

Viedma 8 y 9 de noviembre de 2018



## Desafíos metodológicos en una investigación doctoral sobre discurso digital

**Lucía Cantamutto**

Depto. de Humanidades, Universidad Nacional del Sur, Bahía Blanca/CONICET-Centro Universitario Regional Zona Atlántica, Universidad Nacional del Comahue

**Proyecto:** "Estilo(s) comunicativo(s) y variación pragmática en la interacción verbal del español bonaerense: construcción de identidades, valores y creencias" (SCyT-UNS)

### RESUMEN

En las diferentes etapas realizadas en el marco de la investigación doctoral sobre SMS del español bonaerense, tuvimos que tomar una serie de decisiones metodológicas dado el carácter dinámico (y evanescente) del objeto de estudio. En este trabajo presentaremos un panorama sobre el modo en que se condujo esta investigación que se concretó en la tesis "Estrategias pragmáticas de la comunicación por SMS en español bonaerense". El objetivo de este artículo es identificar algunos de los problemas metodológicos que emergen en el transcurso de una investigación sobre el discurso digital cuya perspectiva teórica es la Sociolingüística Interaccional y la Pragmática Sociocultural. En tal sentido, se exponen reflexiones en torno a las vicisitudes propias de cualquier investigación lingüística sumadas a las inherentes al trabajo con objetos digitales.

**Palabras clave:** Comunicación digital; Humanidades digitales; Sociolingüística; Metodología de la investigación; Discurso digital.

Viedma 8 y 9 de noviembre de 2018

## Introducción

La realización de una investigación doctoral es siempre compleja y las condiciones para que esta concluya favorablemente (es decir, con la defensa de la tesis) son siempre multivariadas (véase, entre otros, Stemberg, 2014). Este proceso, con una duración estimada de 5 años en Argentina<sup>1</sup>, puede extenderse indefinidamente o demorarse durante varios años produciendo diferentes desajustes entre el plan de trabajo y cronograma inicial de la investigación, los alcances de la misma e, incluso, la factibilidad de recoger los datos necesarios para llevar a cabo una investigación de esta envergadura. Por otro lado, además de los factores inherentes a la investigación que pueden influir en la concreción de la tesis, se adicionan aquellos relativos a las relaciones humanas (vínculo con directores y codirectores) y a cuestiones personales sobre las que no es posible anticiparse. En resumen, como cualquier proceso prolongado en el tiempo, realizar una investigación doctoral implica la toma de decisiones tanto en la diagramación inicial como ante las diferentes dificultades coyunturales que aparezcan.

En este trabajo expondremos algunas de las decisiones metodológicas que se debieron tomar en relación a la perspectiva teórica adoptada –interaccional– y el objeto de estudio que nos interesaba –la comunicación digital–. El objetivo es problematizar el diseño y el devenir de la investigación de una tesis doctoral enmarcada en la Sociolingüística Interaccional (Gumperz, 2001; Tannen, 1996), la Pragmática Sociocultural (Bravo & Briz, 2004), la Ciberpragmática (Yus, 2010) y el Análisis del Discurso Digital (Cantamutto & Vela Delfa, 2016a; Thurlow & Mroczek, 2011). La conjetura que guía esta exposición es que toda investigación social presenta desafíos metodológicos que, en este caso, se conjugan los propios de la investigación sociolingüística (definición de variables) y los de la investigación sobre el discurso digital (mutabilidad de las prácticas sociales).

En las siguientes páginas presentaremos, en primer lugar, los rasgos generales de la investigación llevada a cabo entre los años 2012-2017 (§2), en segundo lugar, los principales obstáculos enfrentados junto a las soluciones encontradas (§3). Por último, cierra este trabajo una serie de reflexiones sobre cómo extrapolar esta experiencia en otras investigaciones.

## 2. El plan de investigación

El proceso de elaboración de un plan de investigación comienza mucho antes que su redacción. En tal sentido, Howard Becker señala “cuando alguien se sienta a escribir, ya ha tomado muchas decisiones, pero es probable que no sepa cuáles fueron” (2016, p. 35). Las ideas iniciales, la identificación y el recorte del tema, el planteamiento del problema y el bosquejo de algunos objetivos son instancias previas a la elaboración del plan que guiará el proceso de investigación.

En su libro *Cómo hacer una tesis*, Carlos Sabino indica cuáles son las dificultades que se deben enfrentar en esta etapa inicial. A continuación, replicamos los títulos de lo que el autor desarrolla en esas páginas (1998, pp. 85–90):

- a. Buscar un problema de investigación que resulte de interés para el tesista;
- b. Escoger una temática conocida;
- c. Buscar áreas de trabajo en las que pueda contarse con una ayuda efectiva;
- d. Seleccionar un tema concreto y accesible.

Esta sistematización resulta clarificadora respecto al devenir del proceso inicial de investigación: ordenar las ideas y motivaciones en un tema que se *pueda* abordar en los plazos estipulados (más o menos laxos, según haya o no financiación de por medio). De hecho, sortear la primera instancia de *elegir* el problema de investigación según los intereses del tesista (y no en función de lo que los directores o equipos de investigación deseen) produce una tensión difícil de revertir en centros de investigación pequeños o en los cuales la oferta es limitada. El tesista requiere –casi en la totalidad de los casos– de figuras que dirijan y supervisen su labor. En tal

---

<sup>1</sup> El Consejo Nacional de Investigaciones Científicas y Técnicas otorga becas para la realización de doctorados con una duración de 5 años. En ese período se estipula que un graduado, dedicado exclusivamente a esta tarea, puede comenzar, realizar y defender su investigación doctoral.

Viedma 8 y 9 de noviembre de 2018

sentido, deberá encontrarse el punto en común entre los temas de interés de ambas figuras y, por tanto, el área de conocimiento en la cual se pueda contar con *ayuda efectiva*. La factibilidad es una variable de gran peso en la concreción de una investigación y, en el caso particular de la tesis doctoral, los recursos humanos son parte de los argumentos a tener en cuenta a la hora de elegir dónde y con quién realizar este trayecto de formación.

En la investigación doctoral a la que haremos referencia, el tema se centró, inicialmente, en algunas prácticas lingüísticas de los usuarios telefonía móvil, particularmente, en el intercambio de mensajes de texto (SMS) dentro de la variedad dialectal del español bonaerense. El encuadre teórico, tal como se mencionó, fue el de la sociolingüística y de la pragmática sociocultural, para relevar regularidades discursivas y fenómenos de variación pragmática (Tagg, 2009; Vela Delfa, 2007) y sociolingüística, asociados a variables sociodemográficas y contextuales, en esta variedad dialectal del español de Argentina. La delimitación de este tema tuvo un antecedente de importancia, la tesina de grado, en la cual se focalizó sobre los fenómenos morfosintácticos en la comunicación por SMS en el grupo etario de jóvenes. Esta primera decisión fue fundamental para el éxito de la investigación mayor: contar con un antecedente de relevancia tanto respecto al tema de estudio como sobre las herramientas necesarias para desarrollar una investigación.

La redacción del plan fue prevista tanto como guía de la investigación como para su presentación a diversos programas de becas y para la inscripción en el Programa de Doctorado seleccionado. En cada una de estas presentaciones, diversos ajustes teóricos se fueron realizando pero las cuestiones metodológicas que nos interesan en este texto no pudieron ser previstas anticipadamente. La mayoría de ellas surgieron en el transitar de una investigación y a partir de situaciones que irrumpieron en el trabajo de campo y en tratamiento de los datos.

Por último, antes de exponer las principales decisiones que permitieron llevar adelante la investigación, hay que recordar que –precisamente– para poder tomar estas decisiones, es necesario contar con una *ruta de navegación*, es decir, con un buen *plan de investigación* como texto que guía la investigación. El investigador novel debe volver a este documento una y otra vez tanto para seguir lo planteado inicialmente como para tomar decisiones ante problemas emergentes. En ocasiones, la vorágine de los resultados preliminares alcanzados y los sucesivos *estratos* de interés a los que se va accediendo (al estar en contacto con el objeto de investigación) produce desconcierto respecto a qué era lo que se estaba buscando. Y ahí está el plan: con unos objetivos que se desgranar coherentemente, con hipótesis o preguntas de investigación que no deben perderse de vista, con las variables de interés y relevancia, con el resumen de lo que se quería hacer.

Las siguientes páginas no cuentan solo una experiencia personal de investigación sino que son un conjunto de situaciones *comunes* a las investigaciones doctorales, al menos, en el campo de estudio del discurso digital.

### 3. Tomar decisiones antes, durante y después

Con la lectura de la bibliografía científica necesaria para hacer el plan se anticipan muchos potenciales eventos, y así lo sostienen los manuales de investigación científica (Hernández Sampieri, Fernández-Collado, & Baptista Lucio, 2006, p. 33; Sabino, 1998, pp. 117–118); sin embargo, lo cierto es que hasta que no se está en situación los problemas reales no aparecen. De hecho, estos mismos autores insisten en que, al menos en la vertiente cualitativa, el transcurso de la investigación conduce a replanteársela “el proceso cualitativo no es lineal, sino iterativo o recurrente, las supuestas etapas en realidad son acciones para adentrarnos más en el problema de investigación y la tarea de recolectar y analizar datos es permanente” (Hernández Sampieri et al., 2006, p. 523).

Si bien el plan de investigación es el texto base sobre el cual el investigador *pivotea*, aquello que se haya esgrimido en esa instancia no es el recorrido definitivo. Es el punto de apoyo para avanzar y retroceder respecto a ese objetivo inicial y ese problema de investigación primigenio.

A continuación, haremos un análisis sobre las principales decisiones que se debieron realizar para llevar a cabo (con éxito) este trabajo científico-académico. En primer lugar, las inherentes al objeto de estudio (el discurso digital). En segundo lugar, las relativas a los antecedentes investigación (ante un fenómeno emergente y, al mismo tiempo, evanescente). En tercer lugar, decisiones que surgen de la perspectiva teórica adoptada y el problema de las variables. En cuarto lugar, cuestiones que se desprenden de la ética científica.

Viedma 8 y 9 de noviembre de 2018

## Los problemas de investigar el discurso digital

Las interacciones digitales se caracterizan tanto por su permanencia y por la posibilidad de trazabilidad de los datos (De-Matteis, 2016), como por su mutabilidad y evanescencia (Scolari, 2004, 2009). Los usuarios, por otra parte, tenemos una visión sesgada de los datos que producimos cuyo valor es alto en el mercado del capitalismo digital (Schiller, 2000). Los intercambios comunicativos de los usuarios se desarrollan en el amplio espectro de interfaces digitales y ocurren en distintos tipos de plataformas, aplicaciones y soportes. Cualquier investigación sobre el discurso digital debe plantearse esta problemática al delimitar su objeto de estudio: en la comunicación digital, ninguna práctica (lingüística o social) puede comprenderse en relación a un solo objeto o interfaz digital.

Sin embargo, a pesar de la heterogeneidad de las situaciones comunicativas en las que pueden ocurrir, es posible identificar una suerte de continuum entre las diferentes prácticas comunicativas de los usuarios hablantes y recortar aquello que podría denominarse discurso digital. Más allá de las implicaciones que tengan las propiedades genéricas del discurso digital y, por tanto, las muestras de lengua del discurso digital, el caso particular del intercambio de SMS requiere atención particular.

Dentro de las cuestiones que necesitan ser consideradas (Cantamutto & Vela Delfa, 2016b; Vela Delfa & Cantamutto, 2015, 2016), haremos una breve reflexión en torno a la inestabilidad de las muestras del discurso digital. La profusión, evolución, mutación y/o desaparición de los entornos digitales de comunicación (es decir, dispositivos y plataformas) se evidencia en un lapso breve: para el caso de los SMS, bastaron menos de dos décadas para que transforme su epíteto. El SMS ya no es el vector de la multitudes inteligentes (Rheingold, 2004) sino que se ha transformado en una suerte de casilla SPAM<sup>2</sup> en la que se recibe publicidades, ofertas, avisos y alertas.

Toda investigación que aborde prácticas sociales y/o lingüísticas en torno al discurso digital debe asumir la posibilidad de que su objeto de estudios cambie (en el mejor de los casos) o desaparezca (como sucedió en nuestra investigación). La recogida de datos, necesariamente, requiere comenzar ante los planteamientos iniciales del estudio. Caso contrario, los tiempos –ya mencionados– de un proyecto doctoral suelen ser más extensos que los de determinada práctica digital.

## El estado de la cuestión y las respuestas metodológicas: ¿cuántos datos conforman un corpus representativo?

Como señalamos, muchas de las respuestas surgen de la lectura continua y actualizada de bibliografía. Sin embargo, a la hora de buscar *respuestas* para nuestra investigación es necesario leer situadamente los antecedentes. El modo en que se produce ciencia varía de un país a otro, de una disciplina a otra, debido a diferentes políticas de financiación, dinámica de los equipos de investigación, requisitos de productividad, etc. La posibilidad de replicar estudios científicos es la base para garantizar la validez del conocimiento científico. Sin embargo, la replicación es factible en tanto se cuenten con los recursos (materiales y/o humanos) necesarios para llevarla a cabo. De esto se deriva la necesidad de leer la bibliografía específica teniendo en cuenta el lugar desde donde surge.

En línea con las dificultades intrínsecas al estudio de interacciones digitales, muchos investigadores con interés en los SMS señalan la carencia de corpus de este tipo discursivo de carácter íntimo y privado. Shortis (2016, p. 127), sintetizando aportes realizados entre 2006 y 2008, presenta un panorama general de los estudios de SMS en inglés, en el que claramente destaca el corpus CorTxT (Tagg, 2009), con cerca de 11000 SMS que continúa, en inglés, la tradición iniciada por los estudios en francés sobre corpus extensos de interacciones reales por SMS, con el proyecto Sud4Science (Panckhurst, 2013; Panckhurst & Moïse, 2012).

El predominio de trabajos con datos de interacciones del investigador (es decir, ejemplos personales) emerge en la enumeración realizada por Shortis, donde prevalecen muestras testigo (samples) y datos no especificados. En resumen, los estudios con muestras reales de SMS parecen alinearse entre los que usan un gran conjunto de datos y los que utilizan pequeñas muestras. En un tercer punto se encuentran aquellos que

---

<sup>2</sup> Tascón (2012, p. 489) define “La voz de origen inglés *spam* debe resaltarse en cursiva o, aún mejor, sustituirse por la expresión española correo no deseado o por la forma coloquial correo basura”.

Viedma 8 y 9 de noviembre de 2018

emplean dictados o transcripciones, técnicas que no servían para el interés y objetivos de nuestra investigación.

En nuestro caso, se intentó alcanzar la cantidad de datos utilizados por Tagg (2009) en su tesis doctoral. Siendo el antecedente directo y más cercano al momento de redacción del plan de investigación, el objetivo fue recolectar en la comunidad de habla de estudio (español bonaerense) la misma cantidad para poder comparar algunos resultados. Sin embargo, a pesar de haber utilizado similares estrategias metodológicas, antes de que se pudiera alcanzar la cifra mencionada, el SMS dejó de utilizarse en Argentina. Una década había transcurrido entre una investigación y otra: el objeto de estudio –como ocurre con asiduidad en el discurso digital– comenzó a ser una práctica en desuso y, por tanto, la recolección de datos no pudo llevarse a cabo.

La solución adoptada fue, similar al anteriormente mencionado trabajo de Shortis, la realización de una exhaustiva pesquisa bibliográfica en torno a las metodologías de recolección y cantidad de datos utilizados en trabajos de investigación de maestría y doctorado, así como artículos de investigación publicados sobre SMS. De este modo, se tuvo conocimiento de que, en la mayoría de los casos, los investigadores habían trabajado con pequeñas cantidades de SMS a modo de ilustración o ejemplos (Cantamutto, 2018, pp. 149–154). En tal sentido, se comprobó que los datos de SMS-CEBo utilizados en nuestra investigación eran suficientes.

Por último, la ausencia de corpus representativos de la comunicación digital en español obliga a cada investigador a recoger dificultosamente sus datos. Este aspecto ha sido subsanado a partir de la creación de CoDiCE<sup>3</sup> (Comunicación Digital: Corpus del Español), creado por Cantamutto, Vela Delfa y Boisselier (2015). En este corpus, además de los datos de SMS-CEBo, se encuentran muestras de lengua de diferentes variedades del español y producidas en diferentes dispositivos. El cotejo de los análisis realizados sobre un corpus particular en relación a las prácticas discursivas realizadas en otros corpus permite al investigador validar sus propios resultados: ¿es así la comunicación por SMS o es así la *interacción digital escrita* por teléfono móvil? ¿qué diferencias hay entre las producciones por SMS y por WhatsApp? Si bien son preguntas de investigación diferentes, la descripción de un conjunto de datos puede resultar más significativa si se cuenta con otras muestras de lengua de referencia para sopesar los fenómenos detectados.

### 3.4 La representatividad de las variables

Conectado con lo anterior, otro aspecto a considerar tiene que ver con la representatividad de las variables de investigación. En nuestro caso teníamos, por un lado, las sociolingüísticas y, por otro, las relativas a cuestiones técnicas/tecnológicas. Las conclusiones alcanzadas pusieron de manifiesto que una de las variables de mayor peso en la aparición o no de recursos expresivos era el tipo de teclado y el modelo del teléfono que los usuarios utilizan.

En estudios previos señalamos (Cantamutto, 2013, 2018) la gran dificultad que existía para recabar datos sobre las prácticas comunicativas de quienes habitan la periferia de la ciudad, y son considerados como pertenecientes al estrato sociocultural bajo, de menor instrucción y/o menores recursos. Si las muestras de lengua son donadas por miembros de ese grupo, ¿se autoperceben como tales? ¿en qué grupo de estudio se incluyen?

En tal sentido, respecto a las variables sociolingüísticas, Francisco Moreno Fernández (1998, p. 34) señala que los factores sociales que tienen mayor influencia en la variación lingüística son el sexo, la edad, el nivel socioeducativo, el nivel sociocultural y el origen étnico. La elección de las variables y los criterios para delimitar los grupos que se analicen a partir de ellas dependerán del interés particular de cada investigación. Sin embargo, los factores influyentes de cada variable, necesariamente, variarán también de una comunidad a otra: en cada una, los indicadores de educación, ocupación, ingresos, darán diferentes estratos dependiendo de los parámetros internos de la comunidad. Entre las propuestas metodológicas para la variable clase social, Moreno Fernández (1998, p. 46) recopila distintos modelos de estratificación social, entre ellos, los de Labov basados en una serie de puntos que combinan nivel de instrucción, ocupación e ingresos familiares.

<sup>3</sup> <http://codice.aplicacionesonline.com.ar>

Viedma 8 y 9 de noviembre de 2018

Cada investigador estratificará la sociedad en diferentes niveles según la profundidad y alcance del estudio, y según las características sociológicas de cada comunidad. La forma de denominar esta variable alterna entre clase, nivel sociocultural y nivel socioeconómico. La dificultad radica en recoger muestras de lengua que pertenezcan efectivamente al nivel propuesto. En el caso de la comunicación digital, el uso esporádico de dispositivos personales, la imposibilidad de contar con gatekeepers de la comunidad y la baja participación de los estudiantes de las escuelas periféricas tuvo como consecuencia una menor representatividad de los datos correspondientes al grupo denominado B, con menor nivel de instrucción y menores ingresos. En la investigación, por tanto, se analizaron los porcentajes obtenidos al interior de cada grupo (A y B) y no en relación al conjunto total del corpus.

Por otra parte, Moreno Fernández (1998, p. 55) separa la variable nivel de instrucción, considerándola clave para la sociolingüística por cómo determina la variación lingüística. Esta variable, al igual que las demás, es compleja de delimitar y estratificar así como de equipararla entre las distintas comunidades. Este punto resulta de importancia en nuestro análisis porque, según hemos observado en nuestra investigación, el nivel de instrucción no solo varía entre comunidades sino también entre escuelas. El nivel de escolarización lejos está de ser uniforme en el sistema educativo argentino.

Las dificultades para estratificar la sociedad así como recoger datos de los grupos con una consecuente identificación y autoidentificación de los grupos de pertenencia condujeron a identificar únicamente dos grupos en los que se validaran las variables clase social, nivel sociocultural y nivel de instrucción. De este modo, se evitó trazar criterios demasiado arbitrarios para ubicar en grupos diferentes al conjunto de personas que por su nivel de ingresos podrían pertenecer a una clase social *media-baja* pero cuyo nivel de instrucción es *primario*, por ejemplo. El análisis de los datos en torno a estos grupos produjo resultados que muestran un uso discursivo diferente de los SMS (el grupo A realiza intercambios breves, mientras que el grupo B realiza intercambios extensos), una utilización de mayor cantidad de recursos de abreviación en el caso del grupo B y una mayor complejidad sintáctica en el grupo A.

Al atender a la variable nivel sociocultural hay que considerar los diferentes usos sociales de la comunicación digital y del repertorio de dispositivos de los que el hablante dispone. La propia experiencia (Cantamutto, 2014) demostró que los hablantes del grupo sociocultural más bajo suelen no tener teléfono propio ni haber adquirido hábitos (y por tanto estrategias) en determinado dispositivo. Por el contrario, el tipo de interacciones relevadas dan cuenta de que los usos responden a etapas tempranas de la comunicación por SMS (similar al chat o con intercambios continuos), muchos mensajes indican la falta de crédito y solicitan ser llamados a otro número distinto al que están escribiendo.

En relación a la variable edad, gran parte de la bibliografía revisada sobre SMS se centra en los grupos de hablantes adolescentes y jóvenes, justificando esta elección en que es en este grupo etario donde se centra la mayor riqueza para el análisis. El habla adolescente presenta particular interés para cualquier sociolingüista, ya que en esta etapa de grandes cambios también se producen modificaciones en el plano de la dinámica lingüística con recurrencias al carácter lúdico y críptico.

Sin embargo, si es natural en los jóvenes utilizar formas lingüísticas que innovan en distintos niveles de lengua y construyen discursos identitarios ¿por qué la discusión sobre las consecuencias lingüísticas de las prácticas digitales recae sobre este grupo? Una de las causas principales es que constituyó el grupo etario que mayor cantidad de mensajes enviaba y recibía, además del que más se identificaba con esta tecnología (Rheingold, 2004).

La frecuente asociación entre innovación y jóvenes descansa sobre la idea de que los jóvenes no reconocen lo disruptivo de la novedad respecto a experiencias previas. La corta experiencia vital hace que los jóvenes se relacionen con el mundo sin compromisos con el pasado y, por tanto, “asumen como propia la condición inestable y cambiante del mundo” (Urresti, Linne, & Basile, 2015, p. 17). Esta apertura al mundo también ocurre a nivel del lenguaje, siendo los adolescentes y jóvenes un grupo social con una jerga que los identifica. En esta etapa de grandes cambios, se generan palabras nuevas que penetran con facilidad, debido a su necesidad de diferenciarse de otros grupos etarios (principalmente, los adultos) y de volver más oscuro su forma de comunicarse. Con frecuencia, expresiones que adolescentes y jóvenes generan pasan luego al lenguaje adulto (Zimmermann, 1996).

En la comunicación digital, el modo en que los jóvenes se apropiaron rápidamente de las tecnologías y de las prácticas comunicativas facilitó que se identifique a esta variedad (el *ciberlenguaje*, *netspeak*, *lenguaje*

Viedma 8 y 9 de noviembre de 2018

SMS) con una jerga juvenil que, sin embargo, es imitada por otros grupos etarios al igual de lo que ocurre en rasgos de la interacción cara a cara (Rígano, 1998).

Las investigaciones lingüísticas sobre prácticas digitales tienen que, en primer lugar, definir adecuadamente la variable edad atendiendo al uso que los adultos hacen de la lengua y, en segundo lugar, incorporar entrevistas o test de hábitos sociales para recabar información complementaria sobre las percepciones de estas expresiones “juveniles”. De todos modos, es difícil discernir entre la creatividad inherente al lapso que transcurre entre adolescencia y juventud y los diferentes recursos que los hablantes adoptaron por constricciones propias del dispositivo. Una revisión diacrónica de nuestro corpus indica que las abreviaturas, tan prolíficas en la primera etapa del corpus, ya no se verifican más que en hablantes de estrato sociocultural bajo y algunas abreviaturas estandarizadas. Es necesario, entonces, delimitar qué rasgos detectados responden a restricciones del dispositivo, cuáles a la competencia comunicativa de los jóvenes y cuáles al carácter lúdico y críptico propio del habla adolescente. En tal sentido, nuevamente, la variable dispositivo tecnológico y nivel de uso de la tecnología resultan de relevancia en estos estudios.

Por último, el análisis de las prácticas digitales no debería responder únicamente a la variable del grupo etario. Omitir otros perfiles de interactuantes puede conducir a observaciones erróneas, porque en la variación intragrupal hay una amplia riqueza de elementos para considerar. Por ejemplo, la variable sexo ha registrado ser significativa para detectar los usos y prácticas diferenciadas vinculados a cuestiones identitarias de hablantes femeninos y masculinos.

En términos generales, la estrecha relación entre prácticas sociales y prácticas comunicativas ligadas al teléfono móvil hace que los estudios adolezcan de representatividad en todas las variables. Este aspecto no debe considerarse

### 3.5. El tratamiento ético de los datos: entre anonimizar y ejemplificar

Desde una perspectiva ética, acceder a las interacciones por SMS implica irrumpir en la vida privada de los usuarios e involucrar información que puede perjudicar a otras terceras personas. Específicamente, las investigaciones con un enfoque cualitativo tienden a tener mayor amparo en la forma en que los datos se generan y resguardan, debido a la profundidad que este tipo de estudios suele alcanzar y, por tanto, se requiere la toma de decisiones metodológicas que favorezcan, simultáneamente, la validez externa de la muestra, la calidad de los datos y las consideraciones éticas pertinentes. Cualquiera de los sistemas de comunicación digital tiene mayor persistencia que la comunicación oral (Gobato, 2014, pp. 220–221) y, por lo tanto, será en mayor o menor medida recuperable.

En primer lugar, la investigación sobre los SMS tiene una ventaja respecto al estudio de otros géneros del discurso digital. A diferencia de las muestras de lengua de intercambios ocurridos en la red, los SMS no persisten en la nube, no están indexados a los buscadores web, y, por lo tanto, no dejan huellas que podrían ocasionar el rastreo de quién participó del intercambio. Al ocurrir en un sistema ajeno a internet, las búsquedas en servicios disponibles en los navegadores difícilmente arroje algún resultado ya que los SMS se almacenan en los teléfonos involucrados. Este problema, preferente en los enunciados escritos y en las imágenes, ha sido señalado por de-Matteis (2016, p. 245)

El eje de la trazabilidad, de naturaleza técnica, debe ser considerado de manera paralela, para determinar las posibilidades de que el enunciadore de un texto que el investigador selecciona como dato para su uso y reproducción explícita pueda ser rastreado e identificado.

Es decir, debido a la posibilidad de que, a partir del texto utilizado como ejemplo, pueda ser rastreada la identidad del colaborador, el investigador debe realizar diferentes procesos de anonimización y/o reconstrucción de los enunciados para evitar la identificación del informante.

Diversas discusiones se han desarrollado en torno a la ética de las investigaciones que utilizan datos de fuentes públicas de internet sin pedir consentimiento a los usuarios o a los dueños de las plataformas en las que ocurre el discurso, cuyos límites no están en la dicotomía público/privado (Estalella & Ardèvol, 2007). En tal sentido, la postura ética y epistemológica de los investigadores puede ser contradictoria, es la síntesis alcanzada por Estalella y Ardèvol (2007) de aportaciones de otras investigaciones donde se reflexiona en torno a la cuestión del grado de privacidad/publicidad y los permisos requeridos para trabajar con esas muestras de lengua. En las discusiones iniciales en el campo de investigación, algunos autores, como Susan Herring (1996), señalaban que algo que se encuentra en un espacio público manifiesta carácter público aún incluso cuando las

Viedma 8 y 9 de noviembre de 2018

expectativas de privacidad que suponen los participantes no siempre coinciden con la visión de los investigadores (Walther, 1996). Otro grupo de trabajos proponen criterios más concretos –relacionados con la arquitectura tecnológica del sitio– que aluden a factores como presencia/ausencia de contraseña para el acceso a la información, políticas de resguardo, o sensibilidad del tópico que se trata. Un tercer aspecto a considerar son las cláusulas legales que las páginas, redes sociales y dispositivos tengan respecto a los contenidos que ahí se producen. Estalella y Ardèvol (2007) recuperan la noción expectativas de privacidad de los usuarios. La noción de comunidad complementa esta idea, como otro límite difuso que influye sobre este continuum: es la sensación de relativa privacidad entre los miembros de dicha comunidad, como, por ejemplo, Twitter.

En cualquier caso, cada investigador deberá sopesar la trazabilidad entre el enunciado y el usuario y si se disponen de recursos para evitar la identificación, sin caer en el prejuicio de que los datos de la CMC en internet son fácilmente accesibles (Herring, 2002, p. 110). Si bien un usuario puede utilizar un blog como espacio de interacción público eso no implica que los materiales contenidos ahí estén a disposición de un investigador. Lo mismo sucede con las redes sociales públicas como Twitter, en cuyo caso, la temática abordada será un factor decisivo para solicitar consentimiento y autorización al usuario. En tal sentido, se deben implementar diversas estrategias para proteger a los participantes voluntarios de cualquier investigación, respetando su autonomía y cuidando de no afectar su privacidad. En particular, a partir de la firma de autorizaciones y consentimientos informados por parte de todos los interactuantes y, en el caso de que fueran menores, preferentemente también por parte de sus padres.

El protocolo que hemos seguido en esta investigación no ofrece riesgos potenciales para los participantes ya que los resultados publicados han sido a partir de datos totalmente anonimizados y que no son rastreables a través de los buscadores. En definitiva, toda investigación debe bogar por los criterios éticos básicos y generales indicados en el Informe Belmont (1979) –beneficencia, respeto, justicia– así como las aplicaciones correspondientes: 1) consentimiento informado, 2) evaluación de riesgos y beneficios y 3) selección de sujetos.

Esta decisión, sin embargo, nos obligó a trabajar con tres clases de muestras: la normalizada (para los análisis estadísticos sobre fenómenos pragmáticos), la anonimizada (para la presentación de los ejemplos) y la original (para verificar los fenómenos morfosintácticos presentes). El proceso de normalización y anonimización puede difuminar rasgos lingüísticos que responden a cada fenómeno en particular: por ejemplo, los mecanismos de expresividad con los que se han modificado las fórmulas de tratamiento nominales.

#### 4. Palabras finales

El estudio de la comunicación interacción digital escrita plantea al investigador un reto inmediato: la recolección de muestras de lengua de interacciones reales en un tipo discursivo que se caracteriza por su índole íntima y privada. Esta dificultad se ve acentuada por el retroceso del intercambio de SMS en virtud de la expansión de los mensajes instantáneos como Whatsapp.

En el marco de esta problemática, el objetivo de este trabajo fue presentar las decisiones metodológicas que se han tomado para poder llevar adelante, de manera exitosa, una investigación doctoral sobre una práctica digital en un marco teórico-metodológico particular. A partir de realizar un panorama sobre las formas de recolección y la cantidad de datos utilizados en diversos estudios, tanto para el español como para otras variedades lingüísticas, se observan las limitaciones y virtudes de estos métodos para la investigación en el ámbito de la sociolingüística interaccional, la pragmática y el análisis del discurso.

En resume, la concreción de cualquier investigación es multivariada. Si tenemos en cuenta esta situación compleja, en la que los factores que pueden dificultar el desarrollo de la misma pueden emerger en cualquier momento, no parece imposible alentar a los tesisistas a que el reconocimiento de las dificultades engrandece la investigación. De este modo, cada limitación que aparece se transforma en un desafío y en una toma de decisiones (amparada en la bibliografía de referencia) cuyo resultado tiene que ser la finalización de una investigación. Como hemos dicho muchas veces en nuestro equipo de investigación, cada uno de los trabajos realizados es una *investigación posible* dentro de otras. En cinco años, o más, es esperable que el objeto de investigación mute, surjan nuevas preguntas de investigación y los nuevos resultados obtenidos nos conduzcan a nuevos intereses. Ante esta realidad, inherente a todas las investigaciones, siempre está el *plan*

Viedma 8 y 9 de noviembre de 2018

inicial, la palabra del director y el deseo de finalizar, como dijimos, *una investigación posible*. En un futuro, siempre habrá posibilidad de realizar otras investigaciones relacionadas.

## Agradecimiento

A mi directora, Elizabeth M. Rigatuso con quien, en estos años de tensiones y encantos, nos tuvimos la paciencia necesaria para concretar este proyecto.

## Bibliografía

- Becker, H. (2016). *Manual de escritura para científicos sociales*. Buenos Aires: Siglo XXI editores.
- Bravo, D., & Briz, A. (2004). *Pragmática sociocultural: estudios sobre el discurso de cortesía en español*. Barcelona: Ariel.
- Cantamutto, L. (2013). La conformación de un corpus de mensajería de texto: la interacción verbal mediada digitalmente. In L. Gambón (Ed.), *IV Jornadas de Investigación en Humanidades. Homenaje a Laura Laiseca. Ebook* (pp. 97–106). Bahía Blanca: Universidad Nacional del Sur.
- Cantamutto, L. (2014). La brecha digital: problemas para abordar la variable socioeducativa en el estudio de la comunicación por SMS. *Textos. Revista Internacional de Aprendizaje y Cibersociedad*, 18(1), 55–67.
- Cantamutto, L. (2018). *Estrategias pragmáticas de la comunicación por SMS en español bonaerense*. Universidad Nacional del Sur.
- Cantamutto, L., & Vela Delfa, C. (2016a). El discurso digital como objeto de estudio: de la descripción de interfaces a la definición de propiedades. *Aposta. Revista de Ciencias Sociales*, 69(69), 296–323.
- Cantamutto, L., & Vela Delfa, C. (2016b). Repositorio colaborativo de comunicaciones digitales: aproximación a un corpus para el español. In *Actas de las 1eras Jornadas Nacionales de Humanidades Digitales: culturas, tecnologías, saberes*. Buenos Aires: FyL-UBA.
- De-Matteis, L. (2016). Ejes para una discusión del uso ético de datos interaccionales escritos y orales obtenidos en línea. Buenos Aires: FyL-UBA.
- Estalella, A., & Ardèvol, E. (2007). Field Ethics: Towards Situated Ethics for Ethnographic Research on the Internet. *Forum: Qualitative Social Research*, 8(3), 1.
- Gobato, F. (2014). *La escritura secundaria. Oralidad, grafía y digitalización en la interacción contemporánea*. Bernal: Universidad Nacional de Quilmes.
- Gumperz, J. (2001). Interactional Sociolinguistics: A Personal Perspective. In D. Schiffrin, D. Tannen, & H. Hamilton (Eds.), *The handbook of Discourse Analysis* (pp. 215–228). Malden: Blackwell.
- Hernández Sampieri, R., Fernández-Collado, C., & Baptista Lucio, P. (2006). *Metodología de la investigación* (4th ed.). México, D. F.: McGraw-Hill/ Interamericana.
- Herring, S. C. (1996). *Computer-mediated communication: Linguistic, social, and cross-cultural perspectives* (Vol. 39). John Benjamins publishing.
- Herring, S. C. (2002). Computer-Mediated Communication on the Internet. *Annual Review of Information Science and Technology*, 36(1), 109–168.
- Moreno-Fernández, F. (1998). *Principios de Sociolingüística y sociología del lenguaje*. Barcelona: Ariel.
- Panckhurst, R. (2013). A Large SMS Corpus in French: From Design and Collation to Anonymisation, Transcoding and Analysis. *Procedia - Social and Behavioral Sciences*, 95, 96–104. <https://doi.org/10.1016/j.sbspro.2013.10.627>
- Panckhurst, R., & Moïse, C. (2012). French text messages From SMS data collection to preliminary analysis. *Linguisticae Investigationes*, 35(2), 289–317. <https://doi.org/10.1075/li.35.2.09pan>

Viedma 8 y 9 de noviembre de 2018

Research, N. C. for the P. of H. S. of B. and B. (1979). *The Belmont Report: thical Principles and Guidelines for the Protection of Human Subjects of Research-the National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research*.

Rheingold, H. (2004). *Multitudes inteligentes. La próxima revolución social (Smart Mobs)*. Barcelona: Gedisa.

Rígano, M. (1998). El léxico de los adolescentes. In E. Rojas Mayer (Ed.), *IV Congreso de la Sociedad Argentina de Lingüística*. Tucumán: INSIL.

Sabino, C. A. (1998). *Cómo hacer una tesis y elaborar todo tipo de escritos*. Buenos Aires: Editorial Lumen-Humanitas.

Schiller, D. (2000). *Digital capitalism: Networking the global market system*. Massachusetts: MIT press.

Scolari, C. A. (2004). *Hacer clic. Hacia una sociosemiótica de las interacciones digitales*. (Gedisa, Ed.). Barcelona.

Scolari, C. A. (2009). Alrededor de la(s) convergencia(s) Conversaciones teóricas, divergencias conceptuales y transformaciones en el ecosistema de medios. *Signo y Pensamiento*, 28(54), 44–55.

Shortis, T. F. J. (2016). *Ortographic practices in SMS text messaging as a case signifying diachronic change in linguistic and semiotic resources*. University College London.

Sternberg, D. (2014). *How to complete and survive a doctoral dissertation*. New York: St. Martin's Griffin.

Tagg, C. (2009). *A corpus linguistics study of SMS text messaging*. University of Birmingham.

Tannen, D. (1996). *Género y discurso*. Barcelona: Paidós.

Tascón, M. (Ed.). (2012). *Escribir en internet Guía para los nuevos medios y las redes sociales*. Barcelona: Galaxia Gutenberg. Círculo de Lectores. Fundéu BBVA.

Thurlow, C., & Mroczek, K. (2011). *Digital Discourse. Language in the New Media*. New York: Oxford University Press.

Urresti, M., Linne, J., & Basile, D. (2015). *Conexión total*. Buenos Aires: Grupo Editorial Universitario.

Vela Delfa, C. (2007). *El correo electrónico: el nacimiento de un nuevo género*. Universidad Complutense de Madrid.

Vela Delfa, C., & Cantamutto, L. (2015). Problemas de recogida y fijación de muestras del discurso digital. *CHIMERA. Romance Corpora and Linguistic Studies*, 2, 131–155.

Vela Delfa, C., & Cantamutto, L. (2016). De participante a observador: el método etnográfico en el análisis de las interacciones digitales de WhatsApp. *Tonos Digital*, (31), 1–22.

Walther, J. B. (1996). Computer-mediated communication impersonal, interpersonal, and hyperpersonal interaction. *Communication Research*, 23(1), 3–43.

Yus, F. (2010). *Ciberpragmática 2.0. Nuevos usos del lenguaje en Internet*. Barcelona: Ariel.

Zimmermann, K. (1996). Lenguaje juvenil, comunicación entre jóvenes y oralidad. In W. Oesterreicher & K. Zimmermann (Eds.), *El español hablado y la cultura oral en España e Hispanoamérica*. Frankfurt: Vervuert.