



FILO:UBA
Facultad de Filosofía y Letras
Universidad de Buenos Aires



Las Humanidades Digitales desde Argentina

Tecnologías, Culturas, Saberes



Buenos Aires, 2014

Actas de las I Jornadas de Humanidades Digitales



FILO:UBA
Facultad de Filosofía y Letras

FILODIGITAL
Repositorio Institucional de la Facultad
de Filosofía y Letras, UBA

Ejes para un debate sobre el uso ético de datos interaccionales escritos y orales obtenidos en línea

DE-MATTEIS, Lorena M. A. / CONICET. Universidad Nacional del Sur -lmatteis@uns.edu.ar

» *Palabras clave: análisis del discurso institucional, Sociolingüística, datos disponibles en línea, metodología, ética.*

> **Resumen**

La riqueza de los datos disponibles hace de internet una herramienta y un recurso atractivo para los estudios sociales. Entre estos datos, los escritos son los más empleados en investigaciones de naturaleza lingüística: blogs, foros y grupos de discusión, comentarios de lectores, *Twitter*, *Facebook*, etc., constituyen los entornos virtuales más utilizados para analizar prácticas discursivas. Los aspectos éticos del acceso y el uso de estos datos han sido considerados, con mayor o menor profundidad, por autores como Herring (1996, 1997), Buchanan (2004), Crystal (2005) o Markham (2011). Pero internet también permite recoger datos de naturaleza oral, ya sea con el acuerdo explícito de los sujetos o sin él. En el primer caso, puede considerarse, por ejemplo, el uso de *Skype* y sistemas similares para realizar entrevistas y encuestas a distancia. En el segundo, podemos pensar en el aprovechamiento de las distintas opciones disponibles de *live-streaming*, que resultan de gran utilidad, por ejemplo, para el estudio de la radiotelefonía aeronáutica. El propósito de este breve trabajo, entonces, es traer a consideración algunos aspectos éticos en las decisiones metodológicas que deberían tomarse a la hora de hacer uso de estas opciones de accesibilidad a datos que aparecen como públicos, pero que no siempre se piensan como tales. Con este fin, sistematizaremos nuestras consideraciones a partir de los factores que definen las esferas de investigación en contextos mediados por Internet según los definen Maczewski *et al.* (2004), y organizaremos la discusión sobre la base de los siguientes ejes: origen de los datos, carácter público/privado, identificación de los hablantes.

> **Introducción**

En el marco del *giro computacional (computational turn)* (Berry, 2012) de las Ciencias

Sociales y Humanas, internet constituye para estas disciplinas una herramienta que promueve interrogantes metodológicos nuevos (y modifica otros) y, al mismo tiempo, una rica fuente de datos. Estos datos se caracterizan por su accesibilidad en todo momento y por una diversidad que viene dada por su naturaleza (pueden ser gráficos, escritos y también audiovisuales), su origen sociocultural, su perduración temporal, las temáticas que reflejan, etc. Entre los muchos aspectos metodológicos que Internet viene a poner en cuestión, entonces, las problemáticas éticas que supone su empleo constituyen una de las áreas de mayor interés. En este trabajo, presentamos una reflexión de tipo general e introductoria, de orientación aplicada, acerca del empleo cualitativo de estos datos en investigaciones lingüísticas centradas en los usuarios, proponiendo una serie de criterios operativos elementales que ejemplificaremos, en parte, a partir de reflexiones sobre eventuales usos de datos en línea para el ámbito institucional aeronáutico.

Los datos que provienen del correo electrónico, blogs, foros y comentarios de lectores, entradas en redes sociales -*Facebook* y *Twitter* sobre todo-, son los más utilizados para analizar prácticas discursivas. Con diverso nivel de profundidad, autores como Herring (1996, 1997), Buchanan (2004), Crystal (2005) o Markham (2011) han señalado líneas orientadoras para el aprovechamiento ético de datos escritos producidos en línea en los estudios lingüísticos, tanto en lo que hace al acceso como a su reproducción, pero también pueden encontrarse formulaciones y pautas éticas propuestas desde otras ciencias (Rivers & Lewis, 2014) y de equipos como el que preparó el denominado *Menlo Report* (Bailey, Dittrich, Kenneally & Maughan, 2012) que permiten identificar coincidencias y diferencias de criterio, ampliando la perspectiva disciplinar.

Ahora bien, internet también permite recoger datos de naturaleza oral (tanto en audio como en video), ya sea con el acuerdo explícito de los sujetos o sin él. En el primer caso, puede considerarse, por ejemplo, el uso de sistemas similares a *Skype* para realizar entrevistas y encuestas a distancia. En el segundo, podemos pensar en el aprovechamiento de canales de *live-streaming* de audio y video así como de los archivos disponibles en plataformas para compartir este tipo de archivos.

Estas observaciones preliminares permiten realizar una distinción básica y significativa entre, por un lado, los datos disponibles en línea, en cuyo origen no interviene el investigador, y, por el otro, aquellos elicitados a partir de las diversas alternativas que ofrece internet. Esta distinción supera la dimensión de la naturaleza física o, si se quiere, también semiótica del dato, sino que permite incorporar la problemática metodológica vinculada con su obtención, más pertinente para la reflexión ética que propone este trabajo.

Nuestro propósito, entonces, es considerar algunos de los aspectos éticos en las decisiones metodológicas que deben tomarse a la hora de obtener y hacer uso de datos, sobre todo de los disponibles en línea, que aparecen como públicos pero que *no siempre se piensan* como tales. Para ellos se retoman algunas de las propuestas que otros investigadores han formulado previamente y se intenta, al mismo tiempo, sistematizar algunos interrogantes básicos en ejes centrales de reflexión.

› **Consideraciones éticas en la investigación sociolingüística/discursiva de base etnográfica**

Tanto en la perspectiva teórico-metodológica de la sociolingüística interaccional como en la del análisis del discurso se priorizan los datos de interacciones naturales. Las técnicas etnográficas vinculan, por tanto, al investigador con los grupos sociales objeto de su investigación pues se desarrollan en el contexto interaccional habitual de dichos colectivos. La relación que se establece supone una dimensión ética que no desaparece al trasladar estas técnicas observacionales al mundo virtual.

De manera general, la reflexión sistemática sobre tal dimensión es más reciente en lingüística que en otras ciencias sociales (Newman, 2009), pero estudios ya clásicos en sociolingüística y dialectología urbanas (Labov, 1982; Wolfram, 1993), en documentación de lenguas (Holton, 2009; Newman, 2009; Rice, 2004, etc.) y también en la investigación de entornos institucionales específicos (Hamilton, 1993; Finegan, 1993; de-Matteis, 2013) dan cuenta de esta preocupación y formulan principios generales como los de *corrección del error, deuda incurrida y gratificación lingüística* de Labov y Wolfram, que se suman a los más generales de *autonomía, no-maleficencia, beneficencia y justicia* (Wiles, Heath, Crow & Charles, 2005: 7).

Aunque, como en otras Ciencias Sociales, resulta difícil definir una única deontología o ética normativa (Rice, 2004: 150) en los estudios lingüísticos, existe consenso en que las posiciones y principios éticos más generales deben ser revisados en forma particular por cada investigador frente a la complejidad de su objeto de estudio y la comunidad que investiga. Esta reflexión, además, suele ocurrir dentro del marco de tres grandes paradigmas (Rice, 2004): la *investigación ética*, la *defensa de la investigación (advocacy research)* -en la que se insertan los principios labovianos- y la *investigación para el empoderamiento* de las comunidades estudiadas (*empowering research*), etapas cada vez más colaborativas en la relación investigador/informante (o consultante)/comunidad.

› **La reflexión ética en la investigación etnográfica de contextos virtuales**

Antecedentes

En el marco general de reflexión al que aludimos en la sección precedente, y adscribiendo - en nuestro caso- al paradigma ético y colaborativo de la *empowering research*, resulta conveniente plantearse algunos interrogantes básicos a la hora de proponer el uso explícito de datos producidos por hablantes individuales y que provienen de contextos comunicativos virtuales o que, con independencia del contexto interaccional en el que se originan, se difunden mediante medios informáticos.

Estos interrogantes pueden organizarse teniendo en cuenta las esferas de investigación

identificadas por Maczewski, Storey & Hoskins (2004: 64) para los contextos comunicativos *mediados por internet* (Imagen 1), que combinan los aspectos sociales de toda interacción y proceso investigativo con los técnicos.

El análisis de este esquema alerta sobre dos aspectos. En primer lugar, si los investigadores, participantes, y contextos son factores a considerar tanto en la esfera *online* como en la que denominan *onground* (el mundo material o físico), las relaciones que establecen son diversas. Así, por ejemplo, los datos en línea pueden corresponder a interacciones en las que los participantes:

- a) Interactúan asincrónicamente en el mundo virtual en contextos también virtuales (*Facebook*, *Twitter*, correo electrónico, foros, *blogs*, etc.).
- b) Interactúan (cuasi) sincrónicamente mediante sistemas basados en la transmisión de datos por internet como si se tratara de cualquier otro sistema de comunicación a distancia, tanto oral como escrito (*Skype*, otros programas para chatear).
- c) Interactúan sincrónicamente en el mundo real y estas interacciones son reproducidas en tiempo real a través de la Internet (*live-streaming* de audio o audio y video).



Imagen 1. Esferas de la investigación en contextos mediados por internet. Fuente: Maczewsky, Storey & Hoskins (2004: 64).

Hasta donde llega nuestro conocimiento, el escenario *c* no ha sido abordado por la bibliografía pues queda fuera del ámbito de las CMC (*Computer Mediated Communications*)

tradicionales, pero para los propósitos de este trabajo es importante, pues en ocasiones también estos datos disponibles en línea resultan de interés para el investigador lingüista, por ejemplo, al investigar la radiotelefonía aeronáutica.

Por otro lado, la intersección que concierne a la ética investigativa en este diagrama incluye elementos de una esfera técnica, es decir, las posibilidades informáticas concretas y los saberes de los sujetos referidos al *hardware*, al *software* y a la infraestructura que hacen posibles tanto los contextos comunicativos virtuales como, según hemos observado ya, la obtención de datos elicitados y disponibles en línea.

Si revisamos la bibliografía existente, los postulados éticos más frecuentes coinciden con los aplicables en el trabajo de campo en comunidades no virtuales pero tienen en cuenta la particularidad de la intersección entre las esferas *online* y *técnica*, por ejemplo, observando que, en la mayoría de los casos las intervenciones de los hablantes quedan registradas *por ellos mismos* en forma escrita (Herring, 1996 y 1997; Buchanan, 2004; Crystal, 2005; Markham, 2011, entre otros). Proponen, entonces, que la cita de tales registros debe realizarse teniendo en cuenta la percepción de cada comunidad virtual sobre lo público y lo privado así como las posibilidades técnicas de rastrear en la esfera *onground* al individuo al que corresponde un texto determinado.

Por otra parte, si bien se ha señalado que -a excepción de los enfoques críticos de análisis del discurso-, en general, el análisis lingüístico se concentra más sobre las formas que sobre los contenidos de las intervenciones (Herring, 1996), por lo que no se dañaría a los sujetos al citarlas, es precisamente cuando los intereses de investigación no son formales e interesa conocer y considerar el origen sociocultural o sociolingüístico de quienes han producido esos enunciados seleccionados como datos cuando la discusión ética sobre el uso de materiales en línea adquiere su mayor relevancia. Se trata de investigaciones cualitativas que pueden tener intereses variados como, por ejemplo, estudiar diferentes representaciones sociales, ideologías, cuestiones de poder, o, en nuestro caso, los emergentes discursivos de la (in)seguridad lingüística en un entorno de alto riesgo.

Es en este punto que, a la luz de la naturaleza del dato lingüístico disponible en línea, e integrando las consideraciones previas, pueden proponerse algunas dimensiones descriptivas del dato en línea y su articulación en ejes orientadores para orientar la reflexión ética en torno a su aprovechamiento en investigaciones de tipo cualitativo.

Dimensiones del dato lingüístico en línea y ejes de reflexión ética

Los interrogantes básicos a tener en cuenta a la hora de determinar la conveniencia y modalidad apropiada para aprovechar los datos lingüísticos en línea pueden analizarse de acuerdo con las dimensiones interaccionales y técnicas que permiten describirlos. Entre las primeras, que reflejan las complejas relaciones entre los participantes tanto en las esferas *online* como *onground*, pueden mencionarse las siguientes dimensiones:

- a) *Tipológica*: se trata de la naturaleza predominante de los datos que se producen en una

interacción (*escrita u oral*). Se vincula con la dimensión técnica de *perennidad*.

- b) *Origen metodológico*: central para los fines de este trabajo, esta dimensión considera si el dato en línea fue obtenido por el investigador (*elicitado*) o no (*disponible*).
- c) *Temporal*: contempla si la interacción entre los sujetos observados o, si corresponde, entre estos y el investigador se produce de forma (cuasi) sincrónica/asincrónica.
- d) *Identitaria*: refiere a la posibilidad que puedan tener terceras personas de identificar *a posteriori* al emisor del enunciado x tomado como dato.
- e) *Temática*: trata de la naturaleza más o menos sensible del tópico al que se vincula el dato, evaluado desde una perspectiva lo más cercana posible a la de los integrantes de la comunidad en la que se produce.
- f) *Concepcional*: vinculada con la precedente, refiere a la manera de entender la relación de lo *público* y lo *privado* al interior del contexto comunicativo del que el sujeto productor del dato participa.
- g) *Genérica*: relacionada con las precedentes, las características del género también hacen al tipo de dato porque los géneros en línea se vinculan de manera muy estrecha con las dimensiones técnicas (correo electrónico/posteo de blog/posteo de foro/posteo de red social, etc.).

Entre las dimensiones asociadas con la esfera técnica, por otro lado, un dato en línea puede describirse de acuerdo con:

- a) Su *perennidad*: esto es, la permanencia temporal del dato en línea (a título ilustrativo, en el caso del *live-streaming* esta es efímera o muy baja, mientras que en el caso de comentarios de lector, por ejemplo, es duradera). Esta característica condiciona las posibilidades de registro de los datos y constitución y rasgos del corpus que puede construirse a partir de ellos.
- b) Su *indexación*: alude a la posibilidad de rastrear un enunciado particular a través de un buscador externo a la plataforma en la que se registró.
- c) Su *etiquetado*: posibilidad de rastrear un enunciado dentro de los límites de una plataforma determinada por sus metadatos (por ejemplo, en el caso de *Twitter*, mediante el uso de una búsqueda por *hashtag*).

Estas dimensiones están interrelacionadas y permiten formular, al menos, tres ejes para una reflexión ética, que pueden resumirse en los siguientes interrogantes básicos y sistematizarse en una serie de criterios operativos para guiar la práctica de investigación (*imagen 2*):

- a) *Eje interaccional*: ¿Existe una vinculación entre el sujeto investigador y el sujeto investigado al momento de producirse y registrarse -por el medio que sea- los datos?
- b) *Eje de privacidad*: ¿Cómo se vinculan el sujeto observado, el tópico que trata y el contexto comunicativo?

- c) *Eje de trazabilidad*: ¿Cómo se relaciona la perennidad del dato con su indexación, con la disponibilidad de metadatos y con su registro en un corpus?

Orientación práctica

Dentro del enfoque práctico de este trabajo, en esta sección se propone un posible recorrido reflexivo bajo la orientación de las dimensiones descriptivas del dato en línea y de los ejes propuestos a fin de establecer criterios para un tratamiento ético de los datos. Aunque las pautas ofrecidas en las siguientes secciones se presentan, en su mayoría, como sugerencias de orden general, cabe destacar, que de ninguna manera esta propuesta pretende ser absoluta y que, por el contrario, responde a la reflexión realizada en el marco de una investigación sociolingüística y discursiva de orientación aplicada: la comunicación *en y sobre* el ámbito aeronáutico en Argentina. En la primera orientación, esta labor supone la consideración de datos tanto orales como escritos y, en la segunda, el análisis complementario de datos escritos en línea. Salvo en el caso de entrevistas, que no hemos realizado en línea hasta el momento, el uso de datos interaccionales orales ha concentrado nuestra mayor preocupación sobre los datos disponibles en línea atendiendo a necesidades prácticas de nuestra investigación: aumentar el volumen de las muestras de interacción entre pilotos y controladores de tránsito aéreo reduciendo la necesidad de observaciones *in situ* para detectar fenómenos aislados de interés en una interacción altamente rutinaria. Sin embargo, por su interés intrínseco también consideraremos con cierta extensión el tratamiento de los datos escritos disponibles en línea, dejando de lado un tratamiento profundo sobre los datos elicitados en línea.

Datos elicitados en línea

Si el dato en línea es elicitado por el investigador a través de una interacción con el sujeto investigado, lo más probable es que el dato corresponda al ámbito de las CMC sincrónicas-orales (por ejemplo, vía *Skype*¹, o cuasi-sincrónicas/asincrónicas-escritas (por ejemplo, *chat*, correo electrónico, otros tipos de mensajería) y el trabajo puede realizarse con el consentimiento informado que se emplea como salvaguarda mutua del trabajo etnográfico en contextos no virtuales.

Tanto las condiciones de la interacción, como las del registro, archivo y empleo de estos datos son aspectos que pueden acordarse entre los participantes del proceso de investigación, pues se establece una relación análoga a la de cualquier investigación *onground* y, por tanto, pueden aplicarse los criterios y prácticas éticas generales a las ciencias sociales (véase, entre otros, Wiles *et al.*, 2005; Newman, 2009).

Datos disponibles en línea

¹ Otros conocidos servicios del tipo *voice over IP system* con mensajería instantánea similares a *Skype* son *VSee* (propietario) y *Google Hangout* (libre).

Cuando el investigador accede a datos que, simplemente, están disponibles en línea, las situaciones difieren según su naturaleza tipológica, oral o escrita, pero todos deberían ser utilizados -si se decide que es necesario y ético hacerlo- adoptando las medidas de enmascaramiento o anonimización más ajustadas al específico entorno *onground* bajo estudio al que corresponden.

Datos orales

Entre los datos disponibles, los orales pueden:

- a) Corresponder a un servicio de *live-streaming* público de audio (o, potencialmente, video), es decir, acceso a la escucha (visión) sin necesidad de registrarse como usuario en un sistema.
- b) Ser archivos puestos a disposición por usuarios de servicios para compartir audio y video².

En el primer caso, la perennidad de los datos de *live-streaming* no se correlaciona con su naturaleza oral, puesto que algunos proveedores -es el caso de *LiveATC*³ para la aviación- ofrecen un registro (aunque por tiempo limitado) al que solo pueden acceder usuarios registrados y una transmisión en vivo de carácter abierto.

Para los casos sin registro, el uso de los datos es una posibilidad que, a nuestro juicio, no supone un riesgo importante para los sujetos observados. Sin embargo, tras un proceso en el que experimentamos con distintas alternativas a lo largo del tiempo, en nuestro caso, hemos propuesto modificar sistemáticamente matrículas, compañías números de vuelo, centros de control y, por supuesto, eventuales nombres cuando excepcionalmente decidamos reproducir algunos de estos datos por su relevancia lingüística.

Si los datos transmitidos en vivo son registrados -aunque sea temporalmente- por el servicio que los ofrece, suelen existir restricciones de acceso que hacen necesaria una evaluación más pormenorizada de las condiciones establecidas por el proveedor para su empleo.

Por otro lado, como pauta general, si no hay declaración de las intenciones del sitio, no se establecen condiciones de uso y *no existen* otros proveedores del mismo tipo de datos, cabe que el investigador se pregunte cuál es el derecho de ese proveedor a difundirlos y con qué fin lo hace y, por su propia protección legal, sería conveniente evitar su empleo.

² Servicios populares para *video sharing* son *YouTube*, *Vimeo*, *Flickr*, entre otros.

³ *LiveATC* (www.liveatc.net) permite acceder a *audio-streaming* en tiempo real de distintos centros de control aeronáutico de todo el mundo, exceptuando de forma explícita los de los países donde esta práctica resulta ilegal. Sus Términos de Uso estipulan que el sitio tiene fines educativos y prohíben la explotación de los datos con fines comerciales o judiciales. Otras plataformas menos conocidas permiten acceder a *video-streaming* aunque, en la mayoría de los casos, sin audio (<http://www.webcamhopper.com/es/airport.html>).

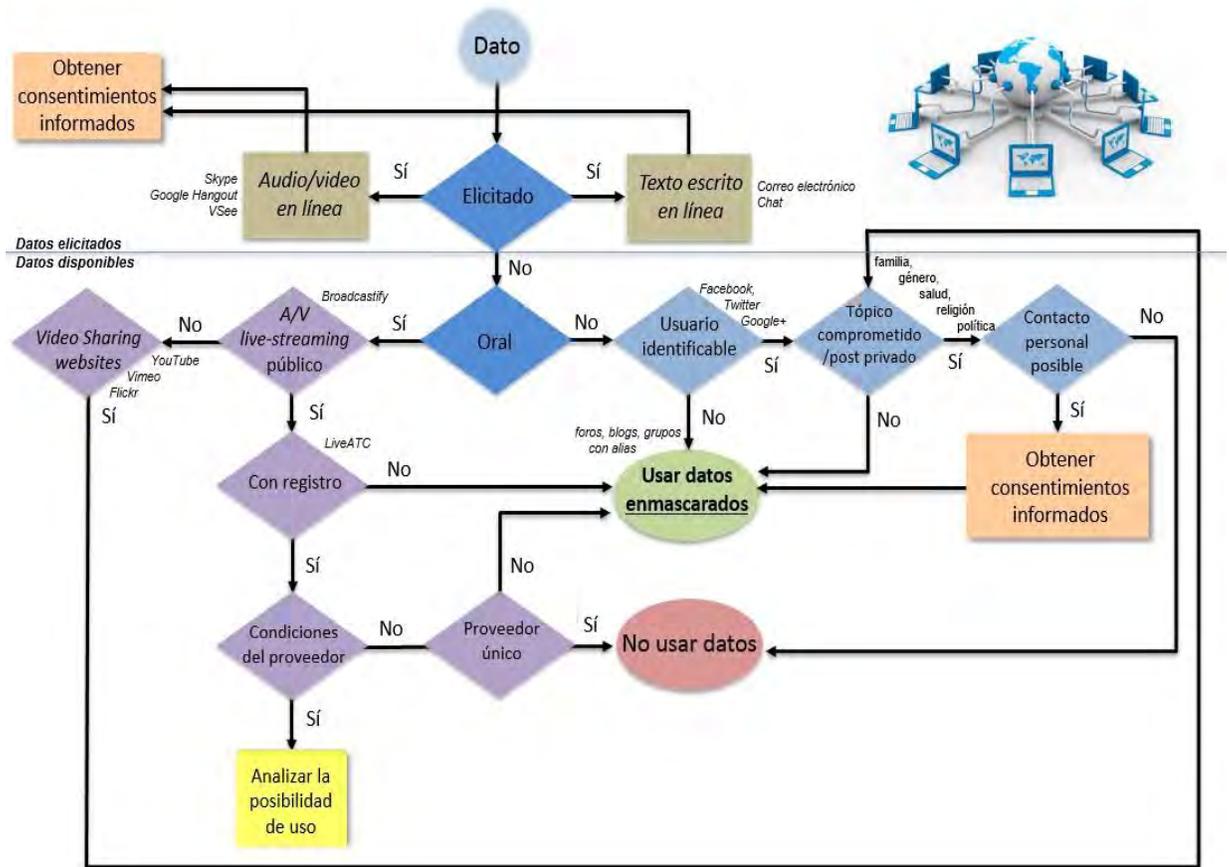


Imagen 2. Pautas orientadoras para el uso apropiado de datos en línea.

En el caso de nuestra investigación, aunque no siempre están operativos y han ido cambiando con el tiempo, algunos centros de control aeronáutico de Argentina -nos referimos, en especial, a las torres de control de los principales aeropuertos-, han tenido más de un sitio *web* desde donde acceder al *live-streaming* de audio. Si bien algunos ya no están disponibles o remiten a los datos de *LiveATC*, lo que nos interesa destacar, a partir de este caso particular, es la utilidad de considerar también la multiplicidad de fuentes como un criterio, aunque de ninguna manera absoluto (de ahí que no lo hayamos incluido en el diagrama de la *imagen 2*), para concluir la ausencia de mala intención por parte de quien pueda administrar canales de *live-streaming* correspondientes a otros ámbitos.

Tanto en el caso de datos de *live-streaming* como también en el de archivos de audio y video compartidos por usuarios, el eje de privacidad y, en particular, la dimensión temática del dato, deberían operar, también, como otro criterio clave para definir la conveniencia de su aprovechamiento. En el caso de videos compartidos con fines sociales, esto no significa que su uso en investigación esté liberado y, como hemos señalado, si se decide emplearlos sería ideal obtener

el consentimiento de los sujetos y, en su defecto, resulta imperativo adoptar medidas para su protección, enmascarando las identidades mediante la omisión de nombres (tanto en la transcripción escrita como en una ocasional reproducción audiovisual), la difuminación de rostros en cuadros o imágenes, etc. La remisión al video en la plataforma resulta, por lo señalado, problemática y no debería realizarse⁴.

Datos escritos

Si la naturaleza del dato disponible en línea es escrita, esto supone una mayor perennidad, aspecto que puede ser más o menos transparente para los usuarios de las distintas plataformas. Por lo tanto, su empleo supone, en potencia, mayor riesgo para los sujetos porque, si no se toman las medidas necesarias de protección, podría rastrearse su identidad.

Aquí adquieren su mayor relevancia los aspectos del dato que hacen al eje de la privacidad y al de la trazabilidad pues, ante tópicos de mayor compromiso (relaciones personales–familiares, sociales, laborales, género, salud, religión, política, para solo mencionar algunos de los de mayor interés para la investigación cualitativa en ciencias sociales), resulta crucial proteger la identidad real de los sujetos observados.

Con respecto a la problemática de la privacidad, hay que tener presente que, si bien el investigador puede ser un partícipe activo en la plataforma de la que toma los datos, esto no significa que haya establecido un acuerdo de investigación con el sujeto observado que lo habilite a utilizarlos como si los hubiera elicitado. Por esto, en principio, un primer aspecto a considerar es la evaluación del compromiso que el tema supone para los sujetos que observa, evaluación en la que debería apelarse a todos los medios disponibles para valorar con honestidad en qué medida estos conciben su contexto de enunciación como *público* o *privado*⁵.

En este sentido, si el tema es comprometedor, puede resultar conveniente considerar un contacto personal para obtener el consentimiento informado antes de reproducir el(los) enunciado(s) de interés (Wiles *et al.*, 2005). Baker (2010) retoma en este sentido, la aplicabilidad del concepto de “consentimiento implícito” ya que, en determinados contextos puede resultar preferible para los hablantes evitar firmar un consentimiento para permanecer completamente anónimos⁶.

Si el contacto personal con los sujetos no fuera posible, convendría evitar el uso explícito de tales enunciados o citarlos de manera acumulativa (a la manera de frases sueltas de naturaleza/ contenido similar que se citan quitando toda referencia al contexto que proporcione elementos

⁴ En este punto, la cuestión de fondo se reduce a discutir si la validez científica del dato frente a la comunidad científica justifica la violación del derecho a la intimidad de los sujetos observados.

⁵ Al proponer pautas para el uso ético de datos de *Twitter*, por ejemplo, Rivers y Lewis (2014) destacan que “Twitter participants can reasonably expect to rely on some anonymity of the crowd to manage privacy. A tweet author discussing his mental health does not do so with the intention of sharing that data with researchers; he does it to communicate with his digital community”.

⁶ “[...] Silence was considered to indicate ‘implied consent’, which Martin and Knox (2000) have argued is preferable to making certain types of participants sign consent forms when they would rather remain completely anonymous” (Baker, 2010: 100).

para su rastreo en línea)⁷. En el caso de que la dimensión temática se valore como no comprometedor, el dato podría ser utilizado sin intentar establecer un contacto personal aunque, por norma general, en uno y otro caso se mantiene la necesidad de enmascarar los datos reales del sujeto⁸.

El eje de la trazabilidad, de naturaleza técnica, debe ser considerado de manera paralela, para determinar las posibilidades de que el enunciador de un texto que el investigador selecciona como dato para su uso y reproducción explícita pueda ser rastreado e identificado. El riesgo, aunque parezca menor, siempre existe pues, en la mayoría de las redes sociales (*social networking websites*), se crean perfiles individualizados, muchas veces incluso con el nombre verdadero de los usuarios. Con cierta frecuencia, además, aunque utilicen un *nickname*, los usuarios no aplican correctas políticas para controlar su exposición o, frente a los cambios adoptados por las plataformas, las tienen desactualizadas (algo que puede asumirse como una premisa metodológica de trabajo). En otros casos, hay que tener en cuenta que las referencias personales que realizan en sus respectivas *timelines* pueden permitir su identificación.

Además de estos factores inherentes a la seguridad de las cuentas en distintas plataformas, en la identificación de la fuente de un enunciado resulta clave la indexación del texto en los principales buscadores en línea. En el caso de las redes sociales, los textos no son indexados y sus autores están, en cierta medida, protegidos. Sin embargo, las plataformas que adoptan el etiquetado de los enunciados (los denominados *hashtags*) introducen una nueva manera de encontrar al emisor de un texto determinado desde *dentro* de la plataforma.

Por otro lado, algunos enunciados registrados en foros y blogs sí pueden estar indexados en buscadores (esto ocurre con frecuencia en los sectores públicos de los foros y de algunos blogs). Sin embargo, en estos casos se trata de espacios donde es más frecuente utilizar *nicknames* para identificarse y solamente los moderadores tienen acceso a las verdaderas identidades de sus usuarios, lo que habilita la cita de los datos (de preferencia, con enmascaramiento). Algo similar ocurre con los comentarios en diarios digitales y otro tipo de participaciones que los usuarios pueden realizar en línea.

› **Conclusión**

La distinción entre datos elicitados y disponibles en línea resulta operativa a la hora de

⁷ Rivers & Lewis (2014: 6) proponen evitar la cita directa de enunciados de *Twitter* y reemplazarla por “A composite of multiple example tweets”.

⁸ En las discusiones de la edición 2014 del MOOC sobre *Corpus Linguistics: Method, Analysis, Interpretation* (Universidad de Lancaster, a cargo de un equipo docente liderado por Tony McEnery), la Dra. Elena Semino propuso como directrices éticas para la cita de datos escritos disponibles en línea las siguientes: considerar solamente producciones escritas a las que puede accederse sin necesidad de ser un usuario registrado con clave, contactar al administrador del sitio para solicitar autorización para recolectar datos permitiéndole rechazar el pedido, usar los datos anonimizados evitando nombres y *nicknames* de usuarios e, incluso, omitir la referencia directa al sitio de internet de donde se toman los datos citados.

orientar la reflexión ética sobre su uso en investigaciones de tipo cualitativo en las que el ejemplo particularizado resulte imprescindible. Resulta claro que en el empleo de los datos *disponibles* en línea, sean orales o escritos, resulta necesario e insoslayable proteger la identidad de los sujetos observados, pero no siempre es suficiente adoptar las mismas medidas de anonimización empleadas en el trabajo en contextos interaccionales no virtuales o con datos elicitados. La reflexión realizada puso en evidencia la necesidad de tener en cuenta una serie de rasgos o dimensiones interaccionales y técnicas adicionales a la hora de establecer cuándo y cómo se anonimizarán los datos en línea. Su carácter en apariencia público no debe engañar al investigador ni lo exime de proteger a aquellos sujetos que -muchas veces sin saberlo- constituyen su objeto de estudio, con independencia del ámbito social al que pertenezcan estos datos. Ya no es suficiente la modificación/omisión sistemática de los nombres de los hablantes, sino que, dadas las características técnicas que hacen a la obtención de datos *en o gracias* a la esfera *online* de investigación, debe realizarse una evaluación pormenorizada de cada información de interés a fin de contar con una descripción acabada del dato a la hora de definir medidas adecuadas para su aprovechamiento ético y adaptadas a cada situación.

Las dimensiones del dato en línea, así como los interrogantes básicos que definen los tres ejes de reflexión propuestos no pretenden ser exhaustivas ni absolutas las orientaciones prácticas descritas para operar con datos disponibles en línea. Este carácter inacabado y tentativo refleja, por un lado, la imposibilidad de una Deontología única en las ciencias sociales, y por el otro, la dinamicidad y los desafíos que supone el nuevo entorno en el que avanza la investigación social desde finales del siglo XX.

› **Bibliografía**

Bailey, M., Dittrich, D., Kenneally, E. & Maughan, D. (2012). The Menlo Report. *IEEE Security & Privacy Magazine*, 10(2), 71-75.

Baker, P. (2010). *Sociolinguistics and Corpus Linguistics*. Edinburgh: Edinburgh University Press.

Berry, D. (Ed.) (2012). *Understanding Digital Humanities*. New York: Palgrave/MacMillan.

Buchanan, E. (Ed.) (2004). *Readings in Virtual Research Ethics. Issues and Controversies*. Londres: Information Science Publishing.

Crystal, D. (2005, febrero). *The Scope of Internet Linguistics*. Recuperado de http://hostmaster.davidcrystal.com/DC_articles/Internet2.pdf el 25/08/2015

de-Matteis, L. M. A. (2013). Deontología en la investigación sociolingüística de las organizaciones aeronáuticas: acceso y finalidades. En Carranza, I. & Vidal, A. (Eds.), *Lingüísticas del uso. Estrategias metodológicas y hallazgos empíricos* (107-124). Mendoza: Universidad Nacional de Cuyo/Sociedad Argentina de Lingüística.

Dwyer, A. M. (2007). Ética y aspectos prácticos del trabajo de campo cooperativo. En Haviland, J. & Flores Farfan, J. A. (Eds.), *Bases de la documentación lingüística*. México: Instituto

Nacional de Lenguas Indígenas, 49-89.

Finegan, E. (1993). Ethical Considerations for Expert Witnesses in Forensic Linguistics. *Issues in Applied Linguistics*, 4(2), 179-187.

Hamilton, H. (1993). Ethical Issues for applying Linguistics to Clinical Contexts: the Case of Speech-language Pathology. *Issues in Applied Linguistics*, 4(2), 207-223. [Traducción al español: (1996). Consideraciones éticas sobre la aplicación de la lingüística en contextos clínicos: el caso de la patología del lenguaje y del habla. *Lengua y Habla*, 1(1), 51-65. Recuperado de <http://dialnet.unirioja.es/descarga/articulo/4848406.pdf> el 29/8/2015.

Herring, S. (1996). Linguistic and Critical Analysis of Computer-mediated Communication: some Ethical and Scholarly Considerations. *The Information Society*, 12, 153-168.

--- (1997). Ethics in Cyberresearch: to cite or not to cite. *The College*, 1(2), 19-23. Recuperado de <http://ella.slis.indiana.edu/~herring/cola.1997.pdf> el 25/08/2015

Holton, G. (2009). Relatively Ethical: a Comparison of Linguistic Research Paradigms in Alaska and Indonesia. *Language Documentation & Conservation*, 3(2), 161-175.

Labov, W. (1982). Objectivity and Commitment in Linguistic Science: The Case of the Black English Trial in Ann Arbor. *Language in Society*, 11(2), 165-201.

Maczewski, M., Storey, M. A. & Hoskins, M. (2004). Conducting Congruent, Ethical, Qualitative Research in Internet-mediated Research Environments. En Buchanan, E. (Ed.), *Readings in Virtual Research Ethics. Issues and Controversies* (62-79). Londres: Information Science Publishing.

Markham, A. (2011). Internet Research. En Silverman, D. (Ed.), *Qualitative Research: Theory, Method and Practices* (111-127). Londres: Sage, Recuperado de www.markham.internetinquiry.org/writing/silverman2011draft.pdf el 06/02/2015

Newman, P. (2009). Fieldwork and Field Methods in Linguistics. *Language Documentation & Conservation*, 3(1), 113-125.

Rice, K. (2004). Ethical Issues in Linguistic Fieldwork: an Overview. *Journal of Academic Ethics*, 4, 123-155. Recuperado de <http://projects.chass.utoronto.ca/lingfieldwork/pdf/2.pdf> el 25/08/2015

Rivers, C. & Lewis, B. (2014). Ethical Research Standards in a World of Big Data. *F1000 Research*, 3, 38.

Wiles, R., Heath, S. Crow, G. & Charles, V. (2005). Informed Consent in Social Research: a Literature Review. *ESRC National Centre for Research Methods*, NCRM/001. Recuperado de <http://eprints.ncrm.ac.uk/85/1/MethodsReviewPaperNCRM-001.pdf> el 25/08/2015

Wolfram, W. (1993). Ethical Considerations in Language Awareness Programs. *Issues in Applied Linguistics*, 4(2), 225-255.