

CORPUS DE CONVERSACIONES DE WHATSAPP: RECOLECCIÓN DE DATOS Y CONSIDERACIONES ÉTICAS

Cantamutto, Lucía
CIEDIS-CONICET/Universidad Nacional de Río Negro
lcantamutto@unrn.edu.ar

Vela Delfa, Cristina
Universidad de Valladolid
cveladelfa@gmail.com

Sociolingüística; Pragmática

El acceso a muestras reales de lengua es un paso fundamental para muchas investigaciones, principalmente, aquellas que, desde una perspectiva pragmática y sociolingüística, buscan analizar los usos lingüísticos contextualizados (Albelda Marco, 2021). En el Análisis del Discurso Digital (Thurlow y Mroczek, 2011; Bou-Franch, P. y Garcés Cornejo Blitvich, 2018), se observa una tendencia a recoger datos de redes sociales cuyas expectativas de circulación se acercan a la esfera pública, tales como Twitter/X o Facebook, y, en menor caso, la atención a las interacciones privadas e íntimas que ocurren a través de diferentes aplicaciones de mensajería instantánea (Pano Alamán y Moya Muñoz, 2015; Vela Delfa y Cantamutto, 2016; Cantamutto y Vela Delfa, 2023). El objetivo del presente trabajo es doble. Por un lado, presentamos una sistematización de las técnicas de recolección de datos de WhatsApp que incluye los tipos de datos que se recuperan en cada caso. Por otro lado, expondremos las consideraciones éticas que el estudio de estas piezas lingüísticas requiere. Respecto al primer punto, atenderemos a las consecuencias metodológicas de optar entre exportar chat, realizar capturas, usar el reenvío, o ser participante-observador. Sobre el segundo objetivo, atenderemos a los aspectos a considerar en la anonimización de los datos y resguardo de la identidad de los participantes, los consentimientos informados y la trazabilidad de los datos. A partir de una revisión sobre los conjuntos de datos utilizados por investigaciones previas, se evidencia un área de vacancia en torno a la constitución de corpus amplios, sistemáticos y de consulta general sobre interacciones digitales en lengua española. Además, estos datos tampoco tienen presencia en los corpus de referencia del español. La combinación de técnicas de recolección ha mostrado ser la opción más segura para recoger los datos digitales y atender a la complejidad que estos tienen. En esta línea, se plantea la importancia de realizar el tratamiento de los datos a los fines de garantizar la anonimidad de los participantes, así como las posibilidades de su explotación a través de programas de análisis lingüísticos, como Antconc.

- Alamán, A. P., & Muñoz, P. M. (2015). CorpusRedEs. Proyecto de creación y anotación de un corpus de comunicación mediada por ordenador en español. *CHIMERA: Revista de Corpus de Lenguas Romances y Estudios Lingüísticos*, 2, 117-129.
- Albelda Marco, M. (2022). Los corpus del español hablado y los estudios pragmáticos. En: Rojo, G. *Lingüística de corpus en español/The Routledge Handbook of Spanish Corpus Linguistics* (pp. 223-238). Routledge.
- Bou-Franch, P., & Garcés Cornejo Blitvich, P. (Eds.). (2018). *Analyzing digital discourse: New insights and future directions*. Springer.
- Cantamutto, L., & Vela Delfa, C. (2023). Corpus de interacciones digitales: sistematización de técnicas para recoger datos en WhatsApp. *Cuadernos. info*, (54), 117-139.
- Thurlow, C., & Mroczek, K. (Eds.). (2011). *Digital discourse: Language in the new media*. Oxford University Press.
- Vela Delfa, C. & Cantamutto, L. (2016). De participante a observador: El método etnográfico en el análisis de las interacciones digitales de whatsapp. *Tonos digital: revista de estudios filológicos*, (31), 34.