XIII Jornadas de Sociología. Facultad de Ciencias Sociales, Universidad de Buenos Aires, Buenos Aires, 2019.

La construcción del big data en la prensa digital argentina.

Gastón Becerra.

Cita:

Gastón Becerra (2019). *La construcción del big data en la prensa digital argentina. XIII Jornadas de Sociología. Facultad de Ciencias Sociales, Universidad de Buenos Aires, Buenos Aires.*

Dirección estable: https://www.aacademica.org/000-023/41

Acta Académica es un proyecto académico sin fines de lucro enmarcado en la iniciativa de acceso abierto*. Acta Académica fue creado para facilitar a investigadores de todo el mundo el compartir su producción académica. Para crear un perfil gratuitamente o acceder a otros trabajos visite: https://www.aacademica.org*.

La construcción del big data en la prensa digital argentina

Becerra Gastón (UBA/ UAI/ CONICET)
gastonbecerra@sociales.uba.ar

Eje 1: Filosofía, Teoría, Epistemología, Metodología
Mesa 9: Big data y ciencias sociales

Resumen: Se presenta una indagación en torno a la construcción temática y discursiva del big data por parte de la prensa digital argentina, siguiendo los lineamientos teóricos de la sociología sistémica de Niklas Luhmann. Haciendo uso de técnicas de datamining y procesamiento computacional integrados a un diseño interpretativo sobre 2.058 noticias, buscamos responder cuáles son los contextos discursivos en los que se sitúa al big data, las distintas maneras en que se lo define, y las metáforas e imágenes con las que se lo evoca e ilustra.
Keywords: big data, medios masivos, prensa digital, construcción mediática, Niklas Luhmann

Introducción

En este trabajo buscamos explorar y reconstruir algunos de los contextos en los que se enmarca al big data en las comunicaciones de los mass media la prensa argentina. Este objetivo se sitúa en un proyecto más amplio que busca describir y comparar las diferentes maneras en que el big data se presenta en distintos ámbitos de comunicaciones sociales, como la ciencia, la política, o la economía (Becerra 2018). Dicho objetivo rector sigue la perspectiva sistémica y constructivista de Niklas Luhmann (1998, 2007) que propone que el sentido de distintos fenómenos sociales complejos –como el big data– no es unívoco sino que se debe buscar en el entrecruce de dichas comunicaciones.

En *La realidad de los medios de masas* Luhmann (2000) nos presenta una descripción de los medios de comunicación masiva en clave de sistemas sociales autopoiéticos. De acuerdo con el autor, el sistema se constituye a partir de aquellas comunicaciones que se sirven de medios técnicos de reproducción masiva para propagar la comunicación más allá de la presencia física, con el problema de que esto implica la imposibilidad de coordinar la interacción y el ajuste de exigencias y expectativas comunicativas entre emisor y receptor. Debiendo suponer la aceptación de su propuesta comunicativa, los mass media estandarizan sus mensajes en formas que modulan las expectativas, toman como punto de referencia a los mismos productos que ellos han generado, y construyen el perfil de sus propios receptores. Aquí nos interesa una forma particular de comunicación que se ha diferenciado al interior del sistema de los mass media: las noticias. En la teorización de Luhmann,

los mass media cumplen la función de dirigir la autoobservación de la sociedad, actualizando constantemente la frontera entre lo conocido y lo desconocido, impregnando la comunicación con marcas y símbolos que le otorgan sus características salientes. Para el caso de las noticias, estas marcas designan a lo observado como novedoso, relevante o urgente.

Las comunicaciones de los mass media se organizan en torno a temas. Los temas funcionan como contextos selectivos más acotados en los que se reduce la complejidad del sentido, y en los que se estructuran las expectativas de la comunicación. Sin temas no habría necesidad de mayor información acerca de algo, ya que no habría trasfondo sobre el cual evaluar el carácter de novedad o sorpresa, o el nivel de coherencia, consistencia o completitud de cada comunicación. Para nuestros intereses comparativos los temas tienen la ventaja de ser designaciones abstractas, de modo que pueden cruzar diferentes sistemas sociales, obteniendo un tratamiento propio en ellos. Los temas subsisten en la medida en que susciten interés y posicionamientos, y el éxito de los mass media depende de su capacidad para imponer la aceptación de temas.

La retórica del big data

En lo que sigue nos interesa explorar la "retórica" del big data, entendiendo a ésta como una construcción discursiva con un cierto fin, o para utilizar el vocabulario teórico de Luhmann, una "semántica", para referir a una estructura discursiva en el medio del sentido, que se almacena en la memoria (cultura), como insumo y como producto de la comunicación. En otras palabras, la semántica es un conjunto de temas con el que entendemos un fenómeno social.

Comúnmente se suele definir al big data mencionado cuestiones tecnológicas, ya sea porque se indica la disponibilidad de grandes volúmenes de datos en diversos formatos, la proliferación de nuevas técnicas para su procesamiento, y el desarrollo de una infraestructura de sistemas capaz de soportar todo esto. Las 3's V –volumen, variedad y velocidad–, un latiguillo que suele ser presentado al público masivo en lugar de definición, son un reflejo de esta limitación. Sin embargo, esta definición tecnológica está lejos de agotar el fenómeno. El big data es, ante todo, un fenómeno social y cultural que se ha ido consolidando junto al desarrollo tecnológico (Becerra and Alurralde 2017). Socialmente, se trata de una nueva expresión del capitalismo que tiene a los datos como su principal activo. Ciertamente la tecnología ha jugado un rol central en el surgimiento de este escenario. Pero esto no es más que decir que socialmente hemos adoptado a la tecnología como medio en nuestras relaciones con otros, con el mundo material o natural, y hasta con nosotros mismos. Culturalmente, asistimos a la irrupción de un discurso polifónico en el que se proyectan diversas fantasías, miedos y angustias en torno al escenario social descrito y a sus usos y potencialidades. En lo que sigue, nuestro foco estará puesto en esta retórica.

La retórica del big data circula en diferentes espacios y contextos de comunicación, de manera fragmentada y hasta conflictiva, y por la incidencia de diversos actores con muy variados intereses. Explorar, sistematizar y criticar estos discursos es un desafío novedoso para las ciencias sociales, uno que, hasta conocemos, no ha sido adoptado empíricamente (tal vez, si ensayisticamente). Luego, el lineamiento metateórico que indica que big data no tienen un significado unívoco o inherente, sino que su significación se debe buscar en forma local, no debe bloquear un análisis del sentido social en el cruce y la sinergia de estos discursos.

Un buen ejemplo de este tipo de análisis se observa en el libro *The Data Revolution* de Rob Kitchin (2014). Allí el autor puntualizó 4 grandes áreas de la comunicación en los que la retórica del big data irrumpe con nuevas tareas y necesidades: el espacio de la comunicación política y el Estado, con la tarea de gobernar personas, y donde se presenta al big data en relación a una administración más eficiente y transparente, así como con necesidades puntuales, como la vigilancia y la seguridad; el espacio de las organizaciones, donde la tarea es el manejo eficiente de decisiones, y su lugar es ofrecer información confiable, detallada y en tiempo real; otro espacio de comunicación que cruza diversos sectores es el del planeamiento y la gestión urbana, y su tarea es generar un mejor espacio, donde el big data se identifica con el proyecto de las smart cities; finalmente, en el espacio de la comunicación comercial, donde la tarea es agregar valor y producir capital, la estrategia discursiva ronda en dotar de inteligencia a toda la cadena comercial. En un segundo momento, y haciéndose eco de la noción de régimen discursivo de Foucault, advierte que a través de estos discursos hay un cierto intento por justificar y naturalizar la adopción de estas prácticas, generando una cierta atmósfera de deseabilidad e interés por una nueva forma de entender y manejar todos los aspectos de la vida social, con la característica saliente de estar dirigido a toda la sociedad en su conjunto y a todos los ámbitos comunicativos o sistemas sociales a la vez. En el plano discursivo este intento cuenta con dos facilitadores centrales (boyd and Crawford 2012; Lupton 2015; Portmess and Tower 2015). En primer lugar, hay una cierta vaguedad en sus términos que se traduce en una apertura de significado: ni “big” ni “data” tienen un sentido intrínseco que los devuelva a un contexto originario y particular. En segundo lugar, comúnmente cuando el big data se introduce en un campo discursivo particular, se recurre a metáforas que “bajan” lo abstracto del concepto a una imagen vívida capaz de interpelar.

Esto estrategia también se puede rastrear en el ámbito más acotado de las comunicaciones de los mass media. Así, por ejemplo, Puschmann y Burguess (2014) reseñaron varios artículos digitales de periódicos, de sitios de difusión científica y tecnológica, y de publicaciones en sitios de negocios, todos ellos publicados en 2012, y registraron dos grandes metáforas recurrentes: la primera, del big data como una fuerza de la naturaleza -entre las imágenes líquidas: una inundación, un océano, un

torrente, un tsunami; o entre las etéreas: una nube, una explosión- a ser controlada; y la segunda, la del big data como un alimento a ser consumido, lo que en cierta forma completa a la anterior en la línea de un "recurso natural", o incluso un combustible capaz de potenciar un artefacto. Todas estas imágenes tienen la ventaja de oscurecer el carácter humano del big data, y de ofrecer un marco épico a las posibles acciones en torno a ellos. Que hay una estrategia discursiva detrás de estas metáforas se vuelve evidente en el momento en que la metáfora encuentra su límite: frente al recurso de las imágenes. Así, en otro estudio, Pentzold, Brantner y Fölsche (2019) analizaron ilustraciones de artículos de The New York Times y The Washington Post del 2010-2016 para identificar (entre otros objetivos) qué tipo de entidades quedaban visibilizadas. En sus resultados destacan la preeminencia de fotografías de diversos actores sociales (usuarios, referentes, profesionales, etc.) en varias situaciones de uso y/o en contextos sociales más amplios (empresas, agencias, juzgados, etc.), y también fotografías de expresiones materiales de la tecnología (computadoras, capturas de apps, logos de empresas, etc.); en segundo lugar, se encuentran los dibujos, infografías y esquemas centrados en números, procesos, códigos y otras visualizaciones artísticas. Los autores sugieren que esta representación busca dar un "rostro humano" al big data. En comparación, es claro que esta estrategia discursiva es más beneficiosa que insistir con la imagen de inundaciones o tsunamis.

En las secciones que siguen presentamos una primera exploración de la tematización del big data por parte de los mass media. Dado que nos encontramos en la etapa inicial de esta indagación, las preguntas que proponemos son de carácter exploratorio:

> **RQ1**. Cuáles son los temas en los que se alude al big data?
>
> **RQ2**. Cómo se lo define explícitamente y cómo se lo alude? Cuál es la orientación semántica de estas definiciones? Y con qué tipo de imágenes se acompañan las notas?

Por cuestiones de espacio, en esta comunicación las preguntas de RQ2 se reportan muy brevemente.

Metodología

Construcción del corpus

Nuestro corpus comprende noticias publicadas en forma digital en periódicos y sitios de noticias argentinos. Para su construcción se listaron los periódicos digitales incluidos en el portal Diarios Argentinos para las categorías general, regional, economía, política y deportes. Luego, se construyó un agregador capaz de buscar en internet los primeros 100 resultados para la búsqueda "big+data" en cada uno de ello, ordenados por su relevancia. Esta es una decisión metodológica criteriosa y discutible: aquí preferimos limitarnos a usos explícitos de este criterio ya que nos interesa aclarar sus límites, sin negar que otros criterios textuales ("grandes datos", "inteligencia artificial") podrían ser relevantes; luego, al dejar el orden en manos del buscador no hacemos más que reinsertar en

nuestro corpus el carácter performativo de tales herramientas en la cultura (Kitchin 2017). De esta base original se eliminaron aquellos resultados provenientes de sitios con menos de 15 noticias. Luego, se recuperó el contenido de estos artículos, y con una rutina dedicada a analizar y recortar su código HTML, filtramos sólo aquellos resultados que incluían nuestro criterio de búsqueda en el cuerpo de la noticia. El resultado es la base de 2.058 noticias, cuya evolución en el tiempo se observa en la figura 1.

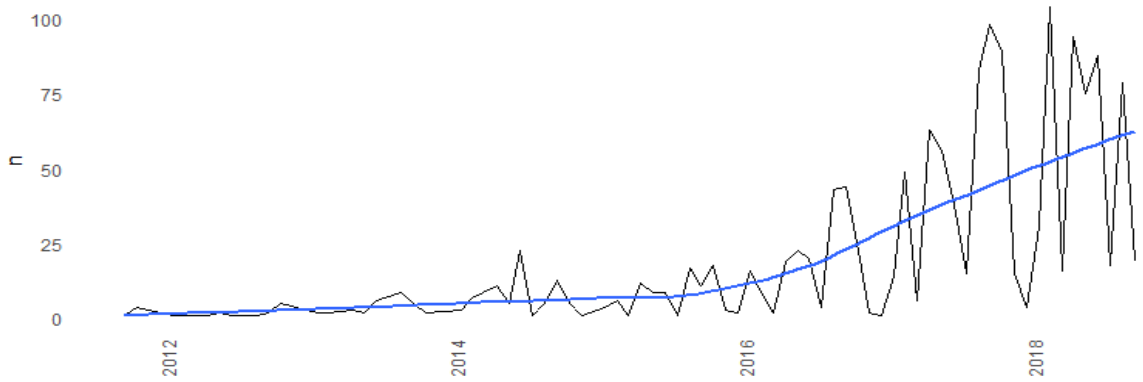

Figura 1. Evolución mensual de 1.476 noticias (582 noticias sin fechas removidas).

Para el análisis de las formas de definir big data, se construyó una base de oraciones con referencias (implícitas y explícitas) al término big data. Esta base incluyó 4.058 oraciones. Para el tratamiento de las imágenes, optamos por recurrir a las ilustraciones posteadas en redes sociales (Facebook o Twitter) para cada artículo. Se construyeron galerías por tópico que incluían las imágenes de los primeros 100 artículos (cada galería incluyó cerca de 50 imágenes).

Técnicas

Para responder la exploración de temas (RQ1) utilizamos un enfoque no supervisado de clasificación, denominado *topic modeling*. Aquí trabajamos con el modelo Latent Dirichlet Allocation (LDA) propuesto por Blei, Ng y Jordan (2003). Este modelo entiende a cada tópico como una distribución probable de todas las palabras (el vocabulario del corpus), considerando su co-ocurrencia en los documentos; y asume que cada documento presenta una cierta proporción de cada tópico. El número de tópicos es introducido por el investigador como un parámetro. Esta es otra decisión metodológica criteriosa y discutible, para que la operamos del siguiente modo: en esta ocasión, hicimos varios intentos con soluciones de 9, 12, 24, 48, 64 y 100 tópicos; en todos los casos se optó por, en un primer momento, seleccionar algunos tópicos significativos, y luego reagruparlos en grandes grupos de tópicos, dado que entendemos –siguiendo a Luhmann– que los sistemas sociales al operar en el medio del sentido tienden a la reducción de complejidad; en todas las soluciones exploradas, los distintos grupos de tópicos construidos siempre se mantuvieron por debajo de 6. Dado este escenario, y por una cuestión de simplicidad, en el análisis final trabajamos

con la solución de 12 tópicos. Luego, para los análisis, se seleccionó una muestra aleatoria de 40 artículos entre los primeros 200 de cada tópico (este es un número un poco mayor al promedio de artículos que rankean primero para cada tópico). El procedimiento interpretativo siguió laxamente los lineamientos metodológicos de la *grounded theory* (Glaser and Strauss 1967). En base a los resultados que fueron arrojando estos análisis, en distintos momentos construímos listas de palabras ad-hoc que denotaban alguna forma de referir al big data, y que fueron utilizadas para comparar frecuencias entre tópicos. Para esta exploración de tópicos el pre-procesamiento (Denny and Spirling 2018) consistió en remover números, símbolos y puntuaciones, convertir a minúsculas, eliminar stopwords, lemmatizar, y finalmente filtrar verbos (excluyendo ser, tener, hacer, estar, haber y decir, todos ellos de gran frecuencia), sustantivos, adjetivos y pronombres (en dicho orden).

Para responder la exploración de definiciones (RQ2) construímos redes asociativas enriquecidas y galerías de imágenes por tópicos. Las primeras son dispositivos gráficos que muestran las principales correlaciones entre palabras, tanto en primer como en segundo grado, de un modo similar al que se usan en algunas técnicas proyectivas de psicología social. Adicionalmente, aplicamos el *lexicon* de Whissell (2009) para evaluar cada término su imaginabilidad (la facilidad para dar sentido a cada palabra), adecuado al castellano (Gravano and Dell'Amerlina Ríos 2014). Para esta exploración el único pre-procesamiento que se realizó fue el de remover números y símbolos, y convertir a minúscula, además de filtrar las palabras por grado de correlación y frecuencia, a fin de dejar un gráfico legible.

Resultados

Contextos temáticos del big data

La figura 2 muestra el resultado de la solución de 12 tópicos y sus 10 principales palabras distintivas, así como el número de artículos que lo tienen por principal tópico (Rank1), la cantidad de oraciones que dicen "big data" y su porcentaje, y el grupo de tópicos al cual fue asignado. La figura 3 muestra la variación entre tópicos en el tiempo de los artículos que tienen un *score* superior a 0.3.

Topic	Términos	Rank1	Oraciones c/ "big data"	% oraciones	Grupo de tópicos
topic_001	ciudad argentina país provincia buenos aires nacional gobierno bigdata proyecto	188	308	8.22%	5
topic_002	detalles Oct equipo jugador real fútbol ganar año partido juego	107	181	4.66%	3
topic_003	tecnología nuevo internet sistema dispositivo inteligente seguridad red tecnológico tiempo	146	224	6.07%	3
topic_004	trabajo profesional tecnología nuevo universidad futuro educación conocimiento desarrollo empresa	195	297	5.79%	4
topic_005	tecnología producción campo productor salud sistema vio médico enfermedad agricultura	128	270	5.74%	3
topic_006	digital contar social medio comunicación nuevo evento edición contenido redes_social	197	304	8.46%	-
topic_007	gobierno política político campaña macri público medio elección presidente poder	179	249	3.41%	5
topic_008	dato información bigdata análisis permitir decisión analizar utilizar grande uso	210	977	16.24%	1
topic_009	usuario facebook dato persona millón empresa información google twitter usar	175	284	5.13%	2
topic_010	año país economía millón nuevo financiero económico dólar sector global	116	159	3.65%	5
topic_011	ver vez saber mismo vida año pasar mundo tiempo persona	168	373	4.55%	-
topic_012	empresa cliente negocio servicio compañía digital producto mercado año tecnología	249	432	5.62%	4

Figura 2. Resultados de los 12 tópicos

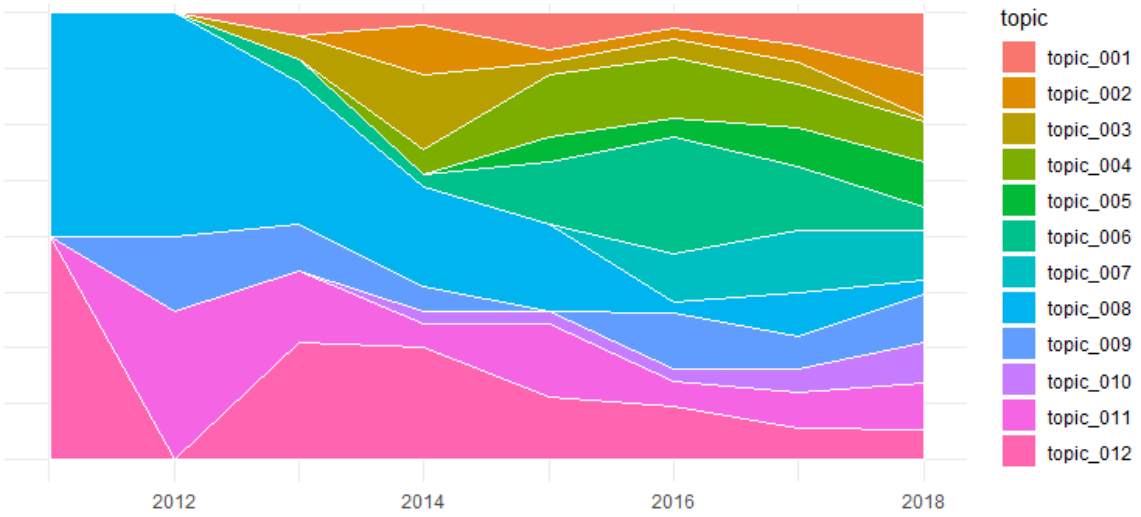

Figura 3. Variación de artículos con score en algún tópico > .3 en el tiempo

Desde aquí desestimamos el tópico #6 que refiere mayormente a un evento particular (Social Media Day) que se presentó en los artículos por medio de contenido no editorial. Con respecto al tópico #11, uno podría inferir un conjunto de noticias tendientes a la reflexión humanística del big data –con palabras tales como mismo, cosa, mundo, saber, persona, humano–. No obstante, el análisis de los artículos refleja una heterogeneidad tan grande que optamos por tratarlo como un tópico no significativo. Esta puede ser una limitación de haber trabajado con un número de tópicos tan bajo, generando una suerte de categoría residual.

Los restantes 10 tópicos se pueden agrupar en 5 grandes grupos que tratan de diversos temas (que en algunos casos remiten a secciones de noticias) y que dan distintos marcos a las menciones al big data:

1. Big data como tema focal (tópico #8);
2. Redes sociales y los algoritmos (tópico #9);
3. Industrias y áreas específicas (tópicos #2, 3 y 5);
4. Management y profesionales (tópicos #12, 4);
5. Política (tópicos #1, 7 y 10).

El primer grupo que nos interesa está constituido sólo por el tópico #8. Los primeros términos que contribuyen a este tópico –dato, información, bigdata, herramienta, permitir, tiempo, análisis, decisión– suelen aparecer en menciones muy genéricas del big data. Se trata, en una proporción notable, de términos que también aparecen entre los más centrales de otros tópicos, de modo que están lejos de ser propios de un contexto específico de discusión o aplicación del big data. El análisis de contenido nos arroja que nos encontramos con piezas focales que tienen por objetivo introducir al big data a un público masivo. Entre los títulos de estos artículos se repiten preguntas como "¿Qué es el Big data?" o definiciones del tipo "Big Data: la ciencia de datos clave dentro de toda

organización" o "Big data, un camino de ida en marketing". Incluso, como se observa en la figura 4, es mucho más elevada la proporción de artículos que incluyen "big data" en el título, en comparación con los demás tópicos.

En estos artículos se observan repetidamente dos momentos discursivos. El primero es la insistencia en números, medidas y otras nociones que no escapan a la idea de una cuantificación. Como se observa en la figura 4, en este tópico es elevado el porcentaje de artículos que hacen mención a una unidad de medida de información, como terabyte o petabyte, cuya manipulación fue el problema originario de aquellas empresas que impulsaron el término big data (Diebold 2012). Es en este contexto en el cual Douglas Laney popularizó las 3 V's, que tienen una presencia notable en los artículos focales. Uno podría conjeturar que estas fórmulas presentan una ventaja insondable para la lógica sistémica que hemos descrito anteriormente: todas ellas parecen pretender ser inherentemente informativas, en tanto suponen un trasfondo de comparación que les da sentido, y a la vez una cierta progresión que les otorga novedad. El segundo momento de la presentación del big data consiste en explicitar posibles beneficios, potencialidades o utilidades. No obstante, esto está lejos de ser privativo de este tópico. Más bien lo propio del tópico es la manera simplificada y genérica en que éstos se presentan, desde simples listas, hasta el recurso de preguntas retóricas –"¿Por qué es importante?", "¿Cuál es su valor?"–. En cualquier caso, se trata de beneficios muy abstractos, en los que se destaca la toma de mejores decisiones. El problema de esta retórica es que sin un contexto particular en el que se encuentre anclado, la toma de "mejores" decisiones sólo puede ser una necesidad relativa, que supone una crítica implícita a una forma "normal" de gestionar organizaciones de todo tipo. Contra este trasfondo, es característico de este tópico que las oportunidades del big data se presenten como un desafío, antes que como una imposición o un deber –como se observa, por ejemplo, en el tópico #4 del grupo management y profesionales–. Uno puede conjeturar aquí que esto es propio de un discurso que busca masificar el interés por el big data. Luego, para reforzar esta estrategia no faltan las citas de autoridad, ya que es común que los artículos incluyan entrevistas a expertos y especialistas, o que directamente tengan contenidos firmados por profesionales de empresas que ofrecen servicios en torno al big data, o de instituciones que la investigan y/o promueven.

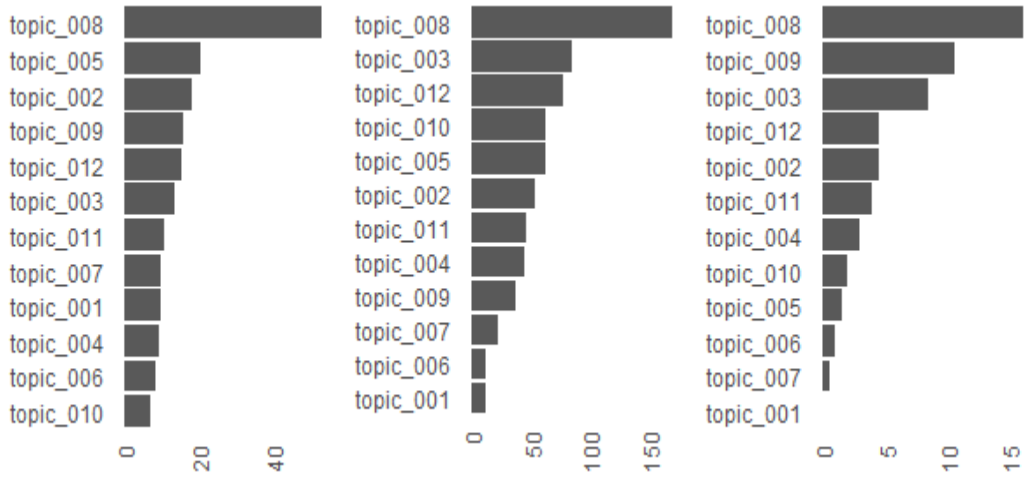

Figura 4. % de títulos que incluyen "big data" en el título (izq.) e incidencia de los términos "volumen", "variedad" y "velocidad" (centro), y % de documentos que incluyen una unidad de medida de datos (*byte) en los primeros 200 artículos de cada tópico.

El segundo contexto en el que se tematiza el big data refiere a las redes sociales y aplicaciones. El tópico #9 incluye términos propios de consumos digitales –como usuario, perfil, redes, personal, gustar Facebook, Twitter, Google–, así como otros ligados al tópico ya reseñado -millón, dato, información-. La revisión de los artículos nos permite observar que se trata de noticias centradas en plataformas y redes sociales, y que estas se pueden entender como casos paradigmáticos del uso del big data (analytics) –con títulos como "Todo lo que Netflix sabe de nosotros" o "Big data en Facebook: así se analizan los datos que publican los usuarios en la red social"–.

Un aspecto que distingue a este tópico del anterior es que se explicita e identifica a las plataformas como un agente, y se tematiza el rol activo que tienen con los datos. Una forma en que se expresa esto es en que las plataformas aparecen "colectando" datos, antes que simplemente "minándolos". Se podría conjeturar que en este contexto en el que el foco está puesto más cerca de las empresas y sus productos como entidades con intencionalidades permite desenmascarar el recorrido completo de producción del dato, haciendo irrelevante metáforas como las del descubrimiento y minado que dan a entender que los datos se encuentran "ahí afuera". Así también se menciona más seguido a los algoritmos, como técnicas ocultas o detrás de las plataformas tendientes a algún fin particular, antes que a beneficios en abstracto –por ejemplo, el artículo titulado "Algoritmo: el nuevo Gran Hermano llegó para quedarse" sostiene en su epígrafe que "Son un conjunto de instrucciones con reglas para lograr atraer a los consumidores"; o el titulado "Así funcionan los misteriosos algoritmos de Google, Facebook, Netflix y Twitter" aclara en su cuerpo que "... son como la fórmula de la Coca-Cola. Todos sabemos que existe, pero pocos la conocen. Los algoritmos operan como puente entre las acciones y los resultados de las máquinas"–. Estos aspectos se pueden observar en la figura 5.

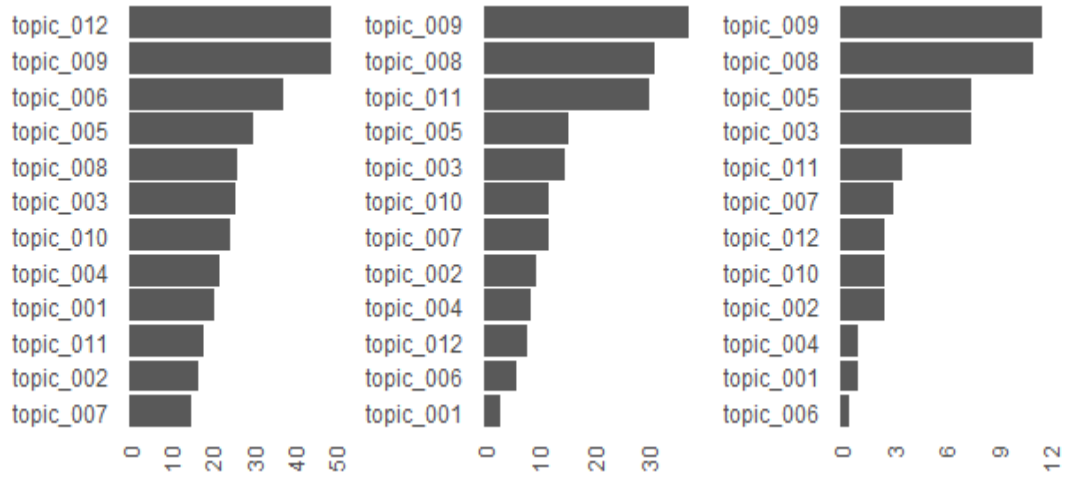

Figura 5. % de artículos que incluyen "plataforma" (der.), "algoritmo" (centro) y "colec*" (der.) en los primeros 200 artículos de cada tópico

El tercer grupo de tópicos referencian al big data en contextos comunicativos específicos: las palabras centrales del tópico #2 alcanzan al ámbito del deporte –equipo, jugador, fútbol, ganar, mundial–; las del tópico #5 presentan una mezcla entre las áreas de la salud, la agricultura y la ganadería –como tecnología, productor, campo, médico, enfermedad–; y las del tópico #3 refieren a innovaciones que el big data introduciría en distintos productos –tecnología, dispositivo, inteligencia, seguridad, aplicación, revolución–, mientras que una revisión de sus artículos más rankeados nos dan una gama que van desde teléfonos hasta autos. Estos anclajes más locales tienden a tomar el lugar de referencias más genéricas, como las "empresas" u "organizaciones" que se observaban en el primer grupo.

El cuarto contexto en el que se enmarca la discusión del big data trata con su problemática profesional. Las palabras del tópico #12 incluyen diversas aristas de la vida económica y empresarial –como cliente, negocio, servicio, compañía, o producto–, cuya tematización puede ser de interés para los ejecutivos y quienes toman decisiones. Los artículos analizados tienen por foco una pretendida "transformación digital" de las empresas, de la cual el big data sería uno de los principales factores. Generalmente, en este contexto el big data es presentado como una innovación tecnológica que ha irrumpido recientemente en las empresas, como una tendencia en marcha de los mercados, o directamente como un rasgo del futuro –por mencionar algunos títulos de los artículos: "Todos los caminos conducen al big data", "Cómo será la tienda del futuro"–. Por su parte, los términos del tópico #4 tratan con la demanda laboral y la oferta de formación de profesionales para estas tareas –trabajo, profesional, universidad, futuro, educación, carrera–. En sus artículos se trata al big data como un factor en los empleos más demandados de este año y los próximos –con títulos como "¿Cuáles son las profesiones que darán más empleo en el 2018?", o "¿Cuáles son las carreras (tecnológicas) con más salida laboral?", o "Las empresas demandan científicos de datos"–; o se lo

menciona como parte de una oferta académica, de las actividades y convenios que las impulsan, o incluso, como parte de una reflexiones acerca del futuro mercado laboral –como "IBM y científicos cordobeses desarrollan proyectos en conjunto", o "El ITBA lanza la primera carrera de analista de datos", o "Dos de cada tres niños que ingresaron en la primaria trabajarán en empleos que aún no existen"–.

El quinto y último contexto discursivo toma al big data como un fenómeno político. En el tópico #1 se incluyen palabras muy generales que refieren al Estado -ciudad, país, gobierno, provincia, Argentina, desarrollo, público, ministerio-. Si nos remitimos a sus artículos, vemos que el big data aparece como un tema de agenda política, de modo que se reporta acerca de iniciativas para su promoción, como la firma de convenios, o el armado de proyectos con otros organismos del Estado –con títulos como "Morales e Ibarra acordaron un foro de big data en Jujuy"–. Por su parte, el tópico #10 trata con temas de economía política, ya sobre mercados internacionales, o sobre política financiera o impositiva. Aquí big data aparece como un término periférico. Los artículos lo mencionan como uno de los factores –entre muchos otros– que están motivando transformaciones en las áreas mencionadas –con títulos como "Hacia un nuevo sistema monetario internacional" o "El Gobierno quiere una AFIP más enfocada en ampliar la base de contribuyentes"–. Este es el tópico donde menos menciones explícitas se registran de big data. Por su parte, en el tópico #7 el foco se corre hacia la política partidaria, donde se tratan temas como las campañas electorales y mediáticas –con palabras como política, político, gobierno, medio, campaña, elección o candidato; y títulos como "Los últimos días de la campaña" o el más explícito "El big data, la receta de Cambiemos para sacarle ventaja a Cristina Kirchner en Octubre"–. En este contexto, el big data también puede aparecer como un arma de vigilancia política y ser referido en una acusación de un partido hacia otro. Estos problemas son generalmente reportados con alguna línea editorial, como se evidencia cuando se analizan las personalidades locales/internacionales mencionadas

Definiciones del big data

Las menciones a big data colectadas en las noticias del primer gran grupo incluyen, como es de esperarse, varias definiciones explícitas. Aquí reencontramos los dos momentos ya mencionados. El primer tipo de definiciones supone un campo semántico en el que abundan las nociones relacionales con un sentido cuantificador: mejorar, tendencia, mayor, o incluso las 3 V's. Se trata de un lenguaje más abstracto y difícil de imaginar. El segundo tipo redunda en la idea de un instrumento que habilita algo a alguien. Sus formulaciones más extendidas recurren a ideas concretas que son fáciles de imaginar: el instrumento se convierte en una herramienta/ ayuda, su uso ofrece/ da/ permite decidir o tomar decisiones a una empresa/ compañia/ ciudad. En cuanto a las imágenes que se incluyen en estos artículos, predominan ilustraciones artísticas del big data, a modo de dibujos o fotos retocadas,

que buscan dar la idea de que los datos se encuentran al alcance de la mano. En algunos casos, los datos aparecen ilustrados como hologramas, haces de luces, redes de íconos y nubes de palabras y números; en otros la ilustración es más literal, si se quiere, ya que aparece el término "big data" impreso en botones y chips. En dichas imágenes los individuos que aparecen de cuerpo entero se encuentran rodeados de inmensos servidores, paredes de pantallas o interfaces en paredes de vidrio, y en estaciones de trabajo; sino el foco está puesto en sus manos y en distintas formas de manipular datos, gracias al recurso de la tecnología. En menor medida se observan algunas fotos de expertos y entrevistados.

Dado el foco más acotado del segundo grupo de artículos –sobre redes sociales y algoritmos– al big data se lo identifica menos con un objeto concreto, como puede ser una herramienta, y más con una situación tecnológica-comercial, que comparativamente resulta más difíciles de imaginar. En las ilustraciones de los artículos abundan los logos de las empresas, capturas de aplicaciones, o imágenes artísticas que buscan ilustrar procesos y funciones típicas de estas tecnologías.

Estos resultados se extienden al tercer grupo de artículos –contextos específicos–. Allí, las imágenes remiten a tecnologías materiales –chips, drones, sensores y nanorobots– e interfaces gráficas que muestran fotos de las áreas en cuestión –el campo, eventos deportivos– con alguna marca de realidad aumentada o con gráficos sobre ellos.

Como mencionamos, en el cuarto grupo de artículos –que trata con tópicos relativos al management y a la problemática laboral– prácticamente no encontramos definiciones explícitas de big data. Uno podría aquí especular que esto se puede deber a dos razones: o el big data es un fenómeno ya conocido por todos; o bien, se utiliza el término sin precisiones para generar una necesidad inespecífica. A diferencia de otros tópicos, las imágenes que ilustran a estos artículos tienen por centro a las personas, especialmente en situación de trabajo, estudio o reunión, y a logos de empresas, antes que de aplicaciones y productos.

En el quinto contexto señalado –la política–, las menciones a big data del tópico #1 lo trata como un objeto de interés y valor para la política y sus actores, con formulaciones generales y sin mayores aclaraciones; mientras que en el tópico #7 se lo define como una herramienta tendiente a analizar información personal de los usuarios de distintas redes sociales para influenciar el comportamiento electoral. Por estas razones existe un cierto solapamiento entre el campo semántico del tópico #1 y 8 (big data en foco), y del #7 y 9 (big data y redes sociales). A nivel general no hay mayores diferencias entre el nivel de abstracción del lenguaje implicado en ambos tópicos. En las imágenes que ilustran estos artículos abundan las fotos de los políticos y sus diversos eventos y reuniones, así como ilustraciones gráficas abstractas que denotan el comercio y la economía.

Discusión

En este trabajo, buscamos describir cómo los medios de comunicación de masas, especialmente las noticias en la prensa digital argentina, informan y tematizan el fenómeno del big data. Nuestros resultados permiten identificar algunos de los significados que se actualizan cuando se coloca big data en diferentes contextos discursivos. Observamos que detrás del significante big data hay significados tan amplios que permiten diferentes estrategias de ilustración y anclaje. Como señala la literatura, la ambivalencia en el término de big data es fundamental para su retórica, y también un factor fundamental para facilitar su difusión en la sociedad. Los diferentes contextos muestran que el big data se presenta por medio de características o cualidades derivadas de su magnitud, y que ofrece beneficios que son significativos solo en la cuantificación y los argumentos instrumentales: el big data es un instrumento "más grande" y "mejor" para un propósito que generalmente no se discute. Podemos hacer algunas conjeturas sobre esto: por un lado, porque dicho propósito responde a un valor supuestamente aceptado en el contexto al que se refiere la noticia (como puede ser el caso, por ejemplo, de los tópicos más “sectoriales” donde se discuten funciones sociales tales como obtener ventajas financieras, lograr un gobierno eficiente, o seguir una trayectoria profesional competitiva, etc.), o por otro lado, y más importante, porque en un discurso más masivo se pretende presentar al big data como un valor en sí mismo.

Referencias

Becerra, Gastón. 2018. “Interpelaciones Entre El Big Data y La Teoría de Los Sistemas Sociales. Propuestas Para Un Programa de Investigación.” *Hipertextos* 6(9):41–62.

Becerra, Gastón and Juan Pablo López Alurralde. 2017. “Big Data y Data Mining. Un Analisis Crítico Acerca de Su Significación Para Las Ciencias Psicosociales a Partir de Un Estudio de Caso.” *{PSOCIAL} Revista de Investigación En Psicología Social* 3(2):66–85.

Blei, David, Andrew Ng, and Michael Jordan. 2003. “Latent Dirichlet Allocation.” *Journal of Machine Learning Research* 3:993–1022.

boyd, Danah and Kate Crawford. 2012. “Critical Questions for Big Data.” *Information, Communication & Society* 15(5):662–79.

Denny, Matthew James and Arthur Spirling. 2018. “Text Preprocessing For Unsupervised Learning: Why It Matters, When It Misleads, And What To Do About It.” *Political Analysis*.

Diebold, Francis X. 2012. “The Origin(s) and Development of ‘ Big Data ’: The Phenomenon , the Term , and the Discipline.” *Penn Economics Working Paper* (12).

Glaser, Barney and Anselm Strauss. 1967. *The Discovery of Grounded Theory: Strategies for Qualitative Research*. Chicago: Aldine Publishing.

Gravano, Agustín and Matías Dell'Amerlina Ríos. 2014. *Spanish DAL: A Spanish Dictionary of Affect in Language*. Buenos Aires.

Kitchin, Rob. 2014. *The Data Revolution: Big Data, Open Data, Data Infrastructures and Their Consequences*. Los Angeles: Sage.

Kitchin, Rob. 2017. "Thinking Critically about and Researching Algorithms." *Information Communication and Society* 20(1):14–29.

Luhmann, Niklas. 1998. *Sistemas Sociales. Lineamientos Para Una Teoría General.* Barcelona: Anthropos / Universidad Iberoamericana.

Luhmann, Niklas. 2000. *The Reality of the Mass Media*. California: Stanford University Press.

Luhmann, Niklas. 2007. *La Sociedad de La Sociedad*. Mexico: Herder / Universidad Iberoamericana.

Lupton, Deborah. 2015. *Digital Sociology*. New York: Routledge.

Pentzold, Christian, Cornelia Brantner, and Lena Fölsche. 2019. "Imagining Big Data: Illustrations of 'Big Data' in US News Articles, 2010–2016." *New Media & Society* 21(1):139–67.

Portmess, Lisa and Sara Tower. 2015. "Data Barns, Ambient Intelligence and Cloud Computing: The Tacit Epistemology and Linguistic Representation of Big Data." *Ethics and Information Technology* 17(1):1–9.

Puschmann, Cornelius and Jean Burgess. 2014. "Metaphors of Big Data." *International Journal of Communication* 8:1690–1709.

Whissell, Cynthia. 2009. "Using the Revised Dictionary of Affect in Language to Quantify the Emotional Undertones of Samples of Natural Language." *Psychological Reports* 105(2):509–21.