdoctorales actualmente en curso, a partir de las cuales daremos cuenta de las potencialidades pero también de las limitaciones que suponen el acceso y el análisis de datos obtenidos en sitios de redes sociales (RSI) como Twitter y Facebook.

Consideramos aquí que 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. Sobre este punto, precisamente, es donde nos interesa detenernos en este artículo, abriendo distintos interrogantes surgidos al calor del trabajo empírico con discursos provenientes de Twitter y de Facebook, plataformas caracterizadas —entre otros atributos— por su inmediatez y su ubicación en un *aquí y ahora*. Estos rasgos no sólo definen las lógicas formales y enunciativas de dichas interfaces, sino que también apuntan a ciertas dificultades inherentes al análisis de los materiales allí contenidos. Las preguntas no tardan en aparecer: ¿cómo elaborar conclusiones sobre los materiales que nos devuelve la web si desconocemos la totalidad —o la representatividad— de ellos?, ¿cómo responder a las preguntas de investigación planteadas si el cálculo algorítmico no nos permite acceder a cantidades aún mayores de datos?, ¿qué técnicas de recolección emplear de acuerdo a las particularidades y a las exigencias de cada interfaz?²

Características como la inmediatez y el presente continuo que imponen las plataformas digitales, llevan a las ciencias sociales en general —y a los estudios en comunicación en particular— a confrontarse con dos cuestiones no menores. Por un lado, las RSI suponen nuevas disquisiciones relativas a la conformación de vínculos en dichos espacios. Por el otro, la construcción de los corpus generalmente supone la recuperación de materiales publicados ya con cierta distancia temporal —es decir, materiales *del pasado*—, cuestión que resulta en extremo dificultosa y sobre la que aquí nos detendremos. Esto es así en tanto, si bien es posible recuperar grandes cantidades de material proveniente de Internet *en tiempo real* o incluso con una previsión de futuro, pareciera que relevar datos del pasado se vuelve una maratónica tarea que se encuentra determinada, por un lado, por las destrezas del investigador y, por el otro, por los límites y potencialidades de la web.

¹ Este artículo amplía la presentación realizada por las autoras en el Coloquio Anual del CIM *El cuerpo de los corpus: pluralidad de métodos en tiempos de big data*, organizado por el CIM — Centro de Investigaciones en Mediatizaciones, radicado en la Facultad de Ciencia Política y RR.II. de la Universidad Nacional de Rosario (Argentina), los días 17 y 18 de agosto de 2017.

² Para la distinción entre metodologías y técnicas remitimos a Marradi, Archenti y Piovani (2007).

Esta particularidad que mencionamos, precisamente, conforma una de las principales inquietudes que nos han llevado a la escritura de este artículo. En este sentido, reflexionamos en torno a aspectos epistemológicos, metodológicos y técnicos ligados a la selección y el análisis de grandes volúmenes de datos, y, además, a las problemáticas específicas vinculadas al acceso y a la gestión de materiales *del pasado* que conforman nuestros respectivos corpus de análisis. Al respecto anticipamos aquí que consideramos que la selección y el abordaje de la materialidad de corpus profusos comporta entonces una serie de complejidades que no necesariamente se corresponden con una mirada sincrónica sobre la labor de investigación y que, si bien podrían considerarse como emparentados con problemáticas propias del *big data*, escapan para su análisis a las herramientas propias de técnicas como la minería de datos, y para los que la investigación social cualitativa actual pareciera requerir de un esfuerzo cuasi artesanal de creatividad metodológica.

Tal como adelantamos, la base de estas reflexiones se encuentra en nuestras propias experiencias de investigación, específicamente en dos investigaciones postdoctorales que trabajan sobre discursividades en redes sociales, una en Facebook y otra en Twitter, y que en tal sentido comparten ciertas problemáticas. Por un lado, ambas investigaciones se plantean desde el abordaje propuesto por el *análisis del discurso*, recuperando tradiciones de tipo interpretativas. Por el otro, y como punto a destacar, en ambos casos se construyen objetos que no existirían por fuera de los materiales generados en espacios online. Es en este sentido que en estas dos investigaciones observamos problemas similares respecto del acercamiento a los corpus de análisis, aunque con diferencias. Por mencionar solo un ejemplo, en el caso de la investigación que toma como corpus de análisis a las publicaciones realizadas en Twitter, la configuración de privacidad suele no ser un problema, dado que la mayoría de las cuentas son públicas; mientras que la situación es variable en el caso de Facebook, lo cual plantea un problema extra para el analista, ligado también a cuestiones éticas sobre la privacidad de la información recopilada.

De estas cuestiones nos interesa poner en común los problemas con los que nos hemos topado hasta ahora y que, creemos, no sólo tienen puntos de contacto entre sí sino que nos permiten mirar más allá de nuestras investigaciones particulares y preguntarnos sobre los problemas metodológicos que están permeando las investigaciones en comunicación. A tal fin, hemos organizado este escrito en las siguientes secciones. En primer lugar, presentaremos las problemáticas ligadas a los universos empíricos y a la construcción de corpus a partir de la experiencia brindada por nuestras investigaciones en curso, desde de las cuales surge la necesidad de estas reflexiones. Seguidamente expondremos las problemáticas teórico-metodológicas relativas al abordaje de grandes cantidades de materiales online que guían este trabajo. En tercer lugar, nos detenemos en la dimensión temporal de esas publicaciones, a partir de recuperar conceptos como los de *memoria* y *archivo*. Por último, retomamos y resumimos las ideas principales sobre las que discutimos en estas páginas.

2. Sobre la construcción de corpus en tiempos extensos

Existe actualmente una mirada un tanto optimista que asume que la fácil y rápida accesibilidad a grandes reservorios de material en línea redundará inevitablemente en mayores potencialidades para la investigación en comunicación. En este trabajo, sin embargo, planteamos cierta mirada crítica en torno a ese optimismo. Precisamente, encontramos que en los últimos años se ha producido un asombroso incremento en investigaciones de variada naturaleza (estudios de comportamientos electorales, de activismos en línea, de establecimiento de vínculos colectivos, etc.) que, empleando distintas miradas disciplinares y dispares abordajes metodológicos, hacen uso del llamado *big data* a partir de una suerte de optimismo que liga la disponibilidad de material con el resultado exitoso de una investigación. Esta actitud, además, deja de lado la necesidad que dichos materiales imponen para una mayor rigurosidad metodológica e incluso para una revisión de los presupuestos de investigación.

En efecto, si bien es destacable la relativa facilidad en la disponibilidad de datos y de distintos tipos de materiales circulantes en la Red, creemos que a la hora de encarar una investigación con estos tipos de materiales es necesario no perder de vista sus riesgos y sus límites, que muchas veces rozan incluso la formulación de las preguntas y las hipótesis de investigación, que deben partir ahora de otros presupuestos epistemológicos y teóricos.

En una de las investigaciones que nos ha despertado los interrogantes que planteamos en este artículo, nos proponemos caracterizar el modo en que se construye el liderazgo político en Twitter, a partir del análisis de la cuenta de la ex presidenta argentina Cristina Fernández de Kirchner (@CFKArgentina). Nos interesa, fundamentalmente, analizar nuestro corpus de forma comparativa a partir de dos grandes paquetes de datos: por un lado, los *tweets* correspondientes al segundo mandato³ (2011-2015); por el otro, aquellos correspondientes al período inmediatamente posterior a la asunción de Mauricio Macri (2015-actualidad). Parte del interés que nos convoca se vincula a poder establecer una cierta tipología de *tweets* que no atienda exclusivamente a los períodos de campaña, sobre los que la mayoría de los trabajos se abocan, sino que responda a una cierta regularidad en el tiempo. Además, nuestra intención intenta recuperar algo del orden del lugar institucional de Fernández de Kirchner, de ahí el corpus comparativo planteado (durante y después de su desempeño como primera mandataria).

En la segunda de las investigaciones que se sitúan en el origen de este artículo, el objetivo es el de analizar la construcción de distintas dimensiones identitarias propuestas en grupos de Facebook de argentinos residentes en el exterior (en un primer recorte, conformado por los grupos *Argentinos en el Mundo*, *Argentinos residentes en el exterior* y *Argentinos desparramados por el mundo*). Las publicaciones realizadas en ese marco asumen como objetivo el compartir experiencias, relatos y reflexiones sobre la propia experiencia migratoria. En ese sentido, los

3 Fernández de Kirchner se desempeñó como primera mandataria durante dos períodos consecutivos. El primero, de 2007 a 2011; el segundo de 2011 a 2015 momento en el que fue sucedida por el actual presidente, Ing. Mauricio Macri.

principales objetivos de esta investigación apuntan a reconstruir y analizar la elaboración de clivajes fundados en atributos sociales, en tradiciones culturales y en dimensiones políticas compartidas realizada por los propios usuarios y propuestos como un horizonte identitario común para los argentinos emigrados en Facebook. A su vez, nos proponemos analizar las estrategias de visibilización de la experiencia migratoria de los argentinos en dichas páginas de Facebook, en particular en lo que refiere a las reelaboraciones públicas y mediatizadas de las vivencias íntimas sobre el proceso migratorio.

El relevamiento de los materiales de ambas investigaciones y la construcción de los corpus de análisis se erige como la primera decisión a la que debemos atender en el curso de nuestro trabajo. En efecto, es allí donde aparece cuanto menos cuestionada aquella apreciación que mencionábamos más arriba: la asunción de que analizar materiales provenientes de Internet no conlleva ulteriores dificultades o problemáticas en lo que respecta al acceso a aquéllos, que se supone disponible y libre por el hecho mismo de encontrarse en la Red. Sin embargo, como veremos a continuación, tanto la accesibilidad como el volumen de los datos recabados no necesariamente son cuestiones sencillas, y no necesariamente conducen a la resolución exitosa de los objetivos planteados. Los cuales demandan, por su parte, un esfuerzo de adecuación del aparato teórico y metodológico del investigador al nuevo quehacer empírico.

En cuanto a la construcción de los corpus, nos interesa mencionar que entendemos aquí al corpus como una construcción ligada a los posicionamientos teóricos del investigador, vinculados a su vez a una problemática de análisis (es decir, ese conjunto coherente de proposiciones hipotéticas o de postulados) que determina un objeto y un punto de vista analítico (Charaudeau, 2011). Tal construcción implica entonces procedimientos sistemáticos de recolección y de análisis del material reunido, y 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, que se construye luego del acceso a un conjunto aún mayor de materiales —el universo de análisis—. Hoy por hoy, sin embargo, el acceso al universo de análisis se vuelve cada vez más problemático dado el volumen de los materiales con los que trabajamos; esto es, la tan invocada *abundancia* propia del también en auge *big data* (Fumero, 2013; Magnani, 2017). Y esto es así en tanto es ese universo de materiales significantes el que se encuentra cada vez más supeditado a mecanismos algorítmicos de los cuales resulta difícil establecer justificaciones, o a partir de los cuales nos resulta complejo avizorar la totalidad en la que se despliegan.

La *representatividad* de los materiales, por su parte, se erige como otra cuestión asociada a las problemáticas que depara la construcción de este tipo de corpus (Mancera Rueda y Pano Alamán, 2014). En efecto, al pensar a los materiales disponibles en la Red como un universo que suponemos potencialmente accesible y disponible a la mano del investigador, inferimos que deberíamos ser capaces de analizar *todo* el material disponible o, cuanto menos, que la selección a realizarse

debería ser lo suficientemente significativa de esa gran totalidad. Sin embargo, como apuntábamos, no debemos perder de vista que esta fácil disponibilidad no es tal: no sólo por lo potencialmente inmanejable de los grandes volúmenes de información en base a métodos de análisis cualitativos, sino también porque se trata de un universo de datos en permanente expansión y del cual no conocemos las fronteras. Como veremos en el apartado siguiente, incluso si miramos hacia el pasado, con la esperanza de dar con un material un poco más sedimentado, no hay certeza de que el material compilado sea la totalidad del material existente⁴.

En una instancia inicial, entonces, los problemas que surgen del abordaje de tales materiales empíricos son dobles. Por un lado, el dificultoso acceso al complejo universo de análisis; por el otro, la aparición de nuevos métodos y técnicas de recolección de datos (como ser, la minería de datos, o incluso aplicaciones como Netvizz, Bluenod o Chorus⁵) para su posterior sistematización y organización. En ambos casos, la abundancia del material obtenido suele volver necesario —o cuanto menos recomendable— el uso de métodos computacionales para la recolección, análisis y presentación de los datos. En consecuencia —y como dato no menor—, la complejidad de algunos de esos métodos obliga a los investigadores en ciencias sociales a conocer y aprehender un conjunto de técnicas que hasta el momento le resultaban ajenas.

En una segunda instancia, por añadidura, las complejidades propias del análisis de este tipo de materiales se acentúan. En efecto, y para el caso de investigaciones como las que aquí nos convocan, abordajes interpretativos como el análisis del discurso demandan un acercamiento que podríamos definir como cualitativo y, siguiendo a Fernández (2017), de nivel *micro*⁶, donde la capacidad intuitiva y analítica del investigador es convocada en pos de dar cuenta de las particularidades de esas discursividades. Sin embargo, la gran cantidad de material que suele conformar los corpus extraídos de RSI dificulta esta empresa, enfrentando al investigador a demandas de tiempo y de exhaustividad que muchas veces sobrepasan las reales

4 Para poner un ejemplo, podemos citar aquí la investigación de Ben-David (2016), quien da cuenta de la desaparición de la memoria en Internet de los sitios ligados a un país ya no más existente como la Yugoslavia. Como menciona con crudeza esta investigadora, se trata de una herencia de la que podría dudarse incluso de su existencia misma, si nos guiamos por las trazas dejadas en Internet: de allí fueron eliminadas todas las páginas con el dominio que indicaba esa pertenencia geográfica (.yu). Como propone esta autora, para estos casos es Internet misma como fuente la que aparece cuestionada, ya que no podemos confiar tampoco en el presente de la Red como punto inicial de la actividad de archivo. Allí pasa a ser enormemente relevante la memoria casi vivencial, casi experiencial de los investigadores-usuarios, impulsando la sospecha de que ese pasado que se nos hace accesible en la Red adolece en realidad de omisiones, las que debemos mapear en base a nuestro propio recorrido por Internet y en base a las posibilidades de reconstruir el ecosistema online existente en el momento de nuestro interés de recopilación.

5 Un posible listado de herramientas o recursos para el análisis de sitios de redes sociales puede encontrarse en <http://www.k-government.com/2016/06/28/100-herramientas-analisis-redes-sna-ars/>

6 El autor distingue tres *distancias de observación* para el abordaje de las *nuevas mediatizaciones*: «una perspectiva *macro* desde la que se describen objetos complejos y extensos como la *sociedad*, la *cultura* o el *sistema discursivo* con sus respectivos *elementos de conflictos claves*: clases sociales, estilos de vida o discursivos, acciones, géneros y modos y medios de comunicación; una perspectiva *medium*, más cercana al fenómeno social, en las que observamos *escenas de intercambio* y *conflicto*: relaciones productivas, situaciones de exposición o lectura de obras de arte o equivalentes y de percepción de medios; y una tercera perspectiva *micro* en la que se enfocan *productos en sus procesos*: objetos industriales o artesanales, textos artísticos o mediáticos» (Fernández, 2017: 13-14).

posibilidades del trabajo *manual* o artesanal al que se apelaba en el pasado; permaneciendo la sospecha que ese tipo de trabajo es aún necesario, so pena de no poder penetrar más allá de la superficie de los fenómenos estudiados.

Justamente, como ya hemos anticipado, el objetivo de esta publicación es exponer una reflexión crítica sobre las potencialidades y los límites de investigaciones en comunicación que, desde el análisis del discurso, se plantean el abordaje de una serie de publicaciones en sitios de redes sociales. Si bien, por una cuestión cuantitativa, probablemente no pueden ser consideradas como pertenecientes al universo del *big data*, sí comparten con este último el modo de abordaje de los materiales y las herramientas de recolección de datos a él vinculadas. En este sentido, y en un intento de superar una mirada particularista que atienda exclusivamente a nuestros problemas de investigación, nos proponemos aportar una serie de hipótesis y preguntas que nos permitan avanzar sobre las perplejidades que conlleva la aplicación de los métodos computacionales, al mismo tiempo que cuestionar aquellas certezas que parecieran inamovibles.

3. Interrogantes teórico-metodológicos en torno al *big data*

Hemos ligado nuestras reflexiones a las problemáticas derivadas del ámbito del *big data*, lo cual exige un esfuerzo por delimitar someramente sus alcances. En efecto, consideramos aquí, siguiendo la reflexión semiótica de Fabbri (2015), que cuando hablamos de *big data* no nos estamos refiriendo únicamente a una cuestión de *cantidad* de datos, sino también a la modalidad estratégica del uso de esa información. Ésta no es totalmente irrestricta ni libre, sino que responde a particulares estrategias algorítmicas de recolección, desconocidas para los usuarios y para los investigadores. Su reconocimiento, además, puede hacer que desde las ciencias sociales se plantee la necesidad de recolectar también otros tipos de datos —los *little data*— si lo que se busca es proponer una comprensión más acabada del objeto a analizar.

La evidencia de crecientes investigaciones en el ámbito de la investigación en comunicación que toman como materia prima los materiales obtenidos mediante técnicas ligadas al *big data* impone la delimitación de nuevas estrategias de abordaje teórico y metodológico. Acercamiento a los objetos de análisis realizado no únicamente en función de los datos recolectados, sino también de la forma en que esa recolección ha sido efectuada y en que la propia investigación es concebida.

En este apartado daremos cuenta brevemente de las certezas e interrogantes en tal dirección que han surgido de nuestras investigaciones en curso, con un énfasis en la dimensión temporal que se pone en juego al proponernos estudios que toman a los datos extraídos de Internet como insumo principal. Por ende, no se realiza aquí una exposición canónica de la teoría o de la metodología empleadas, sino que nos proponemos plantear una suerte de ensayo o de metadiscurso sobre ellas, evidenciando sus puntos de fuerza y aquellas cuestiones que se nos han revelado como más endebles y que creemos necesario poner en debate. Como no puede ser de otro modo al estar centrada en trabajos puntuales de las autoras, tal aproximación conlleva un carácter interpretativo y sesgado. Sin embargo, el carácter relativamen-

te novedoso de las técnicas mencionadas y la mutabilidad que puede endilgarse a los objetos de estudio en cuestión hacen relevante la suma de experiencias de casos particulares de investigaciones que se dediquen al tema, así como los interrogantes y las certezas que vayan surgiendo de éstas. Más allá de, por supuesto, la inevitable asistematicidad que comporte esa puesta en común.

En los últimos años, la necesidad de una reflexión metodológica sería acerca de la recolección de datos provenientes de Internet y su posterior análisis ha sido enfatizada en distintas latitudes de la investigación en ciencias sociales, ante lo que pareciera ser la llegada de una mayor cautela a la hora de presentar las potencialidades ligadas al *big data* (entre otros, y sin pretensión de exhaustividad, mencionamos a Androutsopoulos y Beibwenger, 2008; Giglietto y Rossi, 2015; Lewis, Zamith y Hermida, 2013). Los límites o, más bien, las miradas ingenuas sobre sus potencialidades también han sido puestos en relieve por parte de estudiosos de larga data de las RSI, tales como Boyd y Crawford (2012). Coincidimos con estas autoras quienes afirman, recuperando a Manovich, que el *big data* no tiene tanto que ver con *lo grande* de los datos sino, además y de modo fundamental, con la posibilidad de agrupar esos conjuntos y de establecer múltiples relaciones o referencias cruzadas entre ellos. Justamente, estas autoras plantean que el *big data* puede definirse como un fenómeno cultural, tecnológico y académico que descansa en la interacción de:

- (1) Technology: maximizing computation power and algorithmic accuracy to gather, analyze, link, and compare large data sets. (2) Analysis: drawing on large data sets to identify patterns in order to make economic, social, technical, and legal claims. (3) Mythology: the widespread belief that large data sets offer a higher form of intelligence and knowledge that can generate insights that were previously impossible, with the aura of truth, objectivity, and accuracy (Boyd y Crawford, 2012: 663)

Manovich (2012) es otro de los investigadores que plantea diversas objeciones a las miradas optimistas sobre el *big data*; mencionando, por ejemplo, que sólo las compañías de medios tienen acceso a la totalidad de datos, y poniendo en cuestión la *expertise* necesaria para trabajar con grandes masas de datos, especialmente en el campo de las ciencias humanas, no habituadas a trabajar con herramientas cuantitativas, computacionales, para analizar la actual *explosión de datos*. En el primer caso —el hecho de que sólo las compañías de medios puedan acceder y, en todo caso, publicar sus propios datos— se observa claramente en las propias *Application Programming Interface* (API) de las diferentes RSI, que sólo devuelven un porcentaje de la totalidad de materiales. En relación con el planteo precedente de Fabbri, Manovich (2012) pareciera lanzar una advertencia similar a partir de la distinción de dos tipos de datos sobre los cuales se desarrollan investigaciones en distintas esferas de conocimiento: el *surface data* y el *deep data*. En el primer caso, los *datos de superficie*, se vinculan a metodologías cuantitativas que posibilitan el acceso a datos sobre *muchos*. Por el otro, los *datos profundos*, estos utilizados en metodologías cualitativas y que han permitido conocer problemáticas particulares pero con mayor nivel de especificidad (pensemos, por ejemplo, en los trabajos etnográficos). En principio podría decirse, de acuerdo al autor, que con la proliferación de métodos computacionales, el acceso a innumerables cantidad de datos se vuelve más sencillo, lo que supondría

eliminar esta diferencia entre tamaño y profundidad. Sin embargo, la diferenciación permanece en tanto las preguntas de investigación y las potencialidades del análisis dependen de las disciplinas y las miradas teóricas con las que abordamos nuestros objetos; es decir, y tal como venimos sosteniendo en estas páginas, la posibilidad de acceso no va en desmedro de la capacidad analítica del investigador.

En efecto, a la par que es indudable la facilidad que brindan herramientas de este estilo para la recolección y la sistematización de grandes cantidades de información, también hay que considerar que no resuelven todas las problemáticas asociadas a su implementación, y mucho menos que sean capaces de sustituir a las hipótesis y a las intuiciones de los investigadores. Precisamente, sin despreciar las posibilidades que ofrecen las ciencias computacionales para manejar estos cuantiosos volúmenes de materiales, seguimos considerando que, en tanto nuestras preguntas de investigación se centran en la producción de sentido, se hace necesario enfrentar con mirada crítica las nuevas herramientas ligadas al análisis automatizado, que nos exigen también una atención extra a la hora de delimitar nuestro corpus de análisis.

Cabría preguntarnos si no es acaso el propio objeto de estudio el que cambia, en tanto, tal como afirman Boyd y Crawford (2012), el *big data* no se refiere solamente a un gran conjunto de datos y a las herramientas y procedimientos utilizados para manipularlos y analizarlos, sino también a un giro computacional en el pensamiento y la investigación. Quizá, adelantando una conclusión apresurada, debemos resignarnos ante la imposibilidad de acceder al completo universo de análisis y revisar nuestras operaciones de recorte de los corpus, así como replantear las operaciones y los presupuestos que guían la elaboración de las preguntas de investigación para estos materiales. Justamente, si tales preguntas están vinculadas a los datos que tenemos a disposición, la relevancia de aquellos materiales a los que podemos —o no— acceder en las distintas plataformas en línea toma otro cariz. Las posibilidades de acceso a ellos, las oportunidades de desarrollar distintas estrategias de investigación, cambiarán también nuestra forma misma de delinear las investigaciones que los involucren.

De aquí que solemos hablar de una articulación entre métodos computacionales y métodos *artesanales* que mencionábamos más arriba, como una posibilidad de articulación que redunde en una presencia más fuerte del investigador frente a su objeto de estudio. Porque es en esta masa de datos con enormes volúmenes en donde se vislumbra la necesidad de dicha hibridación, ya que, por un lado, las formas tradicionales de *análisis manual* no fueron pensadas para gestionar grandes bases de datos textuales; y, por otra parte, «algorithmic analyses of content remain limited in their capacity to understand latent meanings or the subtleties of human language» (Lewis, Zamith y Hermida, 2013: 35).

Es en esta hibridación entre ambos acercamientos en donde, creemos, se reactualiza un debate que lo precede: el de los métodos cualitativos y los cuantitativos. Preocupados por la producción de sentido, nos vemos obligados a echar mano de técnicas de análisis que puedan cuantificar y ofrecernos una visualización más clara de los materiales a trabajar. No sin contradicciones, claro, dado que —como

veremos a continuación— a medida que pasa el tiempo se va diluyendo el acceso a muchos de los datos publicados en la web, y que se pierden ante la mutabilidad de los distintos soportes y a la escasa previsibilidad archivística de Internet.

4. La dimensión temporal de las investigaciones en RSI: ¿una cuestión de método?

Retomando las investigaciones que guían esta reflexión, nos detendremos ahora en las inquietudes surgidas a partir de los métodos empleados para la recolección del material, pero en las que interviene como factor determinante la *dimensión temporal* del acceso a los datos. Consideramos que tales problemáticas pueden hacerse extensivas a otras indagaciones del campo, en particular a aquellas del ámbito de la comunicación que privilegian el estudio de redes sociales como Twitter y Facebook. En efecto, recordemos que estas reflexiones surgen a partir de las inquietudes derivadas de la problemática del *big data* en dos investigaciones: la primera, centrada en el análisis de los tweets de la ex mandataria argentina Cristina Fernández de Kirchner; la segunda, en el estudio de los posteos de emigrados en grupos de Facebook de argentinos en el exterior.

En el caso del trabajo que toma a los *tweets* como corpus de análisis, como hemos dicho, se centra en el estudio de dos corpus: el primero, un conjunto de *tweets* correspondientes a la segunda presidencia de Cristina Fernández de Kirchner; el segundo, una vez que esta última es sucedida en el cargo por Mauricio Macri.

Los problemas con los que, hasta ahora, nos hemos topado se vinculan con el hilo de esta presentación en tanto se nos ha vuelto muy dificultoso —hasta el momento, imposible— recuperar la totalidad de *tweets* que inicialmente planteamos. Si bien en ningún momento pensamos siquiera factible analizar ese universo, si pretendíamos establecer recortes a su interior, a partir de conocer esa totalidad. Como el objetivo de nuestra investigación se centra en buscar regularidades en un período largo de tiempo, esta imposibilidad de acceso nos genera, cuanto menos, dudas. Las búsquedas avanzadas de Twitter y las diferentes pruebas con *softwares* de recuperación de datos, resultan infructuosas en tanto realizan una selección de la totalidad de *tweets* que vuelve imposible el acceso a la totalidad y, lo que es aún más problemático, vuelve imposible conocer el porqué de esa selección. ¿Cómo caracterizar, entonces, esa discursividad política en base a datos parciales? Dos cuestiones se revelan aquí. Por un lado, el costo de acceso que establecen algunas compañías para acceder a sus datos. Por el otro, y como observan Boyd y Crawford (2012), aunque podemos acceder a un gran número de *tweets* a través de la API de Twitter, la mayoría de los investigadores no obtenemos la *firehose* del flujo de contenido completo, sino simplemente un *gardenhose* de un número muy limitado de *tweets* públicos⁷ (aproximadamente, el 10%) cuya aleatoriedad es totalmente desconocida:

7 Si bien no es un problema en nuestra investigación, dado que la cuenta de @CFKArgentina es pública, sí se torna una dificultad para aquellos investigadores que pretenden el acceso a datos, por ejemplo, a través de *hashtags*. En estos casos, la API de Twitter sólo devolverá contenido de aquellos usuarios que tengan cuentas públicas.

Without knowing, it is difficult for researchers to make claims about the quality of the data that they are analyzing. Are the data representative of all tweets? No, because they exclude tweets from protected accounts. But are the data representative of all public tweets? Perhaps, but not necessarily (boyd y Crawford, 2012: 669)

Muchos investigadores que trabajan discursividad política en Twitter han tomado un camino de relativa fácil solución: bien el análisis de períodos más cortos de tiempo; bien el trabajo con datos a futuro (los casos más usuales son, por ejemplo, las campañas electorales, que combinan ambos aspectos en tanto, por un lado, se llevan a cabo durante un mes y, por el otro, los períodos de campaña están reglamentados por ley, es decir, es posible saber con antelación cuándo se desarrollará la campaña). Son estas cuestiones y tal como lo venimos trabajando en este artículo, las que nos invitan a pensar en la necesidad de adaptar nuestras preguntas y objetivos de investigación a esta nueva realidad empírica, en tanto continuar mirando atónitos el flujo de *tweets* al que no podemos acceder no soluciona los problemas que se nos presentan.

Para la segunda de las investigaciones, esto es, el caso del análisis de los grupos de Facebook, los objetivos propuestos suponen analizar las publicaciones efectuadas en ellos por sus miembros, a fines de reconocer las distintas configuraciones identitarias que distinguen a los emigrados argentinos participantes de tales RSI. Para ello, y a partir del andamiaje teórico-metodológico brindado por el análisis de los discursos sociales, nos planteamos proponer interpretaciones acerca de las regularidades presentes en el corpus seleccionado a través del reconocimiento de las reglas de funcionamiento de esos discursos, a fines de reconocer las auto y las hetero representaciones, los modos de interpelación y de afiliación desplegadas por los migrantes en tales espacios mediáticos.

Las problemáticas metodológicas ligadas a este particular corpus de análisis, además de tener que dar cuenta de un gran volumen de material relevado, nos ubican también en prácticas cercanas a la denominada *etnografía virtual* —entre cuyos exponentes se destacan Hine (2004) y la netnografía de Kozinets (2010) — y a la candente problemática de la gestión de la información sensible de los usuarios (Van Dijk, 2016). En cierta forma, lo que se plantea es el acceso a espacios en línea a los que también se observa, en lo que implica una «tensión entre accesibilidad y libre uso [que] forma parte entonces de los desafíos con los que se encuentra el investigador en la plataforma Facebook» (Capogrossi, Magallanes y Socaire, 2015: 60). El abordaje postulado, entonces, asume también un planteo en una mirada reflexiva —o autorreflexiva— acerca de las potencialidades y las limitaciones que supone el estudio de espacios online, en lo que supone una actualización de la problemática acerca de los distintos niveles de observación del investigador social (Verón, 2013), en la que éste observa a otros observadores y, a su vez, se auto-observa en tales tareas.

Para esta investigación se privilegió la observación —sin participación— de las publicaciones realizadas en los grupos de Facebook seleccionados, una vez obtenida la autorización de los administradores de dichos espacios online en relación a nuestro

interés académico. Del relevamiento de los materiales disponibles en tales espacios, justamente, se desprende una de las mayores dificultades de esta etapa, ligada a la imposibilidad de contar con recursos automáticos para el relevamiento del material publicado, a pesar de tratarse de un material cuantioso y para el cual sería apropiada la utilización de ese tipo de herramientas o aplicaciones. Para este caso puntual, hemos debido realizar una recopilación manual de las publicaciones efectuadas en los grupos seleccionados; número nada despreciable de posteos, realizados en un período relativamente corto de tiempo (del 3 al 13 de noviembre de 2016). De este modo, y volviendo a las cuestiones que nos convocan en este artículo, nos encontramos con la existencia de casos en los que la recopilación del material obliga a la implementación de técnicas más bien artesanales, que entran en conflicto con el volumen de ese material, y que reactualizan la tensión irresuelta entre las cambiantes configuraciones de privacidad y de exploración y las promesas de acceso irrestricto al material online que ya hemos mencionado como propias de una mirada un tanto ingenua sobre el reservorio habilitado por Internet.

De allí, además, se desprende otra de las dificultades con las que nos hemos topado en este período: al no contar con la posibilidad de una recuperación automática de las publicaciones (ya que no todos los tipos de configuraciones grupales en Facebook habilitan al empleo de aplicaciones automáticas de recuperación), y al irse produciendo posteos con una gran celeridad y —en muchos casos— elevado nivel de interacción, la *vuelta atrás* en la *timeline* de los grupos se complejizaba; sumado a ello que se hacía necesario dejar pasar un tiempo prudencial para la recuperación del material, a fines de evitar que nuevas interacciones o respuestas de los participantes a las publicaciones originales hicieran que éstas abandonaran la línea de tiempo cronológica en el discurrir de los muros.

De este modo, tomando en conjunto las cuestiones surgidas al calor de nuestras investigaciones en Twitter y en Facebook, corroboramos nuevamente que las posibilidades vinculadas con el análisis de grandes cantidades de información propias del *big data*, capaces de abrir inéditas perspectivas de desarrollo para muchos ámbitos de la investigación, nos muestra aquí una arista que pareciera ser más bien la de una promesa incumplida, o donde son evidentes los desafíos y los riesgos que aquella conlleva. En particular, nos encontramos con que la recuperación *en el presente* de materiales producidos *en el pasado* de la Red resulta cuanto menos conflictiva, enfrentando al investigador no sólo a desafíos técnicos —cómo realizar esa *vuelta atrás* en la compilación de los materiales— sino también de índole epistemológica —cómo asumir(se) en relación a la temporalidad de Internet.

No es casual, entonces, que traigamos aquí a colación la inquietud por las formas que asume la *memoria* en la Red, entendida no como una actividad de rememoración por parte de sus usuarios (Feld, 2013; Jelin, 2017), sino como las reflexiones que en tanto científicos sociales realizamos acerca de la actualización de contenidos producidos *en el pasado* y que tratamos de recuperar en un *hoy* de la investigación. Precisamente, la posibilidad —o imposibilidad— de recuperar la totalidad de aquellas publicaciones que no se suceden en el instante de su relevamiento —o

bien *hacia adelante*— nos imponen en cierto modo un olvido obligado, imperativo, distinto a ese olvido propio de la actividad memoriosa, siempre selectiva, del que han dado cuenta pensadores como Halbwachs (2005), Jelin (2002) y Rousso (2000), entre otros.

En efecto, y a contramano de la tentación de asimilar las cantidades masivas de datos presentes en Internet con el convencimiento de que *todo* es accesible y analizable, nuestras prácticas de investigación concretas nos enfrentan a la evidencia de que, si bien el material recabado sigue siendo cuantioso y hasta inabarcable, éste adolece de omisiones, de selecciones no intencionadas, que es sin embargo necesario tener en cuenta a la hora de enfrentarnos a un corpus de análisis, y que a su vez hacen del investigador una especie de nuevo archivista⁸, un *constructor de memoria* de la Red.

Precisamente, las problemáticas y las dificultades halladas en la definición de nuestras estrategias metodológicas, así como el esfuerzo que conlleva afinar las cuestiones técnicas a ellas vinculadas, nos llevan a reflexionar sobre nuestra propia práctica en tanto científicos sociales involucrados en una actividad de resguardo de aquello que sucede en Internet. Si asumimos que nuestras sociedades se han complejizado con el avanzar del proceso de mediatización (Verón, 2013) y que, por lo tanto, es necesario considerar de modo integral la relación y la interdependencia entre los sujetos y las tecnologías, entendemos que parte de nuestra labor de investigación reside también en realizar un aporte —por modesto que sea— para el registro de la actividad de los participantes de la Web, y para el resguardo de las discursividades de los distintos espacios en línea. La actividad archivística que suponemos para el investigador, entonces, apunta también a la conservación de esas memorias singulares que correrían el riesgo de desaparecer en la voráGINE de Internet, con el objetivo de restituirlas al flujo de sentidos allí circulantes; ya que compartimos, como sostiene Freire, que «preservar es dar inteligibilidad» (2009: 211).

En nuestras investigaciones hemos sostenido la propuesta de que Twitter y Facebook constituyen no sólo soportes donde es posible construir particulares formas de liderazgo o de identidad, sino también que se erigen como una especie de memorioso archivo discursivo de una época; entendiendo que nuestra contemporaneidad también está signada por una compleja dinámica entre el recuerdo y el olvido, donde «nos encontramos con la posibilidad de que todo lo adquirido o aprendido pueda llegar a perderse, por ello, hemos de conservar las huellas» (Bedoya y Wapenstein, 2011: 12). Esta convicción, a su vez, también es la que nos habilita para proponer la serie de interpretaciones que hemos realizado en el curso de estas páginas sobre esos materiales y sobre las complejidades de los métodos empleados

8 Entendemos al *archivo* como una construcción realizada por el investigador, con sus particulares lógicas de estructuración y sus mecanismos de resguardo: ambos delimitan la exigencia de un orden de lectura preciso, en estrecha vinculación con el espacio y la época en los que se lo hace intervenir. En palabras de Didi-Huberman, el archivo «no es ni reflejo del acontecimiento ni tampoco su demostración o prueba. Siempre debe ser trabajado mediante cortes y montajes incesantes con otros archivos» (2007: 2).

para su compilación y posterior resguardo; decisiones que parten, como no podría ser de otro modo, por problematizar las formas de conocimiento que se ponen en juego en el abordaje de nuestros objetos.

Probablemente el caso de Facebook y de Twitter que nos convoca en nuestras investigaciones actuales no se emparente tanto con la posibilidad de su rápida volatilización que sí encontramos en otras interfaces en línea —pensemos que se trata de los sitios de redes sociales con más perfiles activos en la actualidad⁹, por lo que podemos suponer que su desaparición, al menos, no será abrupta— pero sí se liga con la presencia de los baches, de las lagunas que hemos mencionado, de las que sólo podemos atisbar su existencia pero sin conocer su profundidad o su alcance. En consecuencia, reconocemos que el sentido que podamos darle a los discursos circulantes en Internet será irremediamente parcial y fragmentario; sin embargo, ello no implica menoscabar su relevancia sino asumir que no es factible la pretensión de abarcar completamente cuanto producido en los espacios virtuales. Pero, al mismo tiempo, el reconocimiento de que constantemente múltiples voces se vuelcan a los espacios posibilitados por distintos soportes de Internet, y de que se trata de voces que podrían perderse de no contar con una actividad de resguardo atenta y dedicada, nos lleva a asumir la labor de investigación en este tipo de materialidades como un desafío con implicancias de cuidado, y que justifica con creces los desvelos teóricos y metodológicos que aquél demanda.

5. Breves conclusiones

En estas páginas propusimos un espacio de reflexión que, tomando como excusa investigaciones particulares en curso, ha planteado una serie de interrogantes que hacen eje en ciertas dificultades para el trabajo con grandes cantidades de datos publicadas en sitios de redes sociales. Dado el carácter ensayístico de estas páginas, no proponemos aquí una exposición de resultados, sino más bien que nuestra intención es retomar y resumir las principales aristas de las que hemos dado cuenta en este artículo, proponiendo además posibles derivas futuras. Como expusimos previamente, las dificultades mencionadas lejos están de ser meras particularidades técnicas, sino que cuestionan los propios principios teóricos y epistemológicos de las investigaciones en comunicación social. Resaltamos, además, que este tipo de trabajo supone la hibridación entre métodos de análisis computacionales y métodos más bien artesanales, requiriendo del investigador, por un lado, la aprehensión de técnicas hasta el momento desconocidas y, por el otro, una atención particular que le permita superar «una especie de naturalización metodológica que de alguna forma oculta, mediante el uso del instrumental técnico y terminológico, la problemática de la validez relativa de las conclusiones que se obtienen en cada investigación» (Miceli, 2008: s/p).

9 Ver al respecto https://www.alexa.com/topsites/category/Computers/Internet/On_the_Web/Online_Communities/Social_Networking. Consultado el 26/09/2017.

Cuanto dicho acerca de las perplejidades técnicas de las que hemos dado cuenta más arriba no debe hacernos circunscribir la problemática de la gestión y el abordaje de corpus únicamente a una cuestión de mayor o menor habilidad del investigador en el manejo de herramientas computacionales. En efecto, reflexionar sobre una problemática del estilo impone también, creemos, reconocer que en ella se abordan cuestiones propias del rol central de las NTICs en la organización y el funcionamiento de nuestras sociedades contemporáneas. En particular, sostenemos que lo instantáneo y lo cambiante propio de los corpus online, entendidos como rasgos de un cierto *espíritu de época* también moldeado por las matrices de vinculación y de subjetividad individual y colectiva propuestos en las distintas interfaces de Internet, impulsan de ese modo también una cierta mirada sociológica acerca de estos objetos que nos desvelan.

En la mirada crítica que propusimos en este artículo, también sostuvimos que los rasgos ligados a lo efímero propio de las plataformas aquí estudiadas, lejos de invalidar una investigación, contribuyen a cimentar su riqueza: no sólo porque su objeto sencillamente correría el riesgo de desaparecer si aguardamos a una mayor estabilización, sino también porque es el estatuto mismo de la temporalidad de los materiales y de los textos el que es puesto en juego en las reflexiones que nos demandan esas cualidades en tanto científicos sociales.

Por último, y dado que una de las problemáticas centrales que abordamos se vincula con el *pasado en Internet*, dimos cuenta acerca de la temporalidad, la memoria y el archivo. En este sentido, si hablamos del registro de una temporalidad en relación con lo efímero no se debe a una búsqueda de contradicciones caprichosas, sino que obedece al convencimiento de que es necesario fortalecer lo que denominamos la *memoria de Internet*, estrechamente ligada a la actividad de conservación y resguardo del material allí alojado, y que va más allá de la problemática de la memoria de los actores sociales que utilizan ese dispositivo.

Sabemos que no se agotan aquí las preguntas que hemos propuesto, así como tampoco las respuestas esbozadas implican un cierre definitivo a estas cuestiones. Sin embargo, creemos que dar cuenta de estas problemáticas puede representar un aporte para las investigaciones del ámbito de la comunicación social que, cada vez más, se basan en el análisis de grandes cantidades de datos. Para ellas, sin dudas, la exigencia de reflexionar y de fortalecer los aparatos teóricos y metodológicos sobre los que se erigen las distintas indagaciones contribuirá a hacerlas más ricas y sólidas, a la par que seguramente permitirá dar cuenta con otra amplitud de miras de los nuevos objetos de estudio que surgen con el complejizarse de las sociedades contemporáneas.

6. Referencias bibliográficas

Androutsopoulos, Jannis; Beißwenger, Michael (2008). Introduction: Data and Methods in Computer-Mediated Discourse Analysis. En: *Language@Internet*, vol. 5. Germany: Ministry of Science and Research of North Rhine-Westphalia, 1-7.

- Bedoya, María Elena; Wappenstein, Susana (2011). (Re)pensar el archivo. Presentación del dossier. En: *Íconos. Revista de Ciencias Sociales*, n° 41. Quito: Flacso, 11-16.
- Ben-David, Anat (2016). What does the Web remember of its deleted past? An archival reconstruction of the former Yugoslav top-level domain. En: *New Media & Society*, vol. 18, n° 7. UK-USA: Sage, 1103-1119. DOI: <https://doi.org/10.1177/1461444816643790>
- boyd, danah; Crawford, Kate (2012). Critical questions for big data. En: *Information, Communication & Society*, vol. 15, n° 5. London: Routledge, 662-679. DOI: 10.1080/1369118X.2012.678878
- Capogrossi, María Lorena; Magallanes, Mariana Loreta; Socaire, Florencia (2015). Los desafíos de Facebook. Apuntes para el abordaje de las redes sociales como fuente. En: *Revista de Antropología Experimental*, n° 15. España: Universidad de Jaén, 47-63. DOI: <https://dx.doi.org/10.17561/rae.v0i15.2390>
- Charaudeau, Patrick (2011). Dize-me qual é teu *corpus*, eu te direi qual é a tua problemática. En: *Diadorim: revista de estudos linguísticos e literários*, vol. 10. Rio de Janeiro: Universidade Federal do Rio de Janeiro - Faculdade de Letras, 1-23. Disponible en <https://revistas.ufrj.br/index.php/diadorim/article/view/3932>. Consultado el 04/09/2017.
- Didi-Huberman, Georges (2007). Das Archiv brennt. En Didi-Huberman, Georges y Ebeling, Knut (eds.). *Das Archiv brennt*. Berlin: Kadmos.
- Fabbri, Paolo (2015). La semiótica del camuflaje. Entrevista a Paolo Fabbri por Autor. En: *L.I.S. Letra, Imagen, Sonido. Ciudad Mediatizada*, n° 13. Buenos Aires: Facultad de Ciencias Sociales — UBA, 87-96.
- Feld, Claudia (2013). La memoria, esa vieja cuchara. En: *Prácticas de Oficio*, n° 11-12. Buenos Aires: Universidad Nacional de General Sarmiento — IDES, 1-4.
- Fernández, José Luis (2017). Las mediatizaciones y su materialidad: revisiones. En Fernández, José Luis (ed.). *Mediatizaciones en tensión: el atravesamiento de lo público*. Rosario: UNR Editora.
- Freire, Cristina (2009). Artistas/curadores/archivistas: políticas de archivo y la construcción de las memorias del arte contemporáneo. En Freire, Cristina; Longoni, Ana (orgs.). *Conceptualismos del Sur/Sul*. São Paulo: Annablume — USPMAC - AECID.
- Fumero, Antonio (2013). Big Data. Retórica de la abundancia en tiempos de escasez. En: *Telos: Cuadernos de comunicación e innovación*, n° 95. Madrid: Fundación Telefónica, 91-93.
- Giglietto, Fabio; Rossi, Luca (2015). Limiti e possibilità degli online social data. En: *Sociologia della Comunicazione*, n° 49. Milano: Franco Angeli, 9-18.
- Halbwachs, Maurice (2005). Memoria individual y memoria colectiva. En: *Estudios*, n° 16. Córdoba: CEA, 163-187.

- Hine, Christine (2004). *Etnografía virtual*. Barcelona: UOC.
- Jelin, Elizabeth (2002). *Los trabajos de la memoria*. Buenos Aires: Siglo XXI.
- Jelin, Elizabeth (2017). *La lucha por el pasado. Cómo construimos la memoria social*. Buenos Aires: Siglo XXI.
- Kozinets, Robert (2010). *Netnography: Doing Ethnographic Research Online*. Londres: SAGE.
- Lewis, Seth; Zamith, Rodrigo; Hermida, Alfred (2013). Content Analysis in an Era of Big Data: A Hybrid Approach to Computational and Manual Methods. En: *Journal of Broadcasting & Electronic Media*, vol. 57, n° 1. London: Routledge, 34-52. DOI: 10.1080/08838151.2012.761702
- Magnani, Esteban (2017). Hacete amigo del big data. En: *Revista Anfibia*. Buenos Aires: Universidad Nacional de San Martín.
- Mancera Rueda, Ana y Pano Alamán, Ana (2014). Las redes sociales como corpus de estudio para el Análisis del discurso mediado por ordenador. En: *Janus. Anexo 1: Humanidades Digitales: desafíos, logros y perspectivas de futuro*. A Coruña: Facultad de Filología - Universidade da Coruña, 305-315.
- Manovich, Lev (2012). Trending: The Promises and the Challenges of Big Social Data. En: Gold, Matthew. *Debates in the Digital Humanities*. Minnesota: University of Minnesota Press.
- Marradi, Alberto; Archenti, Néida y Piovani, Juan Ignacio (2010). *Metodología de las ciencias sociales*. Buenos Aires: Cengage.
- Miceli, Jorge E. (2008). El problema de validez en el análisis de redes sociales: Algunas reflexiones integradoras. En: *REDES-Revista hispana para el análisis de redes sociales*, vol. 14, Barcelona: Universidad Autónoma de Barcelona, 1.
- Rouso, Henry (2000). El duelo es imposible y necesario. Entrevista por Claudia Feld. En: *Revista Puentes*, vol. 1, n° 2. La Plata: Centro de Estudios por la Memoria, 30-39.
- Van Dijk, José (2016). *La cultura de la conectividad. Una historia crítica de las redes sociales*. Buenos Aires: Siglo XXI Editores.
- Verón, Eliseo (2013). *La semiosis social, 2. Ideas, momentos, interpretantes*. Buenos Aires: Paidós.