



Representaciones sociales del big data en la prensa digital argentina

Gaston Becerra

Question/Cuestión, Nro.72, Vol.3, Agosto 2022

ISSN: 1669-6581

URL de la Revista: <https://perio.unlp.edu.ar/ojs/index.php/question/>

IICom -FPyCS -UNLP

DOI: <https://doi.org/10.24215/16696581e726>

Representaciones sociales del big data en la prensa digital argentina

Social representations of big data in the Argentinian digital press

Gaston Becerra

CONICET

Argentina

gaston.becerra@gmail.com

<https://orcid.org/0000-0001-9432-8848>

Resumen

En este trabajo indagamos cómo se construye las representaciones sociales en torno al big data en la prensa argentina, dando cuenta de los imaginarios a los que recurre, las voces que se destacan, y desentrañando los mitos y fantasías que despierta y que contribuyen a su aceptación social. Nos guiamos por la Teoría de las Representaciones Sociales de Serge Moscovici, y ponemos el foco en el proceso de construcción de la representación. Trabajamos con un diseño de corte cualitativo tendiente al análisis de contenido de un corpus de noticias (período 2010-2022) de medios digitales argentinos. Nuestros resultados incluyen la caracterización del big data como una nueva era, objetivaciones en torno a los datasets y los expertos, anclajes en el manejo del conocimiento y la información, y una actitud ambivalente.

En la discusión comparamos estos resultados con otros de corpus extranjeros, y propias desde otros marcos y diseños.

Palabras clave: Big data, representaciones sociales, prensa argentina, inteligencia artificial

Abstract

In this work we investigate how social representations are built around big data in the Argentine press, giving an account of the imaginaries to which it resorts, the voices that stand out, and unraveling the myths and fantasies that it awakens and that contribute to its social acceptance. We are guided by Serge Moscovici's Theory of Social Representations, and we focus on the representation construction process. We work with a qualitative design aimed at content analysis of a news corpus (period 2010-2022) of Argentine digital media. Our results include the characterization of big data as a new era, objectifications around datasets and experts, anchors in knowledge and information management, and an ambivalent attitude towards big data. In the discussion we compare these results with others from foreign corpus, and with our own with different designs.

Keywords: Big data, social representations, Argentine press, artificial intelligence

1. Introducción

En sus orígenes, hacia finales de la década de los '90, el término “big data” se comienza a utilizar para referir a varios desafíos computacionales y técnicos ligados al manejo de grandes volúmenes de datos (Diebold, 2012). Para fines de la década siguiente ya lo encontramos en un espacio discursivo más amplio, vinculado con el management y ofreciendo soluciones en base a conocimientos novedosos, más eficientes y precisos (Lupton, 2015). Desde aquí se ha generalizado al resto de los sistemas y espacios sociales, para ser uno de los elementos tecnológicos centrales de una transformación que promete –como señala el best seller de Mayer-Schönberger y Cukier (2013)– revolucionar la manera en que vivimos, trabajamos y pensamos.

Por su parte, desde las ciencias sociales se constituye un programa de estudios críticos en torno al big data, que busca visibilizar los procesos y las tensiones sociales de fondo, aportar a su entendimiento público y a la discusión de sus fundamentos epistemológicos, y a explorar maneras en que puede ser utilizado con fines progresistas, como potenciar las metodologías de investigación social (Kitchin & Lauriault, 2014). Buscando contribuir a este programa, en este trabajo documentamos y discutimos cómo se construye las representaciones sociales en torno al big data en la prensa argentina, dando cuenta de los imaginarios a los que recurre, las voces que se destacan, y desentrañando los mitos y fantasías que despierta y que contribuyen a su aceptación social.

Para ello, en lo que queda de esta sección introducimos los lineamientos de nuestra perspectiva teórica –la Teoría de las Representaciones Sociales–, desde donde explicitamos nuestras preguntas de investigación, y reseñamos brevemente algunos antecedentes sobre la construcción del big data en la prensa y el sentido común, incluyendo nuestros propios estudios. Luego, reportamos nuestras decisiones metodológicas, correspondientes a un diseño cualitativo con un muestreo que, si bien no es probabilístico, ha sido construido a partir de una clasificación automática de noticias de un corpus más grande, para combinar un criterio de muestreo teórico y uno de cuotas por tipo de noticias. A continuación, presentamos los

resultados de nuestro análisis, poniendo el foco en el proceso de construcción de las representaciones. En el anexo incluimos una tabla extensa con fragmentos y marcas para volver más transparente a nuestra codificación. Finalmente, concluimos el trabajo poniendo en diálogo nuestros resultados con los antecedentes y evaluamos sus alcances y limitaciones.

1.1. Enfoque teórico

La teoría de las representaciones sociales (TRS, en adelante) de Serge Moscovici recupera, en el campo de la psicología social, la idea originaria de Durkheim acerca de las representaciones colectivas, para proponer una forma de conocimiento de carácter práctico, orientado hacia la comunicación y la interacción cotidiana con nuestra realidad (Jodelet, 1985; Wagner & Hayes, 2011). Las RS se construyen activamente en la interacción y en la comunicación entre grupos sociales, y es en este marco que le dan sentido a los objetos del mundo social. Moscovici (1979, p. 18) las define como un “sistema de valores, nociones y prácticas que proporcionan a los individuos los medios para orientarse en el contexto social y material, para dominarlo”. En la actualidad y con varias décadas de desarrollo y debate, la TRS puede ser entendida como un programa de investigación supra-disciplinario, que tiene por principio unificador a una cierta perspectiva sobre la construcción social del conocimiento, y que cuenta con una gran diversidad en cuanto a los temas que ha tratado, los enfoques con los que interroga, y los métodos y técnicas con los que investiga (De Rosa, 2013).

Las RS se construyen en situaciones en las que se debe dar sentido a un fenómeno social novedoso, sobre el que abunda la información desordenada o difícil de sistematizar por parte del individuo, y que se debe amalgamar y delimitar para “llenar los vacíos” y permitir la interacción social (Castorina, 2016). Dispersión de la información, necesidad de focalizar y recortar, y presión a la inferencia son 3 las condiciones que se señalan para la emergencia de una RS. Podemos ejemplificar esto recurriendo al caso ejemplar de las RS: el estudio de Moscovici (1979) sobre la introducción del psicoanálisis en Francia. Se trata de un desarrollo científico que se inserta en el mundo del sentido común, con apropiaciones y recortes propios para cada grupo social, y condicionado por visiones ideológicas que reflejan las relaciones con los otros. No sería arriesgado hipotetizar que la emergencia del big data nos enfrente a un escenario análogo: también es un desarrollo de un espacio particular de la sociedad, que se generaliza hacia otros, promoviendo selección y recorte de ciertas ideas, la introducción de

símbolos y neologismos, y diversas adecuaciones a los posibles contextos simbólicos y de prácticas en los que se le pueden dar uso.

Para dar cuenta del proceso de construcción de la RS, Moscovici propone observar dos procesos complementarios. El primero de estos procesos es el de la objetivación, consistente en la selección y recorte de los elementos que conforman el núcleo figurativo de la RS: figuras, palabras, situaciones, símbolos y designaciones que permiten conformar una imagen del objeto representado. El segundo de estos procesos es el anclaje, consistente en la integración de la RS del objeto en los sistemas de conocimientos pre-existentes de los grupos sociales. Este proceso responde a la tesis por la cual ninguna representación se origina en un vacío de sentido, sino que se suele asimilar la novedad a experiencias pasadas. Así, al enmarcar una representación en la historia de un grupo social, se lo vincula con sus valores e ideologías, y también se lo inserta en un mundo de conflictos sociales. En términos de Moscovici, “la sociedad cambia el objeto social por un instrumento del que puede disponer, y este objeto se coloca en una escala de preferencia en las relaciones sociales existentes” (1979, p. 121). Volviendo al ejemplo de Moscovici (1979), en la objetivación del psicoanálisis, se destacan algunas ideas salientes, como la ubicuidad de lo sexual, o conceptos como el de la represión, que se desprenden de su espacio de formulación teórica y son apropiados en sus propios términos por distintos grupos sociales. En su anclaje, se vislumbra al psicoanálisis como una herramienta para hacer visible y dar claridad a lo que motiva la acción de un individuo, de modo que atañe a la explicación de la conducta de los otros, un problema socialmente muy relevante.

Interesados por la objetivación y el anclaje del big data, a la hora de analizar nuestro corpus nos guiaremos por las siguientes preguntas: por un lado, ¿cómo se hace visible?, ¿con qué idea se lo asocia?, ¿qué lo caracteriza?; y complementariamente, ¿a qué otros conocimientos y prácticas se asimila?, ¿cómo entra a jugar en un mundo de relaciones y conflictos entre grupos sociales?, ¿cómo predispone esto para la acción?

1.2. Antecedentes

Hasta donde conocemos, no se cuenta con trabajos previos que indaguen la RS del big data y que se hayan planteado las preguntas del enfoque procesual. Contamos con algunos

estudios que han realizado preguntas similares, con otros corpus de noticias –aunque no argentinas– sobre big data, desde un enfoque discursivo o semiótico. Así, por ejemplo, Paganoni (2019) trabaja con un corpus de noticias de países de habla inglesa del período 2016-2018, y se pregunta qué voces son recuperadas, qué descripciones del big data se encuentran, y qué narrativas –especialmente, qué usos y qué casos– son presentadas. En dos trabajos en colaboración, Pentzold indagó, por un lado, las imágenes e ilustraciones de los periódicos New York Times y Washington Post para el período 2010-2016 (Pentzold et al., 2019), y por otro, la construcción discursiva en periódicos alemanes, a raíz de un caso –el “Handygate”– que involucró al uso de ciertos datos personales por parte de agencias políticas (Pentzold & Fischer, 2017). También se cuenta con el trabajo de Puschmann y Burgess (2014) que, revisando noticias y publicaciones tecnológicas y económicas norteamericanas del 2012, hacen foco en algunas metáforas recurrentes con las que se introduce al big data, buscando traspasar los límites de sus espacios de comunicación. En la sección de discusiones retomamos todos estos trabajos para comparar con nuestros resultados, señalar convergencias y diferencias, y pensar generalizaciones y particularidades de las distintas representaciones reseñadas.

Luego, contamos con dos antecedentes propios que son partes del mismo programa de investigación en el que se enmarca este trabajo, que también serán puestos en diálogo con nuestros resultados hacia el final del presente trabajo. El primero de estos antecedentes sigue la línea de los antecedentes ya señalados: se trata de un análisis de contenido de más de 2000 artículos de la prensa digital argentina del período 2010-2018 realizado con técnicas estadísticas y de text-mining, mayormente modelado de tópicos (Becerra, 2021). Allí, planteamos (entre otras) las siguientes preguntas: ¿cuáles son los temas en los que se incluye el big data?, ¿cuáles son las tareas y los riesgos a los que se lo asocia?, ¿cuál es el campo semántico con el que se lo define explícitamente? El segundo de nuestros antecedentes se corre de las noticias para poner el foco en el sentido común que ellas generan, y que estudiamos a través de una encuesta (N=335) tomada a estudiantes universitarios argentinos con la técnica de evocación de palabras (Becerra & López Alurralde, 2021a). Nos interesa proponer un diálogo con este antecedente porque el mismo también se guía por la TRS, aunque lo hace con un enfoque analítico alternativo –aunque complementario– al de Moscovici. Nos referimos al enfoque “estructural” propuesto por Abric (1993), que se interroga por la

estructura y la organización de la RS, asumiendo que las mismas tienen un núcleo o sistema central con sentidos “estables, coherentes, consensuados e históricamente condicionados” (p. 76), un sistema periférico “flexible, adaptativo y relativamente heterogéneo” (p. 77). Al integrar estos antecedentes en la discusión de los resultados esperamos dar una mirada más robusta de las RS del big data.

2. Método

Para indagar las RS del big data hemos seguido un diseño de corte cualitativo tendiente al análisis de contenido de un corpus de noticias de medios digitales argentinos del período 2010-2022.

La construcción del corpus de trabajo supuso dos momentos. En un primer momento, siguiendo un directorio de medios digitales argentinos se recolectaron noticias que incluyeran el término “big data” (1). Este corpus se construyó, en un primer momento, en octubre de 2018, y se complementó en abril de 2022. El resultado es una base de datos de 2.955 noticias con texto completo y algunos metadatos. En un segundo momento, se seleccionaron algunos artículos para su lectura y análisis. Para ello se clasificó a las noticias de manera automática con la técnica de modelado de tópicos (2), que las distribuyó en 20 categorías temáticas – noticias en las que el big data es el tema focal; noticias de política, deporte, negocios, etc.– y se exportaron las 20 noticias más representativas de cada categoría. Procedimos de esta manera buscando combinar un criterio de relevancia teórico con uno de cuotas que garantice que distintos “framings” del big data estuvieran incluidos en nuestra muestra, y no sólo aquellos artículos en los que el big data es el tema central o predominante. Luego, ensayamos la inclusión de más artículos, pero como esto no resultó en códigos nuevos, y conforme a los criterios de nuestra técnica de análisis, concluimos haber alcanzado una saturación teórica de la muestra (Vasilachis de Gialdino, 2006). El resultado de este segundo momento es el corpus definitivo de 400 noticias con las que hemos trabajado.

Para el análisis, procedimos siguiendo los lineamientos de la Grounded theory (Glaser & Strauss, 1967) cuya compatibilidad con la TRS se encuentra fundamentada (Restrepo Ochoa, 2013). Específicamente, realizamos una codificación recursiva de las noticias

trabajando con la técnica de comparación constante, y agrupamos y diferenciamos estos códigos a fin de identificar los elementos centrales de una narrativa (Auerbach & Silverstein, 2003). Para asistirnos en estas tareas trabajamos con el software de análisis cualitativo atlas.ti. Luego estos códigos resultantes fueron organizados para responder las preguntas adoptadas en el marco teórico.

3. Resultados

Nuestros análisis sobre el corpus de noticias indican que, generalmente, cuando se implica al big data se pueden encontrar las siguientes ideas:

- en términos más generales, aparece como parte de un conjunto de innovaciones tecnológicas recientes cuyo progresivo alcance sobre los distintos aspectos de nuestra vida marca una nueva era;
- más específicamente, se habla de big data para referir a:
 - la disponibilidad de grandes volúmenes de datos, producto del registro y digitalización de nuestros comportamientos online, o de sensores en el mundo material, lo que llamamos “premisa”,
 - y a la “promesa” de explotar estos datos para generar nuevos conocimientos, pretendidamente más objetivos y accionables para intervenir en el mundo de forma eficiente y certera;
- que involucra a todo tipo de actores, desde Organizaciones como el Estado y las agencias políticas, o las empresas tecnológicas y plataformas digitales, y donde las principales voces que se expresan son las de expertos y consultores;
- existen múltiples ámbitos de aplicación y uso del big data, y en todos ellos aparecen desafíos y riesgos que lo tiñen con un tono y una actitud ambivalente.

En lo que sigue elaboramos sobre estas ideas para responder las preguntas sugeridas por el marco de la TRS. Para no cargar la lectura, en lugar de intercalar fragmentos de las noticias en nuestra exposición, hemos compilado algunos fragmentos en la Anexo indicando los códigos imputados.

El big data como una nueva era tecnosocial

En su tratamiento más general, el big data queda implicado en una definición extensional de un conjunto de desarrollos tecnológicos que incluyen a la inteligencia artificial en general y las aplicaciones de algoritmos y con nuevos dispositivos que permiten el registro y la circulación de datos. Estos desarrollos suelen ser parte de un relato que incluye valores sociales tales como la innovación y la democratización, valores epistémico como la objetividad y la transparencia, o en los que se destaca su capacidad para actuar e intervenir sobre la realidad. El escenario que se describe es el de una ruptura o una revolución con un profundo impacto, y con alcance de transformar todos los aspectos de nuestra vida social, subjetiva y material. Por esto se habla de una tendencia irreversible o una “era del big data” [código “Tendencia” en Anexo], designación que se suma a una lista de etiquetas sociológicas y mediáticas que caracterizan al todo social a partir de generalizar algún desarrollo, tal como sucedió con “la era de la información” (Luhmann, 2007). Esto marca el tono del mensaje en torno al big data que, al presentarlo como un fenómeno global que nos determina, ejerce una presión hacia la inferencia y el desarrollo de una RS.

La objetivación del big data en torno al dataset y sus propiedades

Más específicamente, cuando se define al big data se suele indicar que, gracias al registro de nuestros comportamientos online y al uso de sensores en el mundo material, se han construido grandes datasets, que se encuentran disponibles para ser analizados. El proceso que se implica en esta idea es el de una “datificación” por la cual nuestro mundo social, subjetivo y material son transformados de manera activa en datos por parte de algún agente con cierta capacidad de registro y procesamiento [código “Datificación” en Anexo]. Si bien este proceso se puede rastrear hasta los albores de la modernidad, obtuvo un enorme impulso en los últimos años a partir de la masificación de dispositivos conectados a internet y la participación en plataformas digitales. Esta datificación no suele ser explicitada como un proceso activo y dirigido, sino que más bien se lo da por supuesto, refiriendo a su resultado –la (supuesta) disponibilidad de datasets– o a la proliferación de medios de registros y sensores sin explicitar un fin. La planificación y el diseño de las tareas de registro de la datificación sólo suelen hacerse visibles en campos que muy recientemente se están “datificando”, entre cuyos

ejemplos se mencionan al fútbol o la producción agrícola. Esta poca visibilización y tematización de la datificación tiene dos consecuencias importantes: en primer lugar, se esconde su naturaleza económica y social [código “Económico” en Anexo], que puede incluir la extracción –en ocasiones, coaccionada o forzada– de datos y su transformación en un commodity, o su acumulación, reproducción y circulación como elemento central de un capitalismo de datos (Sadowski, 2019); en segundo lugar, se esconde que detrás del proceso hay poderosos agentes sociales que tienen la propiedad de los datos, generando una nueva forma de desigualdad o “brecha digital”, que pondría en entredicho la prometida disponibilidad y el libre uso (boyd & Crawford, 2012; McCarthy, 2016).

Dada por supuesto la disponibilidad de datos, el foco se posa sobre las características ontológicas y las propiedades mayormente físicas de las bases de datos (Kitchin & McArdle, 2016) [código “Ontología de los datos” en Anexo]. Estas son un elemento central de la narrativa del big data, ya que son las que supuestamente habilitan la capacidad de generar nuevos “insights”. Al respecto, conviene tener en cuenta las advertencias de Leonelli (2020): “el énfasis en las características físicas de los datos oscurece la dependencia de la interpretación y las circunstancias de uso de los datos, que incluyen consultas específicas, valores, habilidades y situaciones de investigación.” Las propiedades de los datasets más mencionadas son aquellas que han sido popularizadas a través de la regla mnemotécnica de las 3 Vs: volumen, variedad, velocidad. Originalmente estas se referían a desafíos de hardware y software implicados en el manejo de grandes datasets. Posteriormente, a esta lista se han ido añadiendo nuevos términos que dan cuenta del avance en los intereses, como valor o veracidad. De todas estas propiedades, volumen es la más prominente, en tanto proporciona lo “big” del big data; pero esto sólo redundaría en abstracciones, ya que tanto “big” como “data” son nociones que suelen adquirir sentido sólo relacional y posicionalmente. De esta manera, la designación misma parece no tener límites intrínsecos, ni estar confinada a un cierto ámbito de prácticas. Y a la vez, sin perder esta generalidad, hay varios recursos retóricos que vuelven al big data un fenómeno que podemos representar como algo inconmensurable y poderoso, como, por ejemplo, la utilización de cifras desorbitantes y unidades de medidas tales como exabytes, o las comparaciones que señalan lo desproporcionado de las capacidades técnicas frente a las de un individuo.

La objetivación a través de lo visible: personajes, infraestructura y actores

La literatura especializada en la retórica del big data y su construcción discursiva mediática suele coincidir en señalar un interesante candidato para la objetivación, al referir a las representaciones de grandes volúmenes de datos con metáforas naturales, tales como la nube, o un “mar de datos”. Este tipo de imágenes se destacan por invisibilizar el carácter social del big data en sus primeros años de masificación. En nuestro corpus este tipo de metáforas, si bien está presente, es menos prominente. En su lugar abundan dos imágenes de la datificación menos metafóricas: los personajes que le dan voz al big data –expertos, consultores y prestadores de servicios– y los equipos, servidores y sensores que configuran su sustrato material; los grandes actores Organizaciones de la vida social -los estados y sus agencias, las empresas, etc.- en donde se registra el impacto del big data.

Sobre los primeros: la mayor parte de las definiciones del big data revisadas provienen de entrevistas y textuales de expertos que, además de introducirnos a la “era del big data”, o describir las propiedades de enormes datasets, también promocionan su aplicación en diversas áreas de la vida social [código “Expertos” en Anexo]. En muchos casos, este encuadre vuelve al big data un fenómeno comercial; y no es infrecuente que se hable de big data para referir a un nombre de producto, línea de servicios, área de la empresa, o a una designación o cargo detentado por un experto [código “Producto comercial” en Anexo]. Una línea narrativa similar se encuentra en un personaje particular: los analistas o científicos de datos, las “estrellas del mercado laboral” que tienen la tarea de cumplir la promesa del big data, es decir, transformar el volumen en conocimiento que permita accionar y generar valor. La datificación también se vuelve visible cuando se describe la proliferación de sensores y cámaras, la disponibilidad de dispositivos, o se mencionan desarrollos informáticos, generalmente referidos como “aplicaciones de inteligencia artificial”. En estos casos lo que se describe es un mundo social donde los individuos conviven con nuevos agentes en la comunicación (Esposito, 2022).

Sobre los segundos: los otros sujetos del big data son actores institucionales y organizaciones [código “Organizaciones” en Anexo] de diversas áreas de aplicación, tales como negocios y gestión, política, ciencia, deporte, medios, etc. [código “Aplicación” en Anexo]. Esto abre el imaginario a casos y ejemplos de organizaciones que estarían liderando la transformación digital: se reporta sobre equipos de fútbol ganadores de mundiales y torneos;

científicos que trabajan en proyecto de salud o gobierno; agencias políticas, tales como la policía o los organismos de control fiscal que buscan predecir y prevenir el crimen o identificar evasores. En un tono más general, se destacan dos grandes personajes, centrales en la narrativa de cualquier desarrollo tecnológico reciente: por un lado, las grandes empresas del capitalismo de plataformas (Srniczek, 2018) –Amazon, Google, Microsoft, Apple, Facebook, y otras redes–; por otro, los estados nacionales y sus gobiernos, como cuando se reporta sobre el big data como problema de agenda política local, o cuando se menciona a las grandes potencias mundiales y se reporta al big data como un problema geopolítico. Estos dos actores aparecen profundamente vinculados en el icónico caso de Cambridge Analytica y la explotación de datos de Facebook para campañas políticas como la de Trump o el Brexit.

Finalmente, se debe notar que en esta objetivación del big data, que pone el foco en los expertos, los actores institucionales y las organizaciones, y el entramado de sensores, dispositivos y aplicaciones, se pierde de vista el sujeto más importante de la datificación: los individuos y grupos sociales cuyas acciones son convertidas en datos, en objeto del big data. Este personaje central sólo aparece en las notas críticas o editoriales que reclaman el resguardo de la información personal, o que cuestionan los límites éticos de la datificación de la economía y la política.

Anclaje en el manejo del conocimiento y la información

A pesar de la novedad del término –que se populariza recién cerca del 2010– la representación del big data recalca en un fenómeno socialmente preexistente que le otorga sentido: la gestión basada en información y conocimiento. Este objetivo es simbiótico con el de la datificación, y también se puede rastrear hasta los albores de la modernidad. Pero ahora, bajo el título de big data, este proyecto pretende finalmente plausibilidad, gracias a un nuevo tipo de conocimiento, marcado por la capacidad predictiva, la posibilidad de una transferencia certera, completitud y hasta objetividad. En este sentido, entendemos que las promesas que se hacen en nombre del big data se fundamentan en supuestos y expectativas de carácter epistémico que involucran lo que aceptamos como conocimiento válido y confiable, quiénes y dónde lo producen, y con qué fines; a la vez, en tanto esto implica una redefinición de las

relaciones entre grupos sociales, que incluye que alguno de ellos pueda accionar sobre otros, esta promesa adquiere un carácter claramente ideológico.

Estos reclamos que combinan lo epistémico e ideológico han sido descritos por van Dijck (2014, p. 198) bajo el concepto de “dataísmo”: “una creencia generalizada en la cuantificación objetiva y el seguimiento potencial de todo tipo de comportamiento humano y sociabilidad a través de tecnologías de medios en línea. El dataísmo también implica la confianza en los agentes (institucionales) que recopilan, interpretan y comparten (meta)datos...” [código “Dataísmo” en Anexo]. Así, se sugiere que el volumen y la variedad de los datos recabados ofrecen una cierta completitud que es menos propenso a errores o supera la incertidumbre de las muestras; que el conocimiento es más complejo porque incluye miles de variables y dimensiones; que utilizar registros automáticos de los comportamientos ofrecen un dato más natural y menos intrusivo que el generado por medio de encuestas y entrevistas; que el cruce de todas las variables disponibles permite generar conocimiento sin la necesidad de hacer preguntas o generar hipótesis, sino que permite a los datos hablar por sí mismos; que esta falta de la guía del investigador lo resguarda de subjetividades, sesgos o manipulaciones, y que por lo contrario, “los datos no se equivocan”; que la granularidad y velocidad de los datos permite captar mejor tendencias, que la disponibilidad de casos antecedentes permite generalizar, y, en definitiva, que se puede predecir el futuro, o hasta direccionarlo. En estos reclamos se implica una supuesta superación del conocimiento científico –máxime del conocimiento de las ciencias sociales–, algo que, en ocasiones se sostiene explícitamente, y se reporta sobre una supuesta revolución científica.

Es en vistas de la valoración de estas promesas que los expertos ganan la credibilidad suficiente como para sostener reclamos tan contraintuitivos que, como señalan algunos artículos, parecen cruzar el límite entre lo real y la ciencia ficción. La manera en que se caracteriza a los expertos y a los actores del big data, también supone una nueva división del trabajo social en torno a la producción y circulación de conocimientos, la cual ya no se limita a los espacios científicos, sino que se desarrolla en el cruce entre empresas, gobiernos y agencias científicas, y presenta problemas éticos y epistemológicos novedosos.

El tono ambivalente del big data

En los distintos espacios a los que alcanza el big data se puede observar alguna “promesa” que especifica el sentido de un conocimiento nuevo con relación a algún objetivo social. Esto ha sido señalado en la literatura por medio de estudios sectoriales y comparativos, como, por ejemplo, cuando Kitchin (2014) reconstruye la narrativa del big data en torno a la política, el comercio y las organizaciones, para finalmente reencontrar los mismos valores: eficiencia, transparencia, optimización. Estos valores generales, junto a promesas concretas, como “mejorar la calidad de vida”, se pueden observar también en nuestro corpus [código “Promesas” en Anexo]. No hay duda que la información y la tecnología son factores centrales para mejorar la forma en que se desarrollan tareas específicas; pero cuando el objetivo que se persigue remite a un problema complejo, se corre el riesgo de estar frente a lo que Mozorov (2015) denomina un discurso “solucionista” que no hace más que simplificar el problema y reducirlo a variables computables: “El solucionismo da por sentado el problema que intenta resolver, en lugar de investigarlo, y así llega a la respuesta antes de haber formulado las preguntas en todas su amplitud” (Mozorov, 2015, p. 24). Incluso si nos remitimos al anclaje señalado –la gestión a través de la información y el conocimiento– y a los principales personajes y características identificados en su objetivación –expertos y organizaciones y su enorme volumen–, encontramos rápidamente antecedentes que advierten acerca de los posibles prácticas riesgosas o perniciosas: se habla así de manipulación genética o manipulación política, y de “Big brother” o “big others” (Zuboff, 2015). En nuestro corpus, esta visión se refleja mayormente en artículos focales, editoriales y de opinión, donde los expertos suelen ser investigadores y académicos [código “Críticas” en Anexo].

Este tono negativo tiñe a varias de las promesas específicas del big data, volviéndolas ambivalentes. Así, por ejemplo, el ideal de la democratización y la transparencia de los asuntos públicos queda en entredicho por las acusaciones de manipulación en campañas políticas recientes, como las de Cambridge Analytica; o cuando “usar de big data” se vuelve una acusación entre partidos políticos; o cuando la vertiginosa innovación de las empresas del sector aparece como un campo de prácticas sin regulación ni resguardo de los datos personales. Además de estos riesgos, el big data también supone algunos desafíos que condicionan o matizan la factibilidad de sus promesas. Por caso, se advierte la necesidad de un profundo cambio cultural en las organizaciones como para hacer valer los datos, o se señala la falta de capital humano calificado para su aplicación. Los fundamentos epistémicos del

dataísmo, por su parte, no se encuentran muy cuestionados; a lo sumo se constata que la revolución de los grandes datos no puede invalidar prácticas tradicionales, o de “small data” en ciertos casos, o que de ninguna manera muchos datos son todos los datos posibles.

4. Discusiones y conclusiones

Las simplificaciones y vaguedades con las que se presenta, que incluyen su mención dentro de un conjunto de tecnologías que estarían marcando una nueva era o tendencia para remarcar su relevancia e importancia, son esperables de una comunicación que, en términos de Moscovici (Moscovici, 1979), busca difundir una representación para lograr la masificación y la formación de una opinión mayoritaria. En las noticias reseñadas el big data se presenta a través de la idea de enormes acervos de datos que permiten explotar información y conocimientos para diferentes tipos de gestiones de una manera novedosa y disruptiva. Los personajes que le dan voz al big data suelen ser expertos que representan o visibilizan a diversas organizaciones de nuestra vida social, entre las que destacan las grandes empresas y los gobiernos. El tono predominante es mayormente positivo, o incluso promocional, ya que el anclaje en la información y el conocimiento para una gestión precisa y eficiente lo asocia a un discurso solucionista, donde el big data traería beneficios importantes. Sin embargo, también se observa un discurso crítico, en el que el big data es asociado con riesgos y prácticas de dominación. Para un público general, alcanzado por ambos mensajes, el big data queda caracterizado como un fenómeno cuya valoración es ambigua, provocando actitudes disímiles.

Hay concordancia con algunos elementos narrativos señalados en nuestros resultados con los de otras investigaciones que han trabajado con corpus de otros países. La preeminencia de las ideas de dato, volumen e información, el foco en los expertos en tecnologías y en grandes actores, y hasta su carácter ambiguo han sido destacados también por Paganoni (2019) sobre un corpus de noticias de USA y Reino Unido, y por Pentzold y Fischer (2017) en uno de Alemania. Lo más fácilmente visible de estos elementos -los contextos de aplicación y los expertos- también han sido destacados en el trabajo de las imágenes que acompañan a las noticias en un corpus de noticias de Estados Unidos (Pentzold et al., 2019). Estas coincidencias sugieren que la representación sigue un discurso que trasciende el plano de lo local. Nuestro trabajo también coincide en señalar –junto con aquellos

antecedentes interesados en un análisis crítico (Paganoni, 2019; Portmess & Tower, 2015)– que las noticias tienden a invisibilizar al proceso de datificación por el cual el big data reporta a un modelo económico particular, con relaciones de explotación y nuevas desigualdades. Y si bien una narrativa “negativa” del big data se ha ido imponiendo desde el caso de Cambridge Analytica en 2018 y en torno a distintos procesos electorales locales e internacionales desde entonces, esta narrativa parece limitarse a casos extremos y no tematizar los procesos sociales de fondo, o discutir los escasos argumentos epistemológicos que darían soporte a los reclamos de un conocimiento nuevo y revolucionario. Paganoni, en relación con un corpus de noticias en inglés, señala: “A pesar de los escándalos recientes y las áreas grises que aún no se han aclarado, el debate tecnológico actual resuena con la promesa de que las ventajas del big data finalmente superarán sus inconvenientes, ante todo en el ámbito vital de la atención médica” (Paganoni, 2019, p. 51).

También debemos señalar algunas concordancias con lo observado en nuestros trabajos anteriores. Uno de ellos corresponde a un análisis textual, realizado con técnicas de text-mining, de un corpus coincidente de más de 2000 de noticias argentinas hasta el 2018 (Becerra, 2021), donde también aparecen algunos de los elementos narrativos antes destacados, como los actores señalados, la insistencia en las promesas, o el tono ambivalente; y donde también coincidimos en destacar las invisibilizaciones señaladas. El otro trabajo se guió por la TRS, aunque con un enfoque alternativo para modelar la representación, ya que nos interesó por describir su estructura, para lo que tomamos una encuesta con la técnica de evocación de palabras (Becerra & López Alurralde, 2021a). Allí señalamos un núcleo figurativo para el sentido común del big data expresado que se expresa en términos como información, dato, control, internet, análisis, grande / volumen, manipulación, bases de datos, y conocimiento (en orden de frecuencia decreciente), que redundan mayormente en la narrativa ambigua señalada. Esta coincidencia en el caso de los dos trabajos guiados por la TRS es particularmente interesante ya que configura una mirada que conecta los dos extremos de los agentes de la comunicación noticias/lector.

Como todo trabajo empírico de corte cualitativo, este estudio tiene sus limitaciones y sus desafíos. La Anexo incluye fragmentos de las noticias y el código que hemos anotado en el análisis, hemos expuesto una narrativa general que ordena las ideas, para luego realizar un análisis guiado por las preguntas explicitadas. En este proceder buscamos justificar nuestra

interpretación de los datos de una manera que resulte transparente y a la vez explicita los puntos ciegos de nuestro análisis, abriéndose a otras interpretaciones alternativas. Luego, la convergencia de resultados con otros estudios de corpus distintos, con enfoques, técnicas y métodos alternativos, le da apoyo a nuestras inferencias y nos permite conjeturar una posible generalización de los resultados aquí señalados, objetivo al que también apuntamos al construir nuestro corpus con una muestra de documentos de un corpus mucho mayor considerando cuotas de tipos de noticias.

Notas

(1) Una presentación más completa del procedimiento de recolección de las noticias y de su preprocesamiento para el topic modeling se puede encontrar en la sección metodológica de un trabajo previo (Becerra, 2021). De hecho, nuestra base de datos (completa) de artículos es una actualización a abril del 2022 de la base construida en octubre de 2018 para ese artículo

(2) Una presentación más completa de la técnica y sus fundamentos, y una breve discusión acerca de su vinculación con la investigación cualitativa, se puede encontrar en otro trabajo (Becerra & López Alurralde, 2021b)

(3) En estas reflexiones seguimos la propuesta de Auerbach y Silverstein (2003) acerca de los distintos criterios de calidad para las investigaciones cuantitativas y cualitativas, que incluye a la transparencia como alternativa para la validez y la confianza, y a la transferencia de casos como alternativa para la generalización estadística.

REFERENCIAS

- Abric, J.-C. (1993). Central System, Peripheral System: Their functions and roles in the dynamics of social representations. *Papers on Social Representations*, 2(2), 75–78.
- Auerbach, C., & Silverstein, L. B. (2003). *Qualitative data: an introduction to coding and analysis*. New York University Press.
- boyd, D., & Crawford, K. (2012). Critical Questions for Big Data. *Information, Communication & Society*, 15(5), 662–679.

- Castorina, J. (2016). La significación de la teoría de las representaciones sociales para la psicología. *Perspectivas En Psicología*, 13(1), 1–10.
<https://dialnet.unirioja.es/servlet/articulo?codigo=5523376>
- De Rosa, A. S. (2013). *Social Representations in the ' Social Arena .'* Routledge.
- Diebold, F. X. (2012). The Origin(s) and development of “ Big Data ”: the phenomenon , the term , and the discipline. *Penn Economics Working Paper*, 12.
<https://doi.org/10.2139/ssrn.2202843>
- Esposito, E. (2022). *Artificial communication. How algorithms produce social intelligence*. The MIT Press.
- Glaser, B., & Strauss, A. (1967). *The Discovery of Grounded Theory: Strategies for Qualitative Research*. Aldine Publishing.
- Jodelet, D. (1985). La representación social: fenómenos, conceptos y teoría. In S. Moscovici (Ed.), *Psicología Social II* (pp. 17–40). Paidós.
- Kitchin, R. (2014). *The Data Revolution: Big Data, Open Data, Data Infrastructures and Their Consequences*. Sage. <https://doi.org/10.4135/9781473909472>
- Kitchin, R., & Lauriault, T. P. (2014). Towards critical data studies : Charting and unpacking data assemblages and their work. *Geoweb and Big Data*, 1–19.
http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2474112
- Kitchin, R., & McArdle, G. (2016). What makes Big Data, Big Data? Exploring the ontological characteristics of 26 datasets. *Big Data and Society*, 3(1), 1–10.
<https://doi.org/10.1177/2053951716631130>
- Leonelli, S. (2020). Scientific Research and Big Data. *Stanford Encyclopedia of Philosophy*.
- Luhmann, N. (2007). *La sociedad de la sociedad*. Herder / Universidad Iberoamericana.
- Lupton, D. (2015). *Digital Sociology*. Routledge. <https://doi.org/10.1057/9781137297792>
- Mayer-Schonberger, V., & Cukier, K. (2013). *Big data. A revolution that will transform how we live, work, and think*. Eamon Dolan/Houghton Mifflin Harcourt.
- McCarthy, M. T. (2016). The big data divide and its consequences. *Sociology Compass*, 10(12), 1131–1140. <https://doi.org/10.1111/soc4.12436>
- Moscovici, S. (1979). *El psicoanálisis, su imagen y su público*. Huemul.
- Mozorov, E. (2015). *La locura del solucionismo tecnológico*. Clave Intelectual.
- Paganoni, M. C. (2019). *Framing big data : a linguistic and discursive approach*. palgrave

macmillan.

- Pentzold, C., Brantner, C., & Fölsche, L. (2019). Imagining big data: Illustrations of “big data” in US news articles, 2010–2016. *New Media & Society*, 21(1), 139–167.
<https://doi.org/10.1177/1461444818791326>
- Pentzold, C., & Fischer, C. (2017). Framing Big Data: The discursive construction of a radio cell query in Germany. *Big Data & Society*, 4(2). <https://doi.org/10.1177/2053951717745897>
- Portmess, L., & Tower, S. (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. <https://doi.org/10.1007/s10676-014-9357-2>
- Puschmann, C., & Burgess, J. (2014). Metaphors of Big Data. *International Journal of Communication*, 8, 1690–1709.
- Restrepo Ochoa, D. (2013). La Teoría Fundamentada como metodología para la integración del análisis procesual y estructural en la investigación de las Representaciones Sociales. *Revista CES Psicología*, 6(1), 122–133. <https://doi.org/10.21615/2579>
- Sadowski, J. (2019). When data is capital: Datafication, accumulation, and extraction. *Big Data and Society*, 6(1), 1–12. <https://doi.org/10.1177/2053951718820549>
- Srnicek, N. (2018). *Capitalismo de plataformas*. Caja Negra.
- van Dijck, J. (2014). Datafication, dataism and dataveillance: Big data between scientific paradigm and ideology. *Surveillance and Society*, 12(2), 197–208.
- Vasilachis de Gialdino, I. (2006). *Estrategias de investigación cualitativa*. Gedisa.
- Wagner, W., & Hayes, N. (2011). *El discurso de lo cotidiano y el sentido común. La teoría de las representaciones sociales*. Anthropos.
- Zuboff, S. (2015). Big other: Surveillance capitalism and the prospects of an information civilization. *Journal of Information Technology*, 30(1), 75–89.
<https://doi.org/10.1057/jit.2015.5>