Breve recorrido sobre los enfoques y métodos de análisis de la comunicación digital en español

Ana Pano Alamán
Université de Bologne, Italie

1 Introducción

El interés sobre el discurso digital en español ha ido creciendo en los últimos veinte años, como demuestran los numerosos proyectos dedicados a explorar la comunicación digital desde diversos enfoques lingüísticos, sociolingüísticos, discursivos y pragmáticos (Pano Alamán y Moya Muñoz 2016; Giammatteo *et al.* 2017; Robles Ávila y Moreno Ortiz 2019). Recientemente, los estudios se han orientado también hacia el diseño y aplicación de herramientas digitales para anotar y analizar automáticamente los datos que se generan, por ejemplo, en las redes sociales, datos que se utilizan como modelo de predicción de opiniones sobre cuestiones sociales, económicas, políticas y culturales.

Ante la diversificación de las investigaciones sobre discurso digital en español y frente a la velocidad con la que evolucionan las tecnologías, parece necesario detenerse a revisar los enfoques y métodos adoptados, desde un punto de vista que tenga en cuenta dicha evolución y que permita reflexionar sobre los desafíos que se presentan en este ámbito de estudio. En este sentido, el objetivo de este trabajo es ofrecer un panorama actualizado de las cuestiones teóricas y metodológicas debatidas en los estudios sobre la lengua española en la red. Delineamos un estado de la cuestión de la disciplina a partir de una revisión de los principales análisis del discurso digital en español realizados en los últimos años. Se identifican los temas y perspectivas abordados de forma más frecuente y los planteamientos metodológicos más recientes. Por último, se discuten algunas propuestas metodológicas y éticas provenientes de proyectos específicos y se identifican los retos a los que se enfrenta el estudio lingüístico y pragmático de la comunicación digital.

2 El análisis lingüístico de la comunicación digital en español

Las primeras investigaciones sobre el discurso digital en español, generalmente centradas en la interacción en chats, *Instant Messaging* (IM), SMS, correo electrónico y foros de debate, coinciden en que la comunicación en estos canales pone en crisis la dicotomía hablado-escrito (Yus 2001, 2010; Pano Alamán 2008). De ahí que se adopten etiquetas como "texto escrito oralizado" (Gómez Torrego 2001) o "lengua oral tecleada" (Laborda Gil 2003) para señalar el peso de la oralidad en la escritura de los mensajes analizados. Con el pasar del tiempo, los estudios han ido adoptando un enfoque que, a través de sintagmas como "conversación escrita" (Sanmartín Sáez 2007) o "interacción digital escrita" (Vela Delfa y Cantamutto 2016), se focalizan en la idea de interacción a través de la escritura. Se asume que la interacción dialógica es constante, gracias a dispositivos tecnológicos que mantienen los canales de comunicación permanentemente abiertos y que facilitan la intervención de expansiones futuras sobre un mismo texto (López Alonso 2006).

Desde la distinción entre medio y concepción y, a partir de la noción de variación concepcional, entendida como una escala gradual constituida por diferentes modalidades de empleo de la lengua, los análisis del español en la red han superado la dicotomía oral-escrito, ampliando el marco teórico en que se apoyan. En ellos se plantea que, como en la interacción cara a cara, en los intercambios comunicativos mediados por ordenador o por móviles, el grado de manifestación de la oralidad o la escritur(al)idad en el discurso está determinado principalmente por las condiciones de producción y recepción de los mensajes, más allá del canal gráfico o fónico que se utiliza. El contexto situacional, la intención comunicativa de los hablantes, las relaciones entre interlocutores, por ejemplo, permiten situar los distintos usos lingüísticos adoptados en la interacción dentro del *continuum* entre proximidad y distancia comunicativa propuesto por Koch y Oesterreicher (2007 [1990]). La perspectiva pluriparamétrica y multidimensional de este modelo contribuye a enriquecer el análisis de los discursos digitales teniendo en cuenta el grado de intimidad entre los interlocutores, su participación emocional en el acto comunicativo, la proximidad de las personas en relación con el *yo-aquí-ahora* de la enunciación, el grado de cooperación y de dialogicidad o posibilidad y frecuencia con la que los interlocutores asumen el papel de locutor, entre otros factores (Mancera Rueda y Pano Alamán 2013; Briz Gómez 2014; Estrada Arráez y de Benito Moreno 2016).

Las investigaciones que tienen en cuenta estos parámetros han demostrado que la mayor parte de los mensajes que circulan por la red se sitúa en el polo de la inmediatez comunicativa y dentro de la modalidad coloquial de la lengua. Así, los rasgos coloquiales se manifiestan en los usos ortográficos, léxicos y gramaticales en los mensajes de chat (Llisterri 2002; López Quero 2003; Sanmartín Sáez 2007) y de correo electrónico (Gómez Torrego 2001; López Alonso 2006), en la interacción en foros (Gouti 2006), blogs (Pano Alamán 2008), redes sociales (Mancera Rueda y Pano Alamán 2013) y *whatsapp* (Vela Delfa y Cantamutto 2016). En general, se señala la reproducción en el texto escrito de fenómenos prosódicos y fonéticos prototípicos de la modalidad oral coloquial, que presentan ya cierta sistematicidad como, por ejemplo, el alargamiento de vocales y el uso de puntos suspensivos; en el plano sintáctico, el recurso a la

brevedad y a la parataxis; y en el léxico, el predominio de neologismos, préstamos y términos y expresiones dialectales (Mancera Rueda y Pano Alamán 2013: 56). Como muestra de un estilo próximo de comunicar, se ha considerado también la inserción frecuente de emojis en los mensajes (Sampietro 2016), lo que obliga a adoptar una perspectiva multimodal.

2.1 Cuestión de tiempo(s)

El estudio de los usos lingüísticos en la red también ha traído consigo una reflexión sobre el medio digital y las transformaciones que ha supuesto en la percepción del tiempo de la interacción. En relación con este aspecto, gran parte de los estudios se ha apoyado en los conceptos de sincronía y asincronía. À partir de los años ochenta, las investigaciones sobre comunicación mediada por ordenador proponen un modelo teórico basado en la distinción entre sistemas síncronos como el chat y sistemas asíncronos como el correo electrónico y las listas de distribución. Herring (2007) explica cómo los sistemas asíncronos, aun reduciendo el tiempo de transmisión, no prevén la interacción directa entre emisor y destinatario, mientras que los síncronos requieren una copresencia mediada entre interlocutores. Sin negar la importancia del carácter síncrono o asíncrono del canal, propuestas más recientes han destacado que atender exclusivamente a la interfaz y a las características técnicas del medio, como su carácter síncrono o asíncrono, supone una limitación en el análisis de las dinámicas discursivas que tienen lugar en los entornos digitales (Pano Alamán 2008; Cantamutto y Vela Delfa 2016). Desde esta perspectiva se ha puesto de manifiesto la necesidad de considerar la progresiva integración de los canales, que lleva a situar parte de los intercambios en una dimensión de simultaneidad (Alcántara Plá 2017).

La comunicación digital se instala, pues, a menudo en el terreno de la inmediatez comunicativa. Esta es una tendencia que han acelerado aún más, si cabe, el desarrollo y expansión de las redes sociales y la comunicación a través de dispositivos móviles. Por ejemplo, a propósito de *Facebook*, Gurevich (2016) afirma que la constante actualización del estado en base a las propias experiencias garantiza un uso intensivo de la plataforma que integra los distintos tiempos de interacción en lo que llama "tiempo todo". Los tiempos a los que se refiere son el tiempo como registro de la vida o de la experiencia vital de cada usuario; el tiempo presente rector de la plataforma, que se manifiesta en "un *streaming* de publicaciones en vivo" (Gurevich 2016: 219); el tiempo contenido en las imágenes, en su consumo y recepción por parte de los usuarios; el tiempo de la interacción por medio de acciones como *gustar, comentar, compartir, etiquetar*, gracias a las cuales las publicaciones son perdurables; y, por último, el tiempo de la visibilidad/interacción.

Facebook, *Twitter*, *Instagram* o *whatsapp*, por mencionar las plataformas y sistemas más utilizados, urgen a relatar lo que acontece en el aquí y ahora más inmediato o a reccionar sin dilación a otro mensaje, mediante un *me gusta* o un emoji. Por otra parte, la convergencia de canales de comunicación digitales permite que la interacción sea ininterrumpida y que pueda llevarse a cabo de forma simultánea, rápida y discontinua desde un mismo dispositivo y desde pantallas reducidas, de modo que muchos de los contenidos que producimos en esos contextos se ajustan a formas hipermediatizadas y

pildorizadas (Slimovich 2016). Asistimos, en definitiva, a un predominio de enunciados breves, coloquiales y conversacionales.

Este cambio de perspectiva fomenta la aparición de investigaciones de análisis lingüístico del discurso digital en ámbitos como el del periodismo, la comunicación política o los lenguajes de especialidad. Algunos de los aspectos explorados en esta línea son los rasgos coloquializadores del discurso en blogs y comentarios de las versiones digitales de medios tradicionales (Mancera Rueda 2011; Sal Paz 2007), los marcadores discursivos y la dinámica conversacional en comentarios en la prensa digital (Moya Muñoz 2015; Pano Alamán 2015a), la función discursiva de las etiquetas dentro del ciberactivismo en Twitter (Menna 2015), las formas de tratamiento (Cautín Epifani 2015) o la coherencia textual en Facebook (Vela Delfa 2017), la brevedad en distintas tipologías de discurso digital (Vela Delfa y Cantamutto 2020), los encuadres frecuentes en lo que se ha llamado "discurso pseudopolítico" en las redes (Gallardo Paúls y Enguix Oliver 2016), los marcos léxico-semánticos de *tuits* electorales (Alcántara Plá *et al.* 2018) y la reconfiguración discursiva de géneros (blogs, páginas Web institucionales, reseñas) de discurso turístico en la red (Sanmartín Sáez 2012). Gran parte de estas investigaciones sugiere que la interacción en estos espacios contribuye a reforzar las relaciones interpersonales y sociales mediante la negociación colectiva del discurso y la adopción de determinadas estrategias pragmáticas por parte de los hablantes.

3 Enfoques pragmáticos

Los primeros estudios sobre comunicación digital de carácter pragmático afirmaban que el acceso a los entornos cognitivos de los interlocutores en la red era limitado o de menor calidad respecto al de los de una conversación prototípica cara a cara. Sin embargo, se ha demostrado que en la comunicación mediada, los interlocutores proporcionan pistas conversacionales e información contextual relevante para garantizar una comunicación eficaz (Yus 2001, 2010). Asimismo, la compenetración de los contextos virtual y real, como se han llamado tradicionalmente, obliga a los investigadores en este campo a considerarlos como complementarios. Desde este enfoque, se han realizado numerosos análisis pragmáticos y sociopragmáticos con una atención particular a las nociones de identidad, comunidad e imagen, y a las manifestaciones de la (des)cortesía en la red.

El fenómeno de la cortesía en la comunicación digital en lengua española cuenta con un número relevante de estudios. Cabe señalar, por ejemplo, el de Laborda Gil (2003) sobre la expresión de la cortesía en el correo electrónico; Alcoba Rueda (2004), sobre imagen y cortesía en páginas Web, blogs y chats; Mariottini (2006), que analiza el uso de los diminutivos como estrategia cortés en el chat en español; y López Quero (2007), sobre los marcadores discursivos *bueno*, *pues* y *pero* en relación con la cortesía en esta misma "conversación" virtual.

La descortesía o "ciberdescortesía" ha sido analizada prevalentemente en relación con las opiniones de los lectores en la prensa digital (Mancera Rueda 2009; Vigara Tauste y Hernández Toribio 2011) y la interacción en las redes sociales (Mancera Rueda 2015). El hecho de que estos sean contextos político-ideológicos en los que una noticia o una opinión ajena se somete a juicio público conlleva una mayor polarización de las

posiciones y una presencia preponderante de críticas negativas y actos descorteses. Sin embargo, se ha afirmado también que la descortesía en este tipo de plataformas permite al hablante no solo reforzar su propia imagen sino también la del grupo del que forma parte, lo que lleva a analizar la actividad de imagen cortés o descortés en estos espacios desde la visión de refuerzo de las relaciones sociales.

Por ejemplo, la agresión verbal a través de una estrategia formalmente cortés como es la ironía, en interacciones entre políticos y ciudadanos en Twitter, puede entenderse como una estrategia al servicio del realce de la propia imagen y de refuerzo de los vínculos con la comunidad discursiva con la que se comparte una posición política e ideológica similar (Pano Alamán 2015b). Asimismo, el *flaming*, entendido como la práctica de insultar o atacar de forma hostil a alguien en medios digitales, puede interpretarse incluso como un mecanismo para entretener a la audiencia imaginada, por ejemplo, en YouTube (Helfrich 2014). Como sugieren Kaul de Marlangeon y Cordisco (2014: 159), es posible analizar este tipo de interacciones, en particular las de contenido político, desde la descortesía de fustigación por "afiliación exacerbada" del hablante a su propio grupo ideológico y por "refractariedad" de aquél al grupo antagonista. En la primera, el hablante produce actividades de imagen en las que se percibe y es percibido como adepto a un grupo que asume su calidad de miembro con conciencia y orgullo, al punto que escoge la descortesía en su defensa. En la segunda, el individuo se percibe a sí mismo y es percibido como opositor al grupo, por ello critica, combate o agrede para expresar que está en una actitud refractaria respecto de aquello que suscita su oposición.

Este tipo de enfoques asumen que en la red los usuarios son participantes activos, críticos y portadores de opiniones convergentes o divergentes, que adoptan estrategias distintas de negociación del poder y de las relaciones sociales. Algunas de las estrategias puestas al servicio de estos objetivos son la atenuación, por ejemplo, en *Facebook*, *TripAdvisor* o *whatsapp* (Helfrich y Pano Alamán 2018), la ironía en los chats o en *Twitter* (Noblía 2004; Pano Alamán 2015b), la agresión verbal en los comentarios a las noticias (Mancera Rueda 2009) o, en el extremo contrario, los cumplidos en Instagram (Placencia 2019).

4 Aspectos metodológicos

En la investigación sobre comunicación digital en español reseñada se han adoptado distintas metodologías propias, por ejemplo, del análisis de la conversación (Sanmartín 2007), el análisis lingüístico del discurso (Mancera Rueda y Pano Alamán 2013), el estudio de los géneros digitales (López Alonso y Séré 2006), la sociolingüística y la lingüística de corte variacional (Estrada Arráez y de Benito Moreno 2016) y la sociopragmática (Fuentes Rodríguez y Placencia 2014). Por otra parte, estos análisis suelen combinar de forma frecuente métodos cuantitativos y cualitativos basados en corpus. En efecto,

> optan por construir corpus que permitan establecer resultados más apropiados, a diferencia de lo que ocurría en las primeras etapas, al menos hasta el 2006, caracterizadas por una tendencia a utilizar ejemplos aislados de interacciones

> mediadas por ordenador, provenientes de segundas fuentes (Pano Alamán y Moya Muñoz 2016: 13).

Se recogen muestras de mensajes de chat, SMS, correo electrónico, foros, comentarios en la prensa digital, entradas y comentarios en blogs, redes (*Twitter*, *Facebook*, *Tuenti*, *YouTube* e *Instagram*, principalmente) y aplicaciones de mensajería como *whatsapp*. Actualmente, la investigación se centra en los cambios que suponen los medios sociales (*social media*) en lo que concierne las prácticas comunicativas. Destacan, por ejemplo, los estudios que reflexionan sobre la actualización del estado en *Facebook* (Gurevich 2016); la función pragmática de la etiqueta en *Twitter* (Menna 2015); la práctica de comentar una noticia, un vídeo o una fotografía en espacios como *Instagram* (Placencia 2019) o *YouTube* (Helfrich 2014) y la gestión de turnos en *whatsapp* (Vela Delfa y Cantamutto 2016). De forma más o menos explícita, estos análisis señalan los límites que presentan en este ámbito los métodos aplicados generalmente al estudio de corpus de lengua escrita y hablada, en particular, de conversaciones orales coloquiales, si se tiene en cuenta que la producción de contenidos en los entornos digitales prevé el uso combinado de palabras clave y enunciados breves con enlaces hipertextuales, imágenes, vídeos, gif o memes que se comparten y se viralizan en pocos segundos, generando un discurso colectivo (Mancera Rueda y Pano Alamán 2013: 26), que puede discurrir en una misma plataforma o a través de distintos canales de forma simultánea.

Estos aspectos en constante evolución obligan al investigador a reflexionar sobre el tipo de dato (verbal, multimodal, hipertextual) que se analiza, los métodos de recolección del dato, la constitución de corpus, su análisis y divulgación. En este sentido, se habla de métodos "nativos digitales" que ponen el foco en los parámetros tecnológicos y sociosituacionales de los intercambios en entornos digitales, como son el formato de los mensajes, el carácter privado o público de la interacción, la estructura de la participación o las intenciones comunicativas de los usuarios (Pano Alamán y Moya Muñoz 2015: 123-124). Invitan, además, a superar la mera descripción de interfaces, considerando las "propiedades transversales de la comunicación digital", como indican Cantamutto y Vela Delfa (2016: 303), quienes plantean un modelo que permita describir y situar los discursos digitales en un *continuum* en base a los modos de realización de la comunicación, que puede ser más o menos monomodal/multimodal, breve/extensa o estática/dinámica; la enunciación, que se sitúa entre los polos continuo/discontinuo o transitivo/intransitivo, entre otros; y las relaciones interpersonales, articuladas de forma gradual también entre lo grupal/no grupal, lo público/privado o el carácter centrípeto/centrífugo de las redes en las que se interactúa. Estas reflexiones acompañan necesariamente también los estudios más recientes sobre los métodos de recolección y anotación de muestras de discurso digital.

4.1 Constitución y anotación de corpus

Construir un corpus de interacciones comunicativas que se originan en Internet puede parecer sencillo, en la medida en que el discurso ya se encuentra digitalizado, es accesible en línea y no requiere transcripción para su posterior análisis. Sin embargo, cuando se elabora un corpus de este tipo, enseguida surgen incompatibilidades entre los métodos asentados de la Lingüística de corpus y los datos recogidos en distintos medios digitales. Como para otro tipo de corpus, cabe interrogarse sobre la extensión

y la representatividad del conjunto de muestras, pero en este caso es necesario también distinguir los datos que genera el sistema –las líneas de entrada y salida de un canal de chat o la fecha de envío de un *tuit*–, de los que genera el usuario, por ejemplo, el texto con enlaces o emojis. Otro aspecto metodológico problemático concierne la posibilidad de conocer y extraer los datos sobre la identidad de los hablantes o su ubicación geográfica, que son esenciales en los estudios lingüísticos variacionales y que, de hecho, algunas plataformas como Twitter pueden ayudar a solventar (Estrada Arráez y de Benito Moreno 2016: 107). Por último, es necesario mencionar la peculiar ortografía que se observa en algunos tipos de discurso digital, un aspecto que plantea desafíos a la anotación lingüística (Mancera Rueda y Pano Alamán 2014: 309).

La mayor parte de los corpus existentes en varias lenguas aborda estas cuestiones aportando distintas soluciones en función de los objetivos de análisis o de los proyectos en los que se enmarcan [1]. En el ámbito del español, la mayor parte de las investigaciones que hemos mencionado se ha apoyado en corpus de distintas dimensiones creados en ocasiones manualmente y diseñados para realizar análisis específicos sobre interacciones en chats, blogs, foros, comentarios, reseñas digitales, redes y *whatsapp*, por lo que la objetividad y representatividad de los datos pueden ser cuestionables (Pano Alamán y Moya Muñoz 2016: 12-13). Asimismo, estos corpus no permiten aprehender las distintas "variedades" de discurso digital existentes ni cubrir, por ejemplo, la diversidad lingüística documentada en el mundo hispanohablante a través de otros corpus de español escrito y hablado.

En los últimos años, han surgido diversas propuestas para paliar estos problemas. Por ejemplo, Vela Delfa y Cantamutto (2015) plantean la creación de un repositorio abierto y colaborativo (Proyecto CÓDICE), cuyos objetivos son poner a disposición de la comunidad científica datos recopilados en investigaciones previas de discurso digital en español, establecer un modelo de descripción de los datos basado en su propuesta de descripción de propiedades, en función de los modos de realización, la enunciación y las relaciones interpersonales, y desarrollar estándares éticos para asegurar la anonimización de los datos personales relativos a las muestras.

Otro proyecto señalable es el de la plataforma Marco Polo de discurso político electoral español (Alcántara Plá *et al.* 2018), que permite a la comunidad científica y al público en general acceder a las coocurrencias de las palabras más frecuentes de un extenso corpus de *tuits* anotados morfológicamente, con la posibilidad de realizar búsquedas por palabras y comparar su frecuencia de uso mediante gráficos. Para la extracción automática de *tuits*, los investigadores emplearon la API (*Application Programming Interface*) de Twitter, que ha sido utilizada también en otros proyectos (Estrada Arráez y de Benito Moreno 2016) en los que se señala la ventaja del microblog respecto a otras redes para la extracción automática de datos [2]. En el proyecto Marco Polo se descargaron durante nueve meses los *tuits* y *retuits* que se publicaban en la cronología de los perfiles seleccionados. Sumando el número de *tuits* de los partidos, por un lado,

1. Destacan, entre otros: Sms4science, corpus multilingüe de sms; Birmingham Blog Corpus, de blogs en inglés; NUSSMSCorpus, con sms en inglés y en chino; DeRiK y CoMeRe, corpus de referencia de discurso digital para alemán y francés, respectivamente; y Web2Corpus*it, corpus de discurso digital en italiano.

2. Sobre la posibilidad de crear *script* con lenguajes de programación como R o Python para extraer *tuits* automáticamente y exportarlos en formatos fácilmente manejables, véanse las propuestas de Benito Moreno y Estrada Arráez (2018: 12-13) y Alcántara Plá *et al.* (2018: 11).

y de los políticos, por otro, se pudo garantizar la representatividad del corpus y equilibrar las diferencias entre los dos grupos de perfiles.

Sobre la cuestión de la representatividad de corpus de discurso digital, las propuestas existentes adoptan soluciones similares o bien tienden a obviar el problema si el número de ejemplos que componen el corpus supera el millón o si se realiza una compilación de datos "idealmente representativa de uno o varios medios virtuales" (de Benito Moreno y Estrada Arraéz 2018: 10) y se constituye un corpus cerrado que permita realizar cuantificaciones y análisis de fenómenos lingüísticos concretos (Mancera Rueda y Pano Alamán 2014: 307).

En lo que respecta al etiquetado de los corpus, la investigación se ha centrado en el nivel microtextual o lingüístico de los datos. Por medio de programas como Freeling, integrado en la herramienta de análisis textual *Sketch Engine* y aplicado también a la lematización y etiquetado de corpus de textos escritos y orales, es posible segmentar los textos de un corpus de discurso digital en palabras que se etiquetan después desde un punto de vista morfológico, utilizando expresiones regulares adaptadas, en este caso, al español. No obstante, etiquetar muestras de comunicación digital con programas como *Freeling* no garantiza que los resultados sean completamente válidos. Es necesario tener en cuenta que los mensajes que intercambiamos en la red contienen, como decíamos, texto escrito, menciones, audios y vídeos incrustados, URL, etiquetas, memes y emojis, además de presentar numerosos rasgos coloquializadores que pueden plantear problemas en el proceso de etiquetado automático, sabiendo que las herramientas que utilizamos están entrenadas sobre corpus de español normativo (Alcántara Plá *et al.* 2018: 15).

En este apartado, podemos mencionar también aquellos proyectos que se ocupan del análisis de la expresión de la opinión en el discurso digital y que, mediante programas como Lingmotif, analizan grandes corpus, por ejemplo, de reseñas de *Amazon* y de *tuits* (Moreno Ortiz 2019). El programa combina Freeling para el análisis morfosintáctico y la herramienta Sentitext, que está integrada por diccionarios y por un método de asignación de polaridad semántica. Estas herramientas permiten llevar a cabo anotaciones léxico-semánticas sofisticadas, sin embargo, la dificultad de señalar la ambigüedad de algunas palabras o expresiones o el significado que aportan los elementos multimodales (véase el pulgar hacia arriba en las reseñas) limitan aún su alcance (Moreno Ortiz 2019: 54).

Por último, cabe señalar que muy pocos estudios de discurso digital en español establecen una distinción clara entre los niveles de descripción microtextual, de los datos propiamente lingüísticos, y macrotextual, referido a la estructura de los mensajes. Si se asume que los parámetros contextuales, sociosituacionales y tecnológicos determinan en gran parte las interacciones en línea, parece necesario anotar metadatos como nombre o apodo del autor del texto, la fecha y hora de publicación, la extensión, si contiene datos multimodales, número de participantes, la pertenencia de una serie de mensajes a un intercambio, etc. En esta línea se sitúa CorpusRedEs, una propuesta de creación de un corpus de discurso digital en español con una vocación general (Pano Alamán y Muñoz Moya 2015), que aporta reflexiones sobre la anotación de estos datos utilizando el estándar de codificación de textos XML-TEI (*Text Encoding*

Initiative), aplicado ya en otros proyectos europeos [3]. Dichos proyectos se orientan hoy en día a mejorar el estándar XML-TEI proponiendo etiquetas y descriptores específicos de comunicación digital válidos para cualquier lengua. Por ejemplo, Beißwenger *et al.* (2012) proponen un modelo de etiquetado de la macroestructura de mensajes en alemán que se relaciona con distintos "modos sociotécnicos" de comunicar en la red (chat, wiki, *tuits*). Para anotar un conjunto de mensajes publicados en un canal de chat, en un blog o en redes sociales establecen una diferencia entre el texto plano, que puede incluir enlaces (URL, *hashtag*), y el archivo de audio, vídeo, fotografía y otros elementos incrustados en el mensaje, asumiendo que este tipo de elementos son predominantes en el discurso digital actual y no pueden ignorarse en el proceso de anotación formal.

En cuanto a las unidades de análisis, Beißwenger *et al.* (2012) proponen anotar el mensaje que llega en bloque a un servidor y que se publica en una plataforma, utilizando el elemento <post> [4]. Como se apuntaba en Pano Alamán y Muñoz Moya (2015: 125), este elemento parece versátil, dado que se basa en un sistema de anotación estándar y facilita la anotación de muestras de discurso digital desde el punto de vista de la estructura de la interacción, utilizando atributos como @who, @when y @from, que remiten al autor del mensaje, la fecha y hora de publicación y el remitente. El elemento <post> se entiende en esta propuesta bien como unidad monológica, aislada, bien como unidad dialógica, cuando constituye una respuesta o una reacción a otro <post>, aspecto que tiene en cuenta para la descripción de la organización textual, la dinámica conversacional observada en gran parte de los entornos digitales.

Sin embargo, antes de aplicar a este tipo de interacciones el modelo prototípico del análisis conversacional, en particular, en lo relativo a la alternancia de turnos, es necesario considerar el contexto comunicativo y los parámetros y propiedades que definen cada entorno y situación, el tiempo o tiempos de la interacción, y el tipo de relaciones interpersonales y sociales que se instauran en la red. Son necesarios, por tanto, elementos y atributos complejos que permitan, por ejemplo, indicar el distinto valor de las pausas en un mismo intercambio, ya que estas pueden responder a factores tecnológicos que hacen que el mensaje tarde en llegar, lo que puede interpretarse erróneamente como un silencio del destinatario; o pueden indicar una intención precisa por parte del hablante y, en ese caso, tener un significado pragmático que cabrá interpretar con la ayuda de otros elementos del contexto (puntos suspensivos, emojis, etc.). Por otro lado, si se considera que las "conversaciones" en los entornos digitales son simultáneas y de límites difusos (Alcántara Plá 2017: 313), parece adecuado revisar los atributos relacionados con el autor (@who) o con las relaciones temporales o espaciales (@next o @prev) que pueden establecerse entre los mensajes dentro de un mismo entorno o en entornos distintos pero integrados.

3. La información sobre estos proyectos está disponible en https://wiki.tei-c.org/index.php/SIG:Computer-Mediated_Communication.

4. La definición completa y los atributos de este elemento están disponibles en https://www1.ids-mannheim.de/fileadmin/kl/cmc-core/cmc-core_short.html\#TEI.post.

4.2 Aspectos éticos

Por último, desde un punto de vista metodológico, es necesario ampliar la reflexión sobre los estándares éticos de recogida y fijación de muestras teniendo en cuenta las diferencias entre extraer, almacenar, gestionar y publicar los datos con fines académicos o a través de medios de divulgación (i.e. Webs de proyectos, medios de comunicación). Los estudios mencionados hasta ahora plantean algunas propuestas válidas en este sentido, aunque se trata de soluciones aplicadas para cada proyecto, que dificultan la adopción de estándares.

Asumiendo que los límites entre público y privado se difuminan en estos contextos, los análisis publicados presentan distintas soluciones en función del propio diseño de investigación. Los estudios sobre interacciones entre internautas en plataformas de acceso abierto y público como *Twitter*, blogs, foros o prensa digital, suelen publicar los mensajes total o parcialmente, eliminando las referencias a los autores, sin solicitar autorización previa o bien solicitándola tanto a los autores del mensaje como a los propietarios de las plataformas, por ejemplo, en foros de debate y comentarios a las noticias (Pano Alamán 2008). En otros estudios, basados en la observación participante en redes de acceso restringido o en intercambios entre privados, como el sistema de mensajería de *Facebook* o la aplicación *whatsapp*, se solicita el consentimiento informado y se anonimizan los datos personales (Vela Delfa 2017) para proteger la privacidad de los participantes, en especial, si son, por ejemplo, menores que interactúan en plataformas como *Tuenti* (Mancera Rueda y Pano Alamán 2013: 309) o la más actual *TikTok*. Por último, en investigaciones de discurso institucional en redes sociales, los datos sobre el perfil y la identidad de los hablantes (políticos, periodistas, instituciones) suelen mantenerse, puesto que se trata de personas que desempeñan actividades públicas. Saber quiénes son y cómo interactúan en los entornos digitales es sin duda relevante de cara al análisis (Pano Alamán 2015b; Gallardo Paúls y Enguix Oliver 2016).

En definitiva, buena parte de las investigaciones se apoyan en corpus cerrados, con objetivos de análisis específicos y con una difusión circunscrita al ámbito académico, por lo que adaptan algunos de los estándares éticos vigentes en la recolección de datos de español escrito y oral. Otros proyectos van más allá. Marco Polo, por ejemplo, considera las restricciones de la Ley de Protección de Datos y las normas que impone Twitter sobre la redistribución de los *tuits* para diseñar una plataforma accesible en línea y divulgar parte de los datos recogidos y analizados (Alcántara Plá *et al.* 2018: 9). Es necesario recordar, en este sentido, que la legislación europea sobre recogida y gestión de datos (Reglamento General de Protección de Datos 2016/679, 27 de abril de 2016) vincula hoy en día los proyectos basados en corpus de comunicación digital, especialmente, en redes sociales, para los cuales existen ya listas de buenas prácticas (Collins 2019). Los desafíos a los que cabe dar respuesta pasan por definir los criterios de selección de los sujetos participantes; las formas de consentimiento informado adaptadas a las distintas formas de privacidad en un mundo de tecnologías cada vez más difusas y ubicuas; el anonimato y confidencialidad de los datos, teniendo en cuenta los términos y condiciones de uso de las plataformas y aplicaciones; la integridad de

los datos; y los problemas que plantea su procesamiento, almacenamiento y difusión a través de la nube o en servidores remotos.

5 A modo de conclusión

El presente trabajo ha intentado ofrecer un panorama actualizado sobre los principales enfoques teórico-metodológicos adoptados en las investigaciones sobre comunicación digital en español aparecidas en los últimos años. Los análisis atienden, en un primer momento, a los aspectos lingüísticos, en particular, a la expresión de la modalidad oral coloquial de la lengua en chats, correo electrónico y foros de debate, para pasar, más tarde, al análisis discursivo y pragmático de las interacciones a través de mensajería instantánea y redes sociales. Los enfoques y las metodologías, cualitativas y cuantitativas y basadas en corpus, tienden hoy en día a converger poniendo en evidencia algunos de los desafíos que se presentan a las futuras investigaciones sobre discurso digital en español. Algunos se relacionan con la definición del contexto, de unidad de análisis y de parámetros que permitan situar las interacciones dentro de determinados géneros textuales y tipologías discursivas y que lleven a superar las dicotomías escrito-hablado y síncrono-asíncrono, demasiado ligadas al medio. Se ha señalado asimismo la necesidad de constituir corpus de discurso digital representativos considerando los procesos de recogida, almacenamiento, creación, anotación y difusión de las muestras y los aspectos éticos de cada fase, en particular, en un momento en el que el análisis de la comunicación digital se incluye en proyectos de minería de datos e inteligencia artificial.

Referencias

Alcántara Plá, M. (2017). "El contexto de los mensajes en la comunicación digital". En P. Gubitosi, M. Giammatteo, A. Parini (éds.) *El español en la red*, Madrid- Frankfurt: Iberoamericana-Vervuert, 303-326.

Alcántara Plá, M., A. Ruiz Sánchez, M. Benito Rey, V. Amessa García, A. Martín Jimeno (2018). "Marco-Polo: una plataforma digital de análisis del discurso electoral en Twitter". *REDD: Revista de estudios del discurso digital* 1: 7-29.

Alcoba Rueda, S. (2004). "Cortesía e imagen en la lengua en Internet". En D. Bravo y A. Briz (éds.) *Pragmática sociocultural: estudios sobre el discurso de cortesía en español*, Barcelona: Ariel, 357-370.

Beißwenger, M., Ermakova, M., Geyken, A., Lemnitzer, L. y Storrer, A. (2012). "A TEI Schema for the Representation of Computer-mediated Communication". *Journal of the Text Encoding Initiative* (3) http://jtei.revues.org/476

Benito Moreno, C. de y A. Estrada Arráez (éds.) (2016). "Introducción. Si Alvar levantara la cabeza: cómo tratar con informantes virtuales", *RILI*, 2(28): 7-17.

Briz Gómez, A. (2014). "Hablar electrónicamente por escrito", *CHIMERA: Romance Corpora and Linguistic Studies* 1: 77-89. https://revistas.uam.es/index.php/chimera/article/view/255/243

Cantamutto, L. y C. Vela Delfa (2016). "El discurso digital como objeto de estudio: de la descripción de interfaces a la definición de propiedades", *Aposta: Revista de ciencias sociales*, 69: 296-323 http://www.apostadigital.com/revistav3/hemeroteca/cvela2.pdf

Cantamutto, L. y C. Vela Delfa (2020). "Mensajes, publicaciones, comentarios y otros textos breves de la comunicación digital", *Tonos digital: Revista de estudios filológicos*, 38 https://digitum.um.es/digitum/bitstream/10201/86441/1/2394-6691-1-PB.pdf

Collins, L. (2019). *Corpus Linguistics for Online Communication.* New York: Routledge.

Estrada Arráez, A. y C. de Benito Moreno (2016). "Variación en las redes sociales: datos twilectales", *RILI*, 8(28), 77-111.

Gallardo Paúls, B. y S. Enguix Oliver (2016). *Pseudopolítica. El discurso político en las redes sociales*, Valencia: Universitat de València.

Giammatteo, M., Gubitosi, P. y Parini, A. (2017). *El español en la red*. Madrid-Frankfurt: Iberoamericana-Vervuert.

Gómez Torrego, L. (2001). "La gramática en Internet", *Lengua y escritura en Internet* http://www.congresosdelalengua.es/valladolid/ponencias/nuevas_fronteras_del_espanol/4_lengua_y_escritura/default.htm

Gouti, G. (2006). "Intercambios en los foros de debate: elementos de reflexión para un acercamiento lingüístico", *Estudios de Lingüística del Español*, 24 http://elies.rediris.es/elies24/gouti.htm

Gurevich, A. (2016). "El tiempo todo en Facebook", *Aposta. Revista de Ciencias Sociales*, 69: 217-238 http://apostadigital.com/revistav3/hemeroteca/gurevich.pdf

Helfrich, U. (2014). "Face-work and flaming in social media". En K. Bedijs, G. Held y C. Maa β (éds.) *Face work and social media*, Münster: LIT-Verlag, 297-321.

Helfrich, U. y Pano Alamán, A. (éds.) (2018) La atenuación en los discurso digitales, Círculo de Lingüística Aplicada a la Comunicación, 73: 5-14. https://revistas.ucm.es/index.php/CLAC/article/view/59056

Herring, S. (2007). "A faceted classification scheme for computer-mediated discourse", *Language@Internet*, 4 http://www.languageatInternet.org/articles/2007/761

Kaul de Marlangeon, S. y A. Cordisco (2014). "La descortesía verbal en el contexto político-ideológico de las redes sociales", *Revista de Filología de la Universidad de la Laguna* 32: 145-162.

Koch, P. y W. Oesterreicher (2007 [1990]). *Lengua hablada en la Romania: español, francés, italiano*, Madrid: Gredos.

Laborda Gil, X. (2003). "Estilo y cortesía en el correo electrónico", *Tonos digital: Revista electrónica de estudios filológicos*, 6. http://www.um.es/tonosdigital/znum6/estudios/Laborda.htm

Llisterri, J. (2002). "Marcas fonéticas de la oralidad en la lengua de los chats: elisiones y epéntesis consonánticas". En R. Almela Pérez (éd.) *Lengua española e Internet. Revista de Investigación Lingüística*, vol. 2, 61-100.

López Alonso C. (2006). "El correo electrónico", *Estudios de Lingüística del Español*, 24 http://elies.rediris.es/elies24/lopezalonso.htm

López Alonso, C. y A. Seré (2006). "Nuevos géneros discursivos, los textos electrónicos", *Estudios de Lingüística del Español*, 24 http://elies.rediris.es/elies24/index.htm

López Quero, S. (2003). *El lenguaje de los chats. Aspectos gramaticales*, Granada: Port-Royal.

López Quero, S. (2007). "Marcadores discursivos y cortesía en la conversación virtual en español", *Language Design* 9: 93-112.

Mancera Rueda, A. (2009). "Manifestaciones de descortesía y violencia verbal en los foros de opinión digitales de los diarios españoles", *Discurso&Sociedad*, 3(3): 437-466 http://www.dissoc.org/ediciones/v03n03/DS3%283%29Mancera.pdf

Mancera Rueda, A. (2011). *¿Cómo se "habla" en los cibermedios? El español coloquial en el periodismo digital*, Bern: Peter Lang.

Mancera Rueda, A. (2015). "Los estudios sobre (des)cortesía y actividades de imagen en las redes sociales", *Textos En Proceso*, 1(1): 50-70. https://www.asice.se/index.php/tep/article/view/19

Mancera Rueda, A. y A. Pano Alamán (2013). *El español coloquial en las redes sociales*, Madrid: Arco/Libros.

Mancera Rueda, A. y A. Pano Alamán (2014). "Las redes sociales como corpus de estudio para el análisis del discurso mediado por ordenador", *Humanidades Digitales: desafíos, logros y perspectivas – Janus*, 1: 305-315

Mariottini, L. (2006). "Los diminutivos en los chats: estrategias de cortesía lingüística", *La Cortesía. Cultura, lenguaje y representación* 3: 103-131.

Menna, L. (2012). "Nuevas formas de significación en red: el uso de las #etiquetas en el movimiento 15M", *Estudios de Lingüística del Español*, 34. http://elies.rediris.es/elies34/Tesina_L-Menna.pdf

Moreno Ortiz, A. (2019). "Mi opinión cuenta: la expresión del sentimiento en la red". En Robles Ávila, S. y A. Moreno Ortiz (éds.) *Comunicación mediada por ordenador*, Madrid: Cátedra, 38-74.

Moya Muñoz, P. (2015). "Los comentarios de los usuarios en la prensa digital: una propuesta para su estudio desde el discurso mediado por ordenador y los estudios periodísticos". *Caracteres* 4 (1): 178-199.

Noblía, V. (2004). "La ironía en los chats: una forma de preservar las imágenes y administrar los conflictos en las conversaciones mediadas por computadoras". En D. Bravo y A. Briz (éds.) *Pragmática sociocultural: estudios sobre el discurso de cortesía en español*, Barcelona: Ariel, 371-397.

Pano Alamán, A. (2008). *Dialogar en la Red. La lengua española en chats, e-mail, foros y blogs*, Bern: Peter Lang.

Pano Alamán, A. (2015a). "*Aproximación a los marcadores del discurso en los comentarios de la prensa digital española y argentina", Revista de Investigaciones lingüísticas y literarias hispanoamericanas*, 20: 89-114.

Pano Alamán, A. (2015b). "Ironía verbal y actividad de imagen en el discurso de políticos y ciudadanos españoles en Twitter", *Pragmática Sociocultural*, 3(1): 59-89.

Pano Alamán, A. y P. Moya Muñoz (2015). "CorpusRedEs. Proyecto de creación y anotación de un corpus de comunicación mediada por ordenador en español", *CHIMERA. Romance Corpora and Linguistic Studies*, 2: 117-129. https://revistas.uam.es/index.php/chimera/article/view/1042

Pano Alamán, A. y Moya Muñoz (2016). "Una aproximación a los estudios sobre el discurso mediado por ordenador en lengua española", *Tonos Digital. Revista Electrónica de Estudios Filológicos*, 30: 1-30. https://digitum.um.es/xmlui/handle/10201/47991

Placencia, M. E. y C. Fuentes Rodríguez (éds.) (2014), *Discurso e identidad en el ciberespacio hispano*, *Discurso&Sociedad*, 8(3). http://www.dissoc.org/ediciones/v08n03/

Placencia, M. E. (2019). "*Qué perfección*: Complimenting Behaviour among Ecuadorian Teenage Girls on Instagram". En E. Ogiermann y P. Garcés-Conejos Blitvich (éds.) *From Speech Acts to Lay Understandings of Politeness*, Cambridge: Cambridge University Press, 93-116.

Robles Ávila, S. y A. Moreno Ortiz (éds.) (2019). *Comunicación mediada por ordenador. La lengua, el discurso y la imagen*, Madrid: Cátedra.

Sal Paz, J. (2007). "Rasgos de oralidad en los foros de discusión de la prensa digital". En L. Granato de Grasso (éd.) *Diálogo y contexto*, La Plata: Universidad Nacional de La Plata, 555-570.

Sampietro, A. (2016). "Emoticonos y multimodalidad. El uso del pulgar hacia arriba en whatsapp", *Apos-ta: Revista de ciencias sociales*, n. 69: 271-295. http://www.apostadigital.com/revistav3/hemeroteca/asampietro.pdf

Sanmartín Sáez, J. (2007). *El chat. La conversación tecnológica*, Madrid: Arco/Libros.

Sanmartín Sáez, J. (éd.) (2012). *El discurso turístico en Internet*, Madrid-Frankfurt: Iberoamericana-Vervuert.

Slimovich, A. (2016). "La digitalización de la política y la vuelta de lo televisivo", *Revista de comunicación,* 15: 111-127.

Vela Delfa, C. (2017). "Coherencia, cohesión y estructura de la interacción en el discurso digital: un análisis de los intercambios en la red social Facebook". En P. Gubitosi, M. Giammatteo y A. Parini (éds.) *El español en la red*, Madrid-Frankfurt: Iberoamericana-Vervuert, 255-278.

Vela Delfa, C. y Cantamutto, L. (2015). "Methodological Approach to the Design of Digital Discourse Corpora in Spanish. Proposal of the CÓDICE Project", *Current Work in Corpus Linguistics (CILC2015)*, 494-499.

Vela Delfa, C. y L. Cantamutto (2016). "De participante a observador: El método etnográfico en el analisis de las interacciones digitales de whatsapp", *Tonos digital: Revista de estudios filológicos*, n. 31. https://digitum.um.es/digitum/handle/10201/50356

Vigara Tauste, A. M. e I. Hernández Toribio (2011). "Ciber(des)cortesía en los foros de opinión de la prensa escrita", *Estudios de Lingüística de la Universidad de Alicante* 25: 353-379.

Yus Ramos, F. (2001). *Ciberpragmática. El uso del lenguaje en Internet*, Barcelona: Ariel.

Yus Ramos, F. (2010). *Ciberpragmática 2.0*, Barcelona: Ariel.