



Cuadernos de Lingüística Hispánica

ISSN: 0121-053X

Universidad Pedagógica y Tecnológica de Colombia
(UPTC)

Bonilla Neira, Laura Cristina

Claves para analizar datos en Twitter. Recolección y procesamiento de corpus*

Cuadernos de Lingüística Hispánica, núm. 39, 2022, pp. 1-21

Universidad Pedagógica y Tecnológica de Colombia (UPTC)

DOI: <https://doi.org/10.19053/0121053X.n39.2022.14283>

Disponible en: <https://www.redalyc.org/articulo.oa?id=322275370006>

- Cómo citar el artículo
- Número completo
- Más información del artículo
- Página de la revista en redalyc.org

UPEM
redalyc.org

Sistema de Información Científica Redalyc


Red de Revistas Científicas de América Latina y el Caribe, España y Portugal
Proyecto académico sin fines de lucro, desarrollado bajo la iniciativa de acceso
abierto

Claves para analizar datos en Twitter. Recolección y procesamiento de corpus*


LAURA CRISTINA BONILLA NEIRA**

Recepción: 15 de mayo de 2022
Aprobación: 13 de junio de 2022

Forma de citar este artículo: Bonilla, L. (2022). Claves para analizar datos en Twitter. Recolección y procesamiento de corpus. *Cuadernos de Lingüística Hispánica*, (39), e14283

 <https://doi.org/10.19053/0121053X.n39.2022.14283>

* Artículo de investigación.

** Licenciada en Español y Literatura, y magíster en Semiótica de la Universidad Industrial de Santander. Candidata a doctora en Lingüística de la Facultad de Filosofía y Letras de la Universidad de Buenos Aires. Becaria doctoral del Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), Argentina. laura.bonilla@uba.ar
 <https://orcid.org/0000-0003-3417-8504>

Resumen

El objetivo de este trabajo es presentar una propuesta metodológica para el análisis de datos de Twitter con un enfoque mixto. Específicamente, el procedimiento de recolección y procesamiento de la información se caracteriza por retomar recursos cualitativos y cuantitativos, así como por la construcción de un corpus manejable para un posterior análisis cualitativo. El procedimiento para abordar los discursos digitales de Twitter consiste en: 1) registro de la etnografía virtual, 2) recolección de los datos por medio de la API de Twitter usando Python; 3) visualización y filtrado de los datos con Open Refine; 4) construcción del corpus 5) categorización y etiquetado de los enunciados verbo-icónicos con Atlas.ti. El trabajo reconstruye el recorrido metodológico llevado a cabo en una investigación doctoral en curso con enfoque cualitativo, de la cual se extraen los ejemplos, con el fin de ofrecer una ruta accesible que pueda ser replicada en investigaciones con este tipo de datos.

Palabras clave: metodología, discurso digital, Twitter, construcción de corpus, análisis del discurso, netnografía.

Keys for Analyzing Data on Twitter. Corpus Collection and Processing

Abstract

This paper aims to present a methodological proposal for the analysis of Twitter data from a mixed approach. Specifically, the procedure for collecting and processing information is characterized by the use of qualitative and quantitative resources, as well as the construction of a manageable corpus for subsequent qualitative analysis. The procedure to approach Twitter digital discourses consists of the following route: 1) virtual ethnography registration; 2) data collection through the Twitter API using Python; 3) visualisation and filtering of the data with Open Refine; 4) corpus construction; 5) categorisation and labelling of the iconic verb utterances with Atlas.ti. The paper reconstructs the methodological path carried out in ongoing doctoral research with a qualitative approach, from which the examples are extracted, to offer an accessible route that can be replicated in research with this type of data.

Keywords: methodology, digital discourse, Twitter, corpus construction, discourse analysis, netnography.

Clés de l'analyse des données sur Twitter. Collecte et traitement du corpus

Résumé

L'objectif de cet article est de présenter une proposition méthodologique pour l'analyse des données Twitter avec une approche mixte. Plus précisément, la procédure de collecte et de traitement des données est caractérisée par la réutilisation de ressources qualitatives et quantitatives, ainsi que par la construction d'un corpus gérable pour une analyse qualitative ultérieure. La procédure pour aborder des discours numériques de Twitter consiste à : 1) enregistrement de l'ethnographie virtuelle, 2) collecte des données au moyen de l'API Twitter en utilisant Python ; 3) visualisation et filtrage des données avec Open Refine ; 4) construction du corpus 5) catégorisation et étiquetage des énoncés verbo-iconiques avec Atlas.ti. L'article reconstruit le parcours méthodologique réalisé dans une recherche doctorale en cours avec une approche qualitative, d'où sont extraits les exemples, afin d'en offrir un parcours accessible et rééritable dans les recherches avec ce type de données.

Mots clés: méthodologie, discours numérique, Twitter, construction de corpus, analyse du discours, netnographie.

Chaves para analisar dados no Twitter. Coleta e processamento de corpus

Resumo

O objetivo deste trabalho é apresentar uma proposta metodológica para a análise de dados do Twitter com uma abordagem mista. Especificamente, o procedimento de coleta e tratamento da informação caracteriza-se pela utilização de recursos qualitativos e quantitativos, bem como pela construção de um corpus gerenciável para posterior análise qualitativa. O procedimento de abordagem dos discursos digitais do Twitter consiste em: 1) registro da etnografia virtual; 2) coleta de dados por meio da API do Twitter utilizando Python; 3) visualização e filtragem de dados com Open Refine; 4) construção do corpus 5) categorização e rotulagem dos enunciados verbo-icônicos com o Atlas.ti. O trabalho reconstrói o percurso metodológico realizado em uma pesquisa de doutorado em andamento com abordagem qualitativa, de onde são extraídos os exemplos, a fim de oferecer um percurso acessível que possa ser replicado em pesquisas com esse tipo de dado.

Palavras-chave: metodologia, discurso digital, Twitter, construção de corpus, análise de discurso, netnografia.

Introducción

Con la llegada de las redes sociales, el debate público ha ampliado sus espacios de interacción. Cada día más personas se conectan con el acontecer nacional e internacional a través de Internet y, particularmente, de las redes sociales. Específicamente, la red social Twitter se ha convertido en un espacio de información y discusión entre internautas que permite enviar y comentar mensajes, reaccionar y encontrar las tendencias de temas de interés.

Twitter ha llegado a ser, además, la red de preferencia de los políticos, donde informan sobre sus agendas y comunican sus posturas sobre determinado tema. Fue lo que sucedió inicialmente con Obama en su disruptiva campaña presidencial de 2008 (Beas, 2010) y más tarde con la predilección de Trump de reemplazar los comunicados de prensa por tuits (Kreis, 2017; Ott, 2017). También los medios de comunicación y periodistas encuentran en Twitter un espacio para masificar sus informaciones (Peirano, 2019) e interactuar con sus audiencias. De modo similar, el activismo social y político ha encontrado un asidero fecundo para manifestar sus ideas (Rovira, 2019) y conectar con otros. Este fenómeno se debe, en parte, a la arquitectura interna de esta red social —en la cual la brevedad, la inmediatez y la viralización (Ventura, 2021) son sus principales características—, así como a la atención que reciben sus contenidos en los medios de comunicación tradicionales.

En ese sentido, Twitter se ha consolidado como una red de intercambio y debate público, que amplía el espacio de participación ciudadana agregando un espacio público digital (Seixas & Nascimento, 2021). Es decir, no se trata de un reemplazo de la plaza pública por el espacio digital, sino de un ensanchamiento del lugar de discusión ciudadana. A este fenómeno se suma el acceso abierto de los datos, a diferencia de Facebook, Instagram o Tiktok, lo que ha influido en que Twitter se convierta en un objeto de estudio para las ciencias sociales en general y particularmente para las ciencias políticas y de la comunicación (Longhi, 2013; Calvo, 2015; Hürlimann et al., 2016; Pallarés & García, 2017). No obstante, muchos de los trabajos realizados desde estas disciplinas tienen un enfoque cuantitativo, por lo que hacen falta más acercamientos desde una perspectiva que incorpore herramientas también cualitativas que podrían enriquecer análisis posteriores de los datos.

En esta línea, el presente artículo se interroga sobre la forma de abordar los enunciados verbo-visuales de Twitter para ser utilizados en posteriores análisis discursivos. Se presenta una propuesta metodológica que orienta la forma de capturar los datos, filtrarlos, organizarlos y construir un corpus manejable susceptible de análisis. Esta alternativa analítica busca aportar al estudio actual del discurso digital, en particular del material extraído de Twitter, y a enriquecer el conocimiento sobre la discursividad política en esta red social.

Este trabajo se organiza del siguiente modo: en primer lugar, se caracteriza brevemente la plataforma Twitter y su importancia como objeto de estudio; en segundo lugar, se desarrolla la propuesta analítica a partir de cinco puntos:

netnografía, aspectos éticos, procedimiento de recolección y procesamiento de los datos, construcción del corpus y análisis descriptivo. Esta ruta metodológica deriva de la reconstrucción del proceso llevado a cabo en una investigación en curso¹, por lo cual los ejemplos de cada paso serán tomados de esta. La investigación aborda los discursos en Twitter de grupos a favor y en contra del Acuerdo de Paz en Colombia en el plebiscito del 2016.

Twitter como objeto de estudio

Twitter, la plataforma de *microblogging* fundada hace más de 15 años (Twitter, 2020), es en la actualidad una de las principales redes sociales. Inspirada en el auge de los blogs de principios de siglo XXI y en los mensajes de texto (SMS), Twitter irrumpe en la web 2.0 para condensar estas dos ideas y ofrecerles a los usuarios una plataforma de promoción personal (Van Dijck, 2016) y, a la vez, de intercambio comunitario. Básicamente, se trata de una red social que permite enviar mensajes de texto plano de 280 caracteres (antes 140) denominados tuits (*tweets* en inglés) que luego se muestran en la página principal del usuario. Con el tiempo, Twitter ha evolucionado sus formatos de interacción y hoy se pueden compartir también imágenes estáticas o animadas (*gifs*), ideogramas (emojis), videos, insertar enlaces de otras plataformas y hasta establecer conversaciones en directo (*spaces*). A pesar de que otras redes como Facebook o YouTube cuenten con más usuarios (DataReportal, 2022), Twitter, con sus más de 400 millones usuarios², se destaca por ser la red más popular de las personalidades políticas y los medios de comunicación y por permitir a los usuarios el seguimiento de las noticias en tiempo real. Esto se debe particularmente a la arquitectura de esta red social, que antepone la información a la formación de redes (Alcántara-Plá, 2020).

Pese a este diseño de la plataforma, donde prima la jerarquía de tuiteros influyentes que generalmente son figuras del espectáculo, políticos o celebridades (Van Dijck, 2016), varios analistas también consideran que en Twitter se encuentra un espacio de participación ciudadana, una especie de “foro político, social y cultural que se mantiene abierto 24 horas al día, los 365 días del año” (Calvo, 2015, pp. 11-12.). Es así como Twitter se configura como un espacio de conversación social en torno a temas de interés y en donde los políticos y otras personalidades encuentran un lugar sin intermediarios para comunicarse con el público. En ese sentido, la plataforma permite “una mayor intervención de los ciudadanos, lo cual ayuda a los movimientos sociales y a las políticas alternativas” (Castells, 2008, p. 1). Freire (2019) sostiene que Twitter se puede entender como un territorio político digital, en cuanto que, además de un despliegue de la acción política, también reúne características distintivas de otras redes sociales, tales como “el espíritu de publicidad, el potencial interactivo

1 Algunos avances se pueden encontrar en Bonilla-Neira (2021, 2022).

2 La plataforma datareportal.com estima que Twitter alcanza los 436 millones de usuarios a enero de 2022. Facebook con 2900 millones y YouTube con 2500 millones lideran las estadísticas con mayor número de usuarios activos en el ámbito mundial.

de la bidireccionalidad, la masividad de la conversación pública, el sesgo positivo, la presencia de capitales simbólicos y el poder de colonización” (p. 68). De este modo, Twitter no es reconocido por el número de usuarios activos, sino por la capacidad que tienen sus mensajes (tuits) de generar tendencias e impactar así en los medios de comunicación masiva.

Desde una perspectiva discursiva, se puede decir que Twitter es un espacio público digital, entendido como “un lugar de interacción discursiva, en que es posible analizar la producción, la circulación y recepción de discursos en que los sujetos online buscan participar, de algún modo, de la vida pública, y más específicamente política”³ (Seixas & Nascimento, 2021, p. 2402). Con lo cual, esta red social permite el ensanchamiento del espacio público (Castells, 2009; Slimovich, 2016; Rovira, 2019) en el ámbito virtual, no como un reemplazo de la plaza pública o del encuentro físico, sino que se trata de un entorno ampliado, puesto que en el espacio virtual también se encuentran las personas ocupando ese lugar. Esto coincide con el carácter conversacional y colaborativo que se ha destacado en las primeras aproximaciones analíticas de Twitter (Honeycutt & Herring, 2009) y que incide en las dinámicas de la sociedad al propiciar la interacción pública de voces (Mancera Rueda & Pano Alamán, 2020). Esta característica enfatiza en el carácter dialógico que promueve la plataforma, dado que, si bien hay una primacía de la información, los usuarios participan con su opinión sobre algún tema, la comparten dando *retuit* a otros o manifiestan acuerdo a comentarios dando “me gusta”. Esta dinámica de comunicación suscita prácticas tecnodiscursivas (Paveau, 2017) que permiten la promoción de una comunidad definida por un *ambient affiliation* (Zappavigna, 2011, 2012), es decir, por medio de la vinculación a un determinado tema donde la comunidad encuentra adhesión con ciertos valores y opiniones.

En este trabajo se plantea el diseño de una investigación cualitativa que, desde la perspectiva de Vasilachis de Gialdino (2009), es de carácter interpretativa, multimetódica, reflexiva y usa diseños flexibles al contexto. Desde este enfoque, en el diseño se incluyen herramientas diversas que permiten un acercamiento al objeto de estudio. En primera instancia, se utiliza la netnografía como método de observación, la cual permite un acercamiento al tema de estudio y retratar de forma panorámica el objeto de análisis. En segunda instancia, se procede a la recolección de datos utilizando herramientas digitales de descarga, visualización y manejo de información tales como API (Application Programming Interface) de Twitter, librerías de Python y Open Refine; recursos que habilitan la exploración de tendencias y patrones en el material. Con los datos organizados, se procede a la construcción del corpus, para lo cual es necesaria la selección de perfiles de acuerdo con criterios propios de cada investigación. A continuación, se detalla cada uno de estos pasos para ahondar en esta propuesta metodológica.

3 Traducción propia.

Propuesta analítica

Etnografía virtual

Con el objetivo de realizar una inmersión inicial en algún tema que se discute en Twitter y explorar el material producido en esta red social, se propone un primer acercamiento a través de una netnografía (Del Fresno, 2011; Kozinets, 2015), que tiene antecedentes en la etnografía virtual (Hine, 2004), y en la ciberetnografía (Robinson & Schulz, 2014). De acuerdo con Hine (2004), la etnografía virtual se erige como un modo de conocer la realidad social que se presenta en internet. Tal como sucede en la etnografía clásica, su forma básica consiste en que el investigador se sumerja en el mundo que estudia por un tiempo determinado y tome nota de las relaciones, actividades y significaciones que encuentre en los procesos sociales en determinado espacio, y de esa forma dar una mirada comprensiva y reflexiva sobre lo que ocurre en internet. La netnografía, como forma de investigación, hace un énfasis en la recolección y el procesamiento de cantidades significativas de datos empíricos para producir conocimiento al respecto.

Del Fresno (2011) afirma que el auge de las nuevas formas de comunicación en red abre un espacio para la netnografía como “una disciplina con gran potencial en el campo de la investigación que ofrece el ciberespacio para las ciencias sociales” (p. 64). Una de las ventajas que brinda la netnografía es la posibilidad de realizar observación no participante, para describir los procesos que se estudian sin las interferencias que la inmersión del investigador puede traer en el espacio *offline*. En otros trabajos (Álvarez-Bornstein & Montesi, 2016; Badillo-Mendoza & Marta-Lazo 2019; Ventura, 2018) ya se ha utilizado la técnica de observación no participante y de registro etnográfico para estudiar la comunicación en Twitter.

Para la exploración etnográfica virtual se propone la elaboración de dos tipos de registros en paralelo. El primero de ellos es un relevamiento de información en periódicos digitales, el cual permite hacer un seguimiento a los eventos noticiosos más destacados durante un periodo determinado al fenómeno que se esté estudiando, por ejemplo, una campaña electoral, una protesta o algún evento. Esta información se puede extraer a través del buscador de cada portal de noticias por medio de palabras clave y así es posible reconstruir determinado contexto o situación. El segundo registro que se sugiere es el de los *trending topics* o tendencias que tuvieron lugar en Twitter. Se trata de las palabras o expresiones que pueden estar agrupadas con el símbolo # (*hashtags*) y sirven para organizar el material dentro de Twitter, pero se basan en temas o ideas, y no en usuarios (Squires, 2016). De manera que, la revisión de las tendencias de *hashtags* permite saber qué temas se estaban discutiendo en la plataforma en determinado periodo de tiempo. Para obtener estos datos se puede utilizar la página web libre Trendinalia⁴ que monitorea los temas del momento en

⁴ El portal de Trendinalia también presenta una serie de gráficos en los que se visualiza la evaluación de los *hashtags* junto a la geolocalización para mostrar la notoriedad que tuvo. <http://www.trendinalia.com/>

Twitter. En este portal solo es necesario seleccionar la ubicación (continente y país), las fechas y a continuación se revela una lista con las tendencias más destacadas, acompañada de una serie de gráficos que muestran el tiempo de duración, la localidad donde tuvieron mayor relevancia y su evolución por horas. Este acercamiento brinda un primer retrato de los acontecimientos y sirve para caracterizar el panorama de la discusión en Twitter.

De este modo, con los registros de información de los periódicos digitales y los *trending topics*, se puede elaborar un chequeo cruzado que permite relacionar los acontecimientos que fueron noticia en la prensa escrita digital con las tendencias de Twitter. Con este procedimiento se pueden observar las hipermediaciones (Scolari, 2008), es decir, procesos de intercambio, producción y consumo simbólico desarrollado en la red. En efecto, con la revisión cruzada de los datos de las dos fuentes mencionadas se observa si aquello que fue tendencia en la red social tuvo alguna repercusión en la prensa o si, en sentido opuesto, alguna noticia fue viralizada en Twitter. En la Tabla 1 se puede ver una muestra de este procedimiento retomado de la investigación en curso⁵. Este registro cruzado de ambas fuentes además posibilita la generación de esquemas o diagramas que sintetizan la información para que pueda ser visualizada de una mejor forma. Por ejemplo, cuando se trata de algún acontecimiento durante un periodo de tiempo, es posible elaborar una línea de tiempo que organice la secuencia de hechos junto con los *hashtags* y, de esa forma, se pueda ver su evolución.

Tabla 1. Ejemplo de registro cruzado de noticias y tendencias por fecha

Fecha	Periódico	Titular	Hashtags (Tendencias)
29/2/2016	ET	Capturan a Santiago Uribe por caso de ‘Los 12 Apóstoles’	#PazTerritorial, #SantiagoUribe
4/3/2016	ET	Tras la captura de su hermano, Uribe arremete contra el presidente Santos	#PazEsJusticia, #TomásYJerónimo
16/3/2016	ET	Consejo de Estado dio concepto favorable a la Corte sobre el plebiscito	#YoPerdono, #PazConGarantías
2/4/2016	ET	Miles acogieron llamado del uribismo y marcharon en contra del Gobierno	#Abril2ALaCalle, #NoMasDesgobierno, #NoMasImpunidad

5 En este caso se presenta una tabla con las noticias y las tendencias sobre un periodo previo a la realización del plebiscito sobre el Acuerdo de Paz en Colombia. Algunos avances se pueden consultar en Bonilla-Neira (2021, 2022).

10/5/2016	EE	“Enfrentaremos el Acuerdo Especial con resistencia civil”: Uribe	#ResistenciaCivil, #NoMeResistoALaPaz
26/5/2016	EE	El choque de posiciones a favor del plebiscito por la paz	#CorteDigaSíALaPaz, #PazSinImpunidad
2/6/2016	ET	Uribismo anuncia ‘firmatón’ nacional contra el plebiscito por la paz	#RealidadDelProcesoDePaz, #DescanseEnPazLaGuerra

Nota: las columnas de esta tabla muestran el registro de las noticias con la fecha, el periódico (en este caso EE, El Espectador, y ET, El Tiempo), titular y los *hashtags* que eran tendencia en ese momento.

Consideraciones éticas

Antes de presentar el procedimiento de extracción y manejo de datos, es preciso referirse al aspecto ético. Un elemento clave que se tiene en cuenta en el uso de información de internet es la evaluación del grado de privacidad o apertura del entorno en línea que se pretende estudiar (Sveningsson Elm, 2009). Twitter es una plataforma de acceso público, cuyas publicaciones en los perfiles individuales o en los de otros usuarios pueden ser vistas por defecto y buscadas por cualquier persona en el mundo⁶. Twitter recopila los tuits de los usuarios para obtener información del contenido que ha leído, retuiteado o al que le ha dado “me gusta”, el idioma y la edad, para de esa forma ofrecerles contenido personalizado (Twitter, 2020). De manera que la mayor parte de los datos que circulan en esta red social son públicos. Además, en su política de privacidad se aclara que la información compartida por los usuarios es objeto de difusión y es susceptible de ser utilizada por otros sitios web o aplicaciones para ser analizada. Tal es el caso de la API de Twitter que permite el acceso a los tuits de forma libre en ciertas cantidades con fines investigativos y de forma arancelada a gran escala para estudios de mercado.

No obstante, como afirman Page *et al.* (2014), el carácter público de los datos no es garantía automática de que quienes publican en determinado sitio web quieran convertirse en material de estudio. Al respecto, De Matteis (2014) arguye que los datos escritos en línea deben ser tratados con mayor precaución para proteger la identidad real de los sujetos observados, verificar si las publicaciones provienen de alguna persona o comunidad vulnerada; por lo cual es necesario detenerse y analizar si la información es sensible y puede comprometerlos, y de ser así, optar por un consentimiento informado. Por ejemplo, en el caso de perfiles de políticos, instituciones u organizaciones civiles en Twitter, se trata de información elaborada y publicada para ser difundida a la comunidad, dado que son perfiles abiertos y oficiales en muchos casos. También, cuando los usuarios emplean los *hashtags* en los tuits

⁶ Excepto si en la configuración el usuario decide hacerlas privadas seleccionando la opción de “tuits protegidos”.

no solo se está etiquetando un tema, sino que se está conectando a la conversación con otros usuarios en lo que Zappavigna denomina *searchable talk* (2012, 2015) o “conversación buscable”; con lo cual el discurso puede ser encontrado, compartido, difundido y, eventualmente, estudiado.

Recolección de datos

Con las observaciones y registros producto de la etnografía virtual antes descrita se pueden reconocer las cuentas de Twitter más activas sobre los temas de interés al utilizar los *hashtags* que fueron tendencia y, de esa forma, no limitarse a las cuentas de los políticos más sobresalientes (a menos que el interés sea solo el de analizar esos discursos). Los datos se obtienen de modo automatizado a través de la API de Twitter, la interfaz de aplicaciones para programadores usando un *script* de Python. Para ello es necesario registrarse en la plataforma y obtener una autenticación en Twitter Developer, lo cual habilita la descarga de datos⁷. Así, se procede a la extracción y el procesamiento de los datos con Python, que cuenta con librerías abiertas especiales para el tratamiento del material de Twitter, como se puede ver en el repositorio que acompaña a este artículo⁸. Se extraen los datos con el paquete *tweepy* de Python, se lee y procesa el material principalmente con las librerías *pandas*, *numpy* y *matplotlib*. La API permite la descarga de aproximadamente 3200 tuits por cuenta. Los datos extraídos se presentan en texto plano y en formato de base de datos, como se muestra en la Figura 1.

Figura 1. Captura de pantalla de los datos como archivo de texto plano

```

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000

```

Nota: la imagen corresponde al texto sin formato que se extrae de la API. Este tipo de archivos está formado solo por caracteres legibles (excluye los símbolos alfanuméricos que dan cuenta de elementos como los emojis), pero aún carecen de formato óptimo para su comprensión.

En la exploración y el procesamiento de los datos capturados es posible encontrar de forma automática, por ejemplo, el tuit más retuiteado, el que más *likes* tuvo y de esa forma empezar a visualizar tópicos recurrentes y ausencias en los datos, elementos que más adelante pueden ser importantes en el análisis discursivo. Los datos obtenidos cuentan con la siguiente información: fecha de creación, número de favoritos (o *likes*),

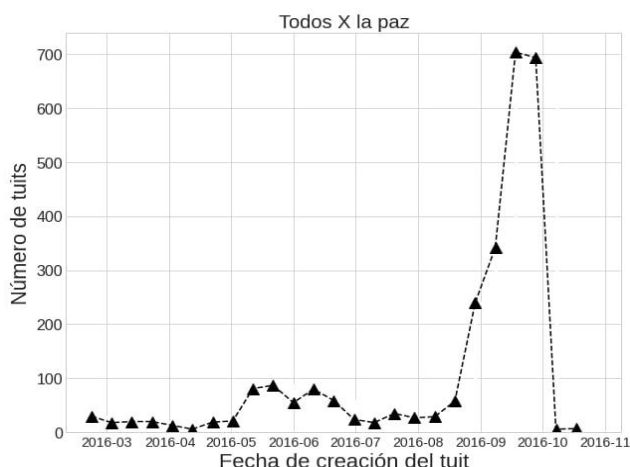
7 En el proceso de autenticación se explica que el uso de los datos tiene fines educativos y de investigación social y esto habilita una descarga masiva o en grandes cantidades de datos. Es un protocolo de seguridad que tiene Twitter para el monitoreo de los datos.

8 En este repositorio se pueden encontrar los *scripts* utilizados para el procedimiento: <https://github.com/AnalisisTuits/DiscursoTwitter>

de seguidores, de identificación, de comentarios, de retuits, el nombre de perfil, el texto como tal y un hipervínculo si contiene algún material multimedia: imágenes o videos principalmente. A través de otra librería de Python es posible organizar los datos en tablas. No obstante, debido al gran número de datos y a que, en algunos casos, los tuits quedan cortados y con caracteres de símbolos, lo cual hace que no se puedan observar de forma integral y así se dificulta la decodificación, es necesario optar por otras herramientas de visualización, como se verá en el apartado 4.

Usando Python también es posible seguir la evolución del número de tuits durante el periodo de tiempo elegido. Así, se pueden diseñar gráficos como el del ejemplo de la Figura 2, en el cual se observa en el eje vertical el número de tuits publicados y en el eje horizontal el mes en el que fueron producidos. En este ejemplo se advierte con claridad el aumento en la producción de tuits en el segundo trimestre (desde mayo hasta julio de 2016) y luego el crecimiento exponencial en el tercer trimestre (desde agosto hasta octubre de 2016).

Figura 2. Gráfico de frecuencia de número de tuits por mes



Nota: el gráfico presenta la frecuencia de los tuits durante el periodo de febrero a octubre de 2016 de una campaña plebiscitaria del perfil “Todos por la paz”. En este caso se observa un número regular para los meses de febrero, marzo y abril; en mayo hay un incremento que luego vuelve a bajar y desde septiembre aumenta. El acrecentamiento final se puede entender por la cercanía a la fecha de la elección.

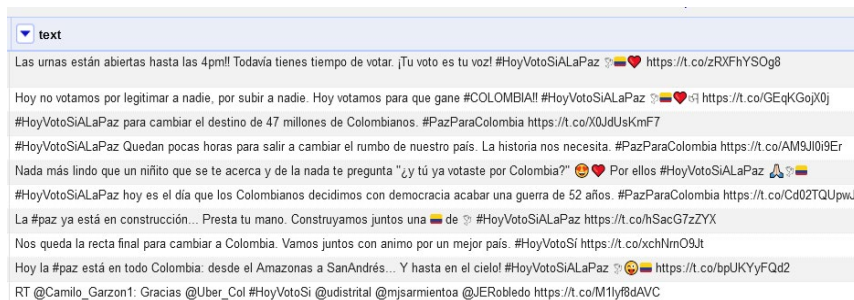
Construcción del corpus

A continuación, y dependiendo del número de tuits extraídos, puede ser necesario hacer un proceso de muestreo, sobre todo al realizar una investigación fundamentalmente cualitativa. La selección de la muestra apunta a señalar con más detalle aquellos casos o unidades de análisis que conformarán el corpus. En primera instancia, se seleccionan las cuentas a partir de criterios inductivos relevantes para cada investigación en particular. Por ejemplo, que las cuentas hayan sido creadas

en un momento determinado, que hayan tenido actividad durante ese tiempo, que sean individuales o colectivas, de perfiles oficiales o no y que tengan determinado número de seguidores (información que puede ser extraída de netnografía realizada previamente). Con la idea de no sesgar la muestra y con los perfiles seleccionados a partir de esos criterios, se propone utilizar un muestreo combinado o mixto en el cual se incluye una muestra por aleatoriedad y otra dirigida (Hernández-Sampieri & Mendoza-Torres, 2018). Por ejemplo, en Bonilla-Neira (2021) se optó por seleccionar un 20 % de los tuits al azar y los 10 tuits con más retuits (RT) de los datos. Si se tiene en cuenta que el análisis que se pretende hacer es discursivo, resulta importante conocer las voces que tuvieron más eco en los datos, es decir, aquellos tuits que más fueron compartidos. Por ese motivo, se toman en cuenta los *retuits* que son entendidos como heterogeneidades discursivas (Authier-Revuz, 1984; Grossmann, 2019). No obstante, el muestreo aleatorio permite mitigar el posible sesgo en los datos seleccionados. La librería *pandas* ejecuta la operación de muestreo de forma automática, solo se debe indicar el porcentaje de tuits que se quiere seleccionar.

En seguida, con el corpus constituido, se procede al refinamiento y a la mejora de la visualización. Como se mencionó en el apartado anterior, los datos capturados se encuentran en texto plano, es decir, sin ningún tipo de formato tipográfico ni ideogramas, de manera que para mejorar su presentación se utiliza la herramienta de visualización Open Refine. Se trata de una aplicación de código abierto (Open Refine, 2022) para limpieza y transformación de datos. Funciona como una base de datos con columnas y filas, que permite organizar y filtrar los datos, lo cual habilita su exploración en un entorno amigable, como se muestra en la Figura 3. Este *software* libre opera como una aplicación web local, es decir, se inicia en un servidor web y se abre en el navegador predeterminado. La herramienta permite la organización de datos al combinar, separar, agregar o eliminar columnas para mejorar su visualización. Por ejemplo, separar del texto el enlace o los *hashtags* incluidos y hacer otras filas para que puedan verse a simple vista. Además, las columnas tienen unas pestañas para buscar y filtrar determinada información, por ejemplo, para seleccionar palabras de interés en los tuits. Este entorno gráfico también sirve para observar los emojis y recuperar el color, con lo cual se puede apreciar mucho mejor el material. En ese sentido, Open Refine es ideal para el manejo de información como la de Twitter por el tipo de formato de texto en el que se descarga (generalmente json y cvs). En el visualizador se pueden encontrar las columnas con la enumeración de los datos, luego el texto, el número de identificación, la fecha, el nombre de la cuenta, el número de retuits, el número de favoritos y los seguidores que tenía la cuenta al momento de la descarga.

Figura 3. Captura de pantalla del entorno de Open Refine



Nota: el programa toma el texto plano que se extrajo de Twitter, lo convierte y organiza en columnas con la información correspondiente. En esta captura de pantalla solo se observa la columna del texto para que se vea mejor el detalle de los tuits.

Análisis descriptivo del corpus

Después de la construcción del corpus se da paso al análisis descriptivo, que implica hacer una revisión detallada de los tuits. Cabe recordar que los tuits son enunciados plurisemióticos complejos (Paveau, 2017) en los que es necesario apreciar las relaciones establecidas entre, al menos, los modos verbal y visual. Si bien es cierto que Twitter es una red centrada en la escritura (Squires, 2016), su algoritmo impulsa cada vez con más fuerza las publicaciones con material visual, sean imágenes estáticas o en movimiento. En la Figura 4 se puede apreciar un tuit típico de ejemplo con una radiografía básica de los elementos que lo componen. Entre sus partes se encuentran el nombre del usuario resaltado en negrilla, el identificador de usuario codificado con la @, la imagen de perfil, la fecha de producción del tuit y los modos verbal y visual que componen el tuit. El texto también puede estar acompañado por un enlace de otra plataforma, una noticia o un video.

Figura 4. Radiografía de un tuit



Nota: ejemplo de un tuit compuesto por texto e imagen, en este caso un afiche. Se presentan cada uno de sus módulos: nombre de usuario, identificador, imagen de perfil, datos de producción (hora, fecha y plataforma de gestión), botones de interacción (casillas de comentario, retuit y me gusta) y tuit (con los componentes verbal y visual).

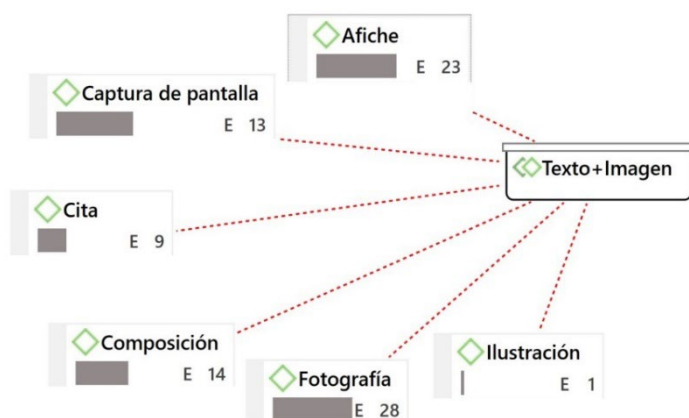
Para este procesamiento se sugiere el apoyo de algún programa de análisis cualitativo de datos como Atlas.ti, que permite el etiquetado sistemático de forma ágil a través de códigos definidos por quien investiga. Además, es importante que en este *software* es posible observar los tuits junto con las imágenes si las contienen. De esta manera, se estaría considerando la forma ecológica (Paveau, 2017) del tuit, en la medida en que se atiende a más elementos del contexto en la comprensión. El procedimiento de análisis que se sugiere consiste entonces en almacenar y agrupar los tuits en documentos (pueden ser Word o PDF), en este caso los tuits por perfiles, en un proyecto, también llamado unidad hermenéutica y, luego, dar paso a la codificación de estos.

En primera instancia, los tuits se pueden clasificar en cuatro categorías generales siguiendo el modelo propuesto por Gallardo-Paúls *et al.* (2018) de acuerdo con la estructura semiótica: texto, texto con imagen fija, texto con imagen en movimiento y texto con enlace. A su vez, dentro de la categoría “imagen fija” se utiliza la caracterización sobre el formato de Marchal *et al.* (2021): fotografía, que incluye las *selfies*, las que el usuario toma a otros; ilustración, que incorpora dibujos o caricaturas; captura de pantalla, que puede ser a la pantalla del computador o del celular, por ejemplo, de noticias o publicaciones de otras redes sociales; infografía, que consiste en la representación visual de datos, se incluyen también estadísticas, mapas o gráficas con resultados de encuestas; composición, que es una mezcla de varias categorías, por ejemplo, un collage o serie de fotos, también un meme o un video y un cartel a la vez, y afiche, que supone carteles promocionales, folletos y volantes. En la Figura 5 se observan algunos ejemplos de estas categorías y en la Figura 6 se aprecia una visualización básica en un documento de Atlas.ti. En cuanto a las imágenes en movimiento, se puede atender a categorías emergentes halladas en el corpus, tales como videos testimoniales, *spots* o animaciones. Sobre el texto con enlace, en algunos casos puede que se encuentre el vínculo embebido y aparezca la fotografía del portal web enlazado; en otros casos, solo se presenta la URL. Esta codificación inductiva inicial permite examinar la constitución general del corpus y reconocer las temáticas más recurrentes y los formatos en que esta se presenta.

Figura 5. Ejemplos de tuits con imágenes fijas



Nota: ejemplos de tipos de tuits basados en la categorización sobre el contenido visual compartido en Twitter (Marchal *et al.*, 2021). De izquierda a derecha y de arriba a abajo, en la primera imagen se encuentra un tuit con fotografía; el segundo, cita; el tercero, afiche; el cuarto, infografía (tipo encuesta); el quinto, ilustración; sexto, composición.

Figura 6. Ejemplo de un gráfico básico de red con Atlas.ti

Nota: el gráfico muestra una red básica de un documento categorizado en Atlas.ti. Se observa el código: “Texto + imagen” con los subcódigos que lo componen: Afiche, captura de pantalla, cita, composición, fotografía e ilustración. La letra “E” da cuenta del enraizamiento, es decir, el número de tuits asignados con el código respectivo.

En segunda instancia, debido a la naturaleza del corpus: tuits sobre un asunto público de carácter político-social, se considera importante codificarlo a partir de categorías que den cuenta de la interacción multimodal, para lo cual se utiliza la propuesta de Roque (2016, 2017). Si bien los planteamientos del autor han sido expuestos para analizar imágenes como afiches sobre la guerra o anuncios publicitarios por la evidente intención argumentativa, se tiene en cuenta que también se pueden emplear para los tuits, porque en la mayoría de los casos se presenta una interferencia entre los modos de composición. Específicamente, resulta relevante establecer una codificación balanceada de la interacción entre el encuadre verbal y el elemento visual que conforma el tuit, con el fin de comprender su funcionamiento.

Se trata entonces de distinguir cuatro tipos de interferencias verbo-icónicas, como podemos observar en los ejemplos de la Figura 7: 1) la bandera visual, *visual flag* (Birdsell & Groarke 2007), cuando la imagen atrae la atención hacia un argumento que se presenta verbalmente, estrategia muy utilizada en los tuits que tienen la imagen como enganche para atraer la atención de los usuarios; 2) el paralelismo, cuando tanto la imagen como el texto presentan la misma idea y se podría decir que es redundante, por ejemplo, un tuit descriptivo de una acción que se encuentra exactamente en la imagen; 3) la complementariedad o argumento conjunto, cuando los dos enunciados se complementan, aportando cada uno información importante para el sentido general del enunciado, como el caso del enunciado verbal de un tuit que indica ver lo que se encuentra en la imagen; 4) el antagonismo, cuando los dos enunciados presentan una relación opuesta, es decir, mientras el enunciado visual muestra una situación, el enunciado verbal lo niega, tal es el caso de un tuit en el

que, por ejemplo, se pide que cese la violencia y la imagen es justamente una escena violenta.

Figura 7. Ejemplos de funciones de la imagen en los tuits



Nota: de izquierda a derecha y de arriba a abajo se encuentran ejemplos de bandera visual, paralelismo, complementariedad y antagonismo.

Finalmente, se puede dar paso a un análisis discursivo exhaustivo de acuerdo con los intereses y objetivos de cada investigación. Por ejemplo, en el caso del trabajo con el que se ejemplifica este recorrido metodológico, se analiza la construcción del *ethos* como dimensión retórico-argumentativa, así como el registro discursivo de la violencia verbal para lo cual se emplean, entre otras, categorías de la lingüística de la enunciación y constituyentes del discurso polémico (Bonilla-Neira, 2020, 2021). Aunque con otro proceso de recolección de datos, Arrieta y Avendaño (2018), con una perspectiva sociodiscursiva y sociopragmática, analizan mensajes de Twitter de dos expresidentes de Colombia desde la metodología del análisis crítico del discurso (ACD). También, Mazucchino (2020) estudia, desde el análisis de discurso de tendencia francesa, los tuits de un expresidente y un presidente mexicanos, y describe el tuit como el sustituto actual del panfleto tradicional. En otros trabajos como el de Ventura (2021) se caracteriza la estrategia de “presentar propuestas de campaña en Twitter” en los tuits de candidatos presidenciales en Argentina, en el marco del análisis estratégico del discurso a partir del estudio de recursos gramaticales y pragmático-discursivos de los modos semióticos. De esta forma, una vez construido el corpus de análisis de tuits, se pueden hacer distintos tipos de planteamientos según sea el objetivo de la investigación.

Conclusiones

Este trabajo permitió la reconstrucción de un recorrido metodológico llevado a cabo en una investigación que integra elementos cualitativos y cuantitativos para estudiar el discurso digital en datos provenientes de Twitter. En un panorama con escasos artículos enfocados en el componente metodológico para afrontar este tipo de datos, resulta relevante facilitar esta ruta procedimental. En esa línea, este artículo detalló el tratamiento de la información, que por momentos puede ser abrumadora, razón por la cual recupera el proceso de inmersión a los datos y de construcción de un corpus manejable apto para el análisis. De esta manera, se ofreció una alternativa metodológica en la cual se presenta de forma conjunta información sobre métodos, técnicas y herramientas que pueden ser utilizadas para procesar datos de Twitter y construir el corpus.

La construcción de esta propuesta metodológica combina técnicas para la observación, como la netnografía; para la recolección, como la captura de tuits por medio de la API de Twitter, y para el procesamiento de los tuits con librerías de Python y a través de herramientas de visualización como Open Refine y análisis, como el *software* para el análisis cualitativo de datos Atlas.ti. La propuesta aquí desarrollada además plantea el uso de un registro cruzado de datos (noticias y tendencias) en el momento exploratorio, lo cual permite ahondar en la situación de comunicación de los acontecimientos, y propone el uso de dos clasificaciones de imágenes recuperadas de distintas propuestas, lo cual al reunir las en una única clasificación permite catalogar los tuits en función de su relación entre imagen y texto. Asimismo, se reconoció en Twitter un espacio digital del cual se pueden tomar muestras de la discusión pública que allí se genera. Para esto se tuvieron en cuenta los aspectos éticos del carácter público del material y, al mismo tiempo, el tratamiento cuidadoso de los datos.

En síntesis, esta propuesta recupera un proceso que ha sido de idas y vueltas, y este trabajo ha permitido conectar pasos y trazar caminos para dar un esbozo del diseño metodológico que se ha estado desarrollando en una investigación doctoral que profundiza en el análisis discursivo del corpus. A partir de esa experiencia se buscó ofrecer líneas posibles para organizar un diseño de investigación con ejemplos que permiten orientar mejor el proceso en un área que aún sigue reflexionando sobre este asunto. Se espera que este procedimiento pueda servir para otro tipo de investigaciones, dado que la organización de estos recursos y su procedimiento permitió la construcción de un corpus flexible con mitigación del sesgo y preparado para su posterior análisis.

Agradecimientos especiales al Dr. Christian Sarmiento-Cano por la colaboración en el soporte computacional para el manejo de los datos y a la Lic. Noelia Stetie por el estímulo y las lecturas de los borradores que me ayudaron a mejorar la presentación del trabajo.

Referencias

- Alcántara-Plá, M. (2020). Metodología híbrida para el análisis del discurso digital. El ejemplo de «democracia» en Twitter. *Cuadernos Aispi*, 16, 25-44. <https://doi.org/10.14672/2.2020.1696>
- Álvarez-Bornstein, B., & Montesi, M. (2016). La comunicación entre investigadores en Twitter. Una etnografía virtual en el ámbito de las ciencias de la documentación. *Revista Española de Documentación Científica*, 39(4), 1-15. <https://doi.org/10.3989/redc.2016.4.1352>
- Arrieta, L. E. & Avendaño, G. (2018). El discurso del tuit: un análisis lingüístico, sociodiscursivo y sociopragmático. *Cuadernos de Lingüística Hispánica*, 32, 107-130. https://revistas.uptc.edu.co/index.php/linguistica_hispanica/article/view/8123/6591
- Authier-Revuz, J. (1984). Hétérogénéité(s) énonciative(s). *Langages*, 19(73), 98-111.
- Badillo-Mendoza, M. & Marta-Lazo, C. (2019). Ciberciudadanía a través de Twitter: caso Gran Marcha Carnaval y consultas populares contra la minería en La Colosa. *Cuadernos.info*, 45, 145-162. <https://doi.org/10.7764/cdi.45.1454>
- Beas, D. (2010). *La reinención de la política: Obama, internet y la nueva esfera pública*. Península.
- Birdsell, D. S., & Groarke, L. (2007). Outlines of a Theory of Visual Argument. *Argumentation and Advocacy*, 43(3-4), 103-113. <https://doi.org/10.1080/0028533.2007.11821666>
- Bonilla-Neira, L. (2020). Tópicos y violencia verbal en la convocatoria a la marcha #NoMásDesgobierno en Colombia. *Revista de Estudos da Linguagem*, 28(4), 1747-1777. <https://doi.org/10.17851/2237-2083.28.4.1747-1777>
- Bonilla-Neira, L. (2021). Construcción de la imagen colectiva de grupos a favor del Acuerdo de Paz de Colombia en Twitter. *Revista de Estudos da Linguagem*, 29(4), 2225-2257. <https://doi.org/10.17851/2237-2083.29.4.2225-2257>
- Bonilla-Neira, L. (2022). Polémica en Twitter por la “resistencia” al Acuerdo de Paz en Colombia. En C. Pineira-Tresmontant (ed.), *Dire et ne pas dire la violence* (pp. 113-130). L'Harmattan.
- Calvo, E. (2015). *Anatomía política de Twitter en Argentina*. Capital Intelectual.
- Castells, E. (2008, ene.-mar.). Los medios y la política. *Revista TELOS, Revista de Pensamiento, Sociedad y Tecnología*, 1-15. <https://telos.fundaciontelefonica.com/archivo/numero074/los-medios-y-la-politica/?output=pdf>

- Castells, M. (2009). La comunicación en la era digital. En *Comunicación y poder* (pp. 87-189). Siglo XXI.
- DataReportal. (2022). *Twitter Stats and Trends. Essencial Twitter Stats for 2022*. <https://datareportal.com/essential-twitter-stats>
- Del Fresno, M. (2011). *Netnografía. Investigación, análisis e intervención social online*. Editorial UOC.
- De Matteis, L. (2014). Ejes para un debate sobre el uso ético de datos interaccionales escritos y orales obtenidos en línea. En *Las humanidades digitales desde Argentina. Tecnologías, culturas, saberes* (pp. 235-247). Repositorio digital UNS. [https://repositoriodigital.uns.edu.ar/bitstream/handle/123456789/2851/De-Matteis%2C L. Ejes para un debate....pdf?sequence=1&isAllowed=y](https://repositoriodigital.uns.edu.ar/bitstream/handle/123456789/2851/De-Matteis%2C%20L.%20Ejes%20para%20un%20debate....pdf?sequence=1&isAllowed=y)
- Freire, N. (2019). Por qué es Twitter el territorio político digital. *Polis*, 15(2), 39-74.
- Gallardo-Paúls, B., Enguix-Oliver, S. & Oleaque-Moreno, J. (2018). Estilos de gestión de los perfiles políticos en Twitter: imagen y texto en las cuentas de los partidos políticos en la campaña del 26J. *Revista de Investigación Lingüística*, 21, 15-51. <https://doi.org/https://doi.org/10.6018/ril.21.367371>
- Grossmann, F. (2019, 4 de feb.). Discours rapporté versus Discours partagé : convergences, différences, problèmes de frontières. En *Conférence invitée dans le cadre du colloque Ci-dit, Université libre de Bruxelles*. <https://hal.archives-ouvertes.fr/hal-02005379>
- Hernández-Sampieri, R. & Mendoza-Torres, C. P. (2018). *Metodología de la investigación. Las rutas cuantitativa, cualitativa y mixta*. McGrawHill Education.
- Hine, C. (2004). Etnografía virtual. En *Etnografía virtual*. Editorial UOC. <https://doi.org/10.1111/j.1468-5930.1987.tb00214.x>
- Honeycutt, C., & Herring, S. C. (2009). Beyond Microblogging : Conversation and Collaboration via Twitter. In *Proceedings of the Forty-Second Hawai'i International Conference on System Sciences (HICSS-42)* (pp. 1-11). <https://doi.org/10.1109/HICSS.2009.89>
- Hürlimann, M., Davis, B., Cortis, K., Freitas, A., & Handschuh, S. (2016). A Twitter Sentiment Gold Standard for the Brexit Referendum. In *Proceedings of the 12th International Conference on Semantic Systems* (pp. 193-196). ACM.
- Kozinets, R. V. (2015). *Netnography: Doing Ethnographic Research Online* (2nd ed.). Sage Publications.

- Kreis, R. (2017). The “Tweet Politics” of President Trump. *Journal of Language and Politics*, 16(4), 607-618. <https://doi.org/10.1075/jlp.17032.kre>
- Longhi, J. (2013). Essai de caractérisation du tweet politique. *L'information Grammaticale*, 136, 25-32.
- Mancera Rueda, A. & Pano Alamán, A. (2020). *La opinión pública en la red. Análisis pragmático de la voz de ciudadanos*. Iberoamericana/Vervuert.
- Marchal, N., Neudert, L. M., Kollanyi, B., & Howard, P. N. (2021). Investigating Visual Content Shared over Twitter During the 2019 EU Parliamentary Election Campaign. *Media and Communication*, 9(1), 158-170. <https://doi.org/10.17645/mac.v9i1.3421>
- Mazzuchino, M. G. (2020). El tuit como micropanfleto: el caso de la asunción del nuevo presidente mexicano. *Lexis*, 44(2), 525-574.
- Open Refine. (2022). *Open Refine*. GitHub. <https://github.com/OpenRefine/OpenRefine#contact-us>
- Ott, B. L. (2017). The Age of Twitter: Donald J. Trump and the Politics of Debasement. *Critical Studies in Media Communication*, 34(1), 59-68. <https://doi.org/10.1080/15295036.2016.1266686>
- Page, R., Barton, D., Unger, J. W., & Zappavigna, M. (2014). *Researching Language and Social Media. A Student Guide*. Routledge.
- Pallarés, S. & García, C. (2017). Análisis comparativo del discurso de Mariano Rajoy en Twitter durante las dos campañas electorales de 2015 : las autonómicas del 24M y las generales del 20D. *Revista Dígitos*, 3, 119-136.
- Paveau, M.-A. (2017). *L'analyse du discours numérique. Dictionnaire des formes et des pratiques*. Hermann.
- Peirano, M. (2019). *El enemigo conoce el sistema. Manipulación de ideas, personas e influencias de la economía de la atención*. Penguin Random House.
- Robinson, L., & Schulz, J. (2014). *New Avenues for Sociological Inquiry. August 2009*. <https://doi.org/10.1177/0038038509105415>
- Roque, G. (2016). *Esquisse d'une rhétorique des interactions verbo-iconiques*. Open Edition Journals. <https://doi.org/10.4000/imagesrevues.3434>.
- Roque, G. (2017). Rhetoric, Argumentation, and Persuasion in a Multimodal Perspective. In A. Tseronis & C. Forceville (Eds.), *Multimodal Argumentation and Rhetoric in Media Genres* (p. 313). John Benajmins Publishing Company.
- Rovira, G. (2019). *Activismo en red y multitudes conectadas. Comunicación y acción en la era de internet*. Icaria- UAM-X.

- Scolari, C. (2008). *Hipermediaciones. Elementos para una teoría de la comunicación digital interactiva*. Gedisa.
- Seixas, R., & Nascimento, L. (2021). Impeachment or Death: The Rhetorical Configuration of a Polemic Event in the Digital Public Space. *Revista de Estudos da Linguagem*, 29(4), 2397-2428. <https://doi.org/10.17851/2237-2083.29.4.2397-2428>
- Slimovich, A. (2016). La digitalización de la política y la vuelta de lo televisivo. El caso de los candidatos argentinos en Facebook. *Revista de Comunicación*, 15, 111-127. http://udep.edu.pe/comunicacion/rcom/es/articulos/2016/6_Art111-127.html
- Squires, L. (2016). Twitter. Desing, Discourse, and the Implications of Public Text. In A. Georgakopoulou & T. Spilioti (Eds.), *The Routledge Handbook of Language and Digital Communication* (pp. 239-255). Routledge.
- Sveningsson Elm, M. (2009). How Do Various Notions of Privacy Influence Decisions In Qualitative Internet Research? In A. Markham & N. Baym (Eds.), *Internet Inquiry: Dialogue Among Researchers* (pp. 69-98). Sage.
- Twitter. (2020). *Global Impact Report*. <https://about.twitter.com/content/dam/about-twitter/en/company/global-impact-2020.pdf>
- Van Dijck, J. (2016). *La cultura de la conectividad. Una historia crítica de las redes sociales*. Siglo XXI.
- Vasilachis de Gialdino, I. (2009). La investigación cualitativa. En I. Vasilachis de Gialdino (ed.), *Estrategias de investigación cualitativa* (pp. 23-64). Gedisa.
- Ventura, A. (2018). *¿Cómo analizar discursos de 140 caracteres? Propuesta metodológica para el estudio del discurso estratégico con una perspectiva multimodal y crítica*. CHIMERA: Revista de Corpus de Lenguas Romances y Estudios Lingüísticos, 5(2), 275-287. <https://doi.org/10.15366/chimera2018.5.2.006>
- Ventura, A. (2021). De presupuestos e implicaturas: las propuestas en la campaña presidencial argentina 2015 en Twitter. *REDD Revista Estudios del Discurso Digital*, 4, 1-48. <https://doi.org/10.24197/redd.4.2021.1-48>
- Zappavigna, M. (2011). Ambient Affiliation: A Linguistic Perspective on Twitter. *New Media and Society*, 13(5). <https://doi.org/10.1177/1461444810385097>
- Zappavigna, M. (2012). *Discourse of Twitter and Social Media*. Continuum. <https://doi.org/10.11139/cj.31.3.412-414>
- Zappavigna, M. (2015). Searchable Talk : The Linguistic Functions of Hashtags. *Social Semiotics*, 25(3), 274-291. <https://doi.org/10.1080/10350330.2014.996948>