

# Las redes sociales como herramienta para la investigación de neologismos

## *Social Media as a Research Tool in Neology*

**Mar CAPILLA MARTÍN**

*Universidad Autónoma de Madrid*

mar.capilla@gmail.com

**Resumen:** El presente artículo tiene dos objetivos principales. Por un lado, mostrar la utilidad de las redes sociales como herramientas para la investigación lingüística. Para ello, nos centraremos en algunos neologismos formales que constituyen una evidencia clara de la creatividad léxica y de la productividad de los procesos morfológicos (*tiktoker*, *pyme*, *presintomático*, entre otros). Por otro lado, determinar si constituyen las redes sociales un corpus lingüístico o si se consideran meras fuentes de acceso a la información. En este marco, trataremos de establecer cómo es de representativo Twitter comparado con el CORPES XXI con el fin de descubrir qué herramienta resulta más útil para el estudio de neologismos. En este trabajo, además, se presentan y discuten las ventajas y los inconvenientes de la incorporación de redes sociales como herramienta de estudio para la investigación lingüística.

**Palabras clave:** lingüística de corpus; neologismos; redes sociales; morfología; Internet.

**Abstract:** Social networks allow us to interact with people, create content, share it and, ultimately, spread it. This social universe continues to be another means of communication through which the language is transmitted. If we consider the reference corpus, whose purpose is to obtain global features of the language at a certain moment in its history, social networks represent a novel source of information. This chapter will present through some formal neologisms —which constitute clear evidence

of lexical creativity and the productivity of morphological processes (tiktokers, pyme, presintomático, and others)—, the usefulness of social networks as tools for linguistic research and will analyse the problems and limitations of using the social network Twitter as a corpus for linguistic research.

**Palabras clave:** corpus linguistics; neologisms; social media; morphology; Internet.

## 1. INTRODUCCIÓN

Tras la normalización de Internet y la entrada de nuevas tecnologías que permiten la interacción entre los hablantes, surgen las redes sociales, herramientas de difusión que suponen una revolución en el mundo de la comunicación, así como en el ámbito social y digital. Plataformas como Facebook, Twitter o Instagram ofrecen información valiosa sobre el comportamiento del lenguaje en redes sociales, un contexto más relajado y coloquial, próximo al nivel conversacional. Tal es su importancia que en muchas investigaciones, incluso, se han utilizado las redes sociales como un corpus de estudio en sí mismo de donde extraer datos relevantes para un posterior análisis. En este trabajo, por un lado, trataremos de determinar si una red social como Twitter puede considerarse un corpus lingüístico. Por otro lado, trataremos de mostrar las ventajas, o desventajas, de la utilización de las redes sociales en la investigación lingüística frente a otros corpus lingüísticos. En este sentido, intentaremos demostrar que la falta de anotación y la cantidad de datos dificultan y ralentizan el análisis lingüístico. Para ello, analizaremos la presencia de distintos neologismos en el CORPES XXI y Twitter, considerando la selección y recuperación de datos para comprobar en qué espacio se muestra más este tipo de formaciones. Este contraste es significativo, puesto que, tanto el CORPES XXI como la red social Twitter se consideran en la actualidad dos grandes recursos para las investigaciones lingüísticas por su vigencia y dimensión.

En el ámbito de la neología, las redes sociales suponen una novedosa fuente de documentación a tener en cuenta para el estudio de la lengua. Gracias a ellas, podemos analizar diferentes tipos de neologismos que constituyen una evidencia clara de la creatividad léxica y de la productividad de los procesos neológicos. Este tipo de voces será posible encontrarlo en fuentes actuales que estén en continua transformación, debido a su carácter novedoso. Por este motivo, el uso de redes sociales supone una valiosa fuente documental a la que acudir, aunque, como veremos, la falta de transparencia en estos medios pueda en ocasiones no servir como herramienta para un análisis lingüístico exhaustivo. Si bien es cierto que la cantidad de material y de datos que puede recopilarse de los discursos en las redes sociales ofrece información sobre los neologismos de una lengua determinada, en este caso el español. A este respecto, la variedad de los datos obtenidos en las búsquedas ayuda a establecer las tendencias predominantes en los procesos de formación de palabras: unidades léxicas nuevas de

significantes y significados, solo de significante, solo de significados o importadas de otro sistema lingüístico. Estas son algunas de las tareas más enriquecedoras para la investigación lingüística, que extrae conclusiones de usos reales de la lengua española de estas novedosas formas de comunicación, fotografiando la lengua y su evolución a cada paso.

## 2. NEOLOGÍA Y NEOLOGISMO

El estudio de la neología en español ha sido abordado desde diferentes perspectivas, aunque con un claro predominio de la corriente descriptiva. Así, en numerosas investigaciones se ofrecen listados de unidades léxicas que podrían considerarse neologismos, sin profundizar y analizar qué es una unidad neológica y qué criterios se siguen para determinar la neologicidad de una nueva palabra. Según el *Diccionario de la lengua española*, un *neologismo* es un 'vocablo, acepción o giro nuevo en una lengua'. Se trata, entonces, de una unidad de nueva creación o de un significado o acepción que antes no se había otorgado a un significante.

Ya desde 1952 Matoré concibe neologismo como «una acepción nueva introducida en el vocabulario de una lengua en una época determinada». Entendiendo como acepción nueva, toda forma creada, toda forma con un nuevo sentido o toda forma con un cambio categorial. Siguiendo su planteamiento, existirían dos tipos de neología: formal y semántica. En el primero, encontraríamos unidades léxicas son significantes nuevos o con significante y significado nuevos. Al segundo pertenecen las palabras existentes que tienen nuevos significados. Mantiene dicha distribución Pottier-Navarro (1979) que insiste en la existencia de un componente nuevo en la neología formal y la inexistencia de este en la semántica. Por su parte, Auger y Rousseau (1978) ofrecen años antes un nuevo tipo de neología: la llamada *neología de préstamo*, donde se incluirían aquellas voces procedentes de otras lenguas que se han adquirido en la lengua receptora. En esta línea también Guerrero Ramos (1995) distingue tres clases de neologismos: neologismo formal, neologismo por préstamo y neologismo semántico. Atendiendo también a la formación de neologismos, Cabré Castellví (2006) propone una clasificación más extensa donde se distinguen cinco tipos: neologismos de forma, neologismos sintácticos, neologismos semánticos, préstamos y otros. Distinta clasificación ha elaborado el grupo *NEOFRAN*, que diferencia entre neologismos morfosemánticos, neologismos sintáctico-semánticos, neologismos morfológicos y neologismos pragmáticos. Dentro de cada tipo, se hace una subclasificación más detallada donde se incluyen fenómenos como la prefijación, la composición, el truncamiento, la metonimia, la conversión, entre otros.

En cualquier caso, Guerrero Ramos (1995) y Vega y Llopart (2017) defienden el uso de *neología* para hacer referencia al proceso de creación y la utilización de la voz *neologismo* para aludir al producto resultante del proceso, es decir, la nueva unidad. Desde

una perspectiva lingüística, establecer esta distinción o limitación en los términos es relevante, pues no siempre se han concebido como tal. Respecto al estudio de la neología o de los neologismos, Cabré Castellví (2006) destaca la importancia de tener en cuenta diferentes perspectivas, no solo lingüísticas, sino también socioculturales o psicológicas, para entender el neologismo como producto, un producto esencialmente sincrónico, pues su creación está vinculada estrechamente a la actualidad, a la evolución y a la dinamicidad de la lengua. También Lavale-Ortiz (2019) apoya los estudios neológicos desde una perspectiva sincrónica. No obstante, destaca la correlación entre sincronía —basada en el ahora de la lengua—, y la diacronía —fundamentada en un punto de vista histórico—. Entendiendo así ambos enfoques como inseparables.

En cuanto a la clasificación de los neologismos, son importantes los estudios de Cabré (2006), Álvarez de Miranda (2009) y Casado Velarde (2015). Álvarez de Miranda centra su investigación en las modalidades del cambio léxico, incluyendo la neología, la pérdida léxica y el cambio semántico; e introduce fenómenos como la discontinuidad, presente en la poligénesis temporal de préstamos y derivados, resurrecciones de carácter exclusivamente literario y el reciclaje intencionado. Más centradas en la clasificación de los neologismos son los estudios de Cabré (2006) y Casado Velarde (2015). La propuesta de Cabré establece la siguiente clasificación para los procedimientos neológicos en español: neologismos formales (por sufijación, por prefijación, por sufijación y/o prefijación, por composición, por composición culta, por lexicalización, por conversión sintáctica, por sintagmación, por siglación, por acronimia y por abreviación), neologismos por variación —presentan una variante formal ortográfica), neologismos sintácticos —implican un cambio de subcategoría gramatical—, neologismos semánticos —conllevan un cambio de significado—, préstamos y otros, donde se integran aquellas voces difíciles de etiquetar.

Difiere de esta clasificación Casado Velarde, que plantea la organización de neologismos de diferente forma: neologismo morfológico o morfosintáctico, donde se incluyen procedimientos como la derivación, composición, parasíntesis y otros como siglas, acrónimos y acortamientos; neologismo semántico —autóctono o alógeno—, calco, neologismo sintáctico-semántico, neologismo léxico por préstamo de otra lengua (préstamo léxico), creaciones fonosimbólicas, creaciones *ex nihilo* y otras.

En relación con los estudios sobre neología, la mayoría de ellos se centran en examinar los distintos procesos de formación de palabras, generalmente en la publicidad, en la prensa y los medios de comunicación. Son interesantes los estudios de Romero Gualda (1993), Guerrero Ramos (2006) y Vega Moreno (2010), que abordan la presencia de creaciones neológicas como un recurso imprescindible en el lenguaje publicitario, así como en los medios de comunicación, donde se pretende captar la atención del público.

Otra línea de investigación muy fructífera es la lexicográfica. La relación existente entre la lexicografía y los criterios para la inclusión de neologismos en los diccionarios generales, así como las propuestas de significados para estas nuevas palabras. Investigaciones como las de Cabré (2004), Sánchez Manzanares (2013), Díaz Hormigo

(2015) o Clavería (2016) apuestan por la revisión de criterios para la incorporación de nuevas formas. Asimismo, tratan de determinar y seleccionar qué información se debe ofrecer en los artículos lexicográficos de las nuevas creaciones.

Teniendo en cuenta lo aquí señalado, entendemos por *neologismo* no solo una acepción nueva de una determinada unidad léxica (neologismo semántico), sino también una forma nueva creada (neologismo formal). Por ello, para nuestra investigación asumiremos el concepto de neologismo planteado en Matoré (1952) y nuestro análisis, por tanto, se establecerá dentro de lo que el autor clasifica como neología formal, donde se ubican unidades léxicas con significantes nuevos o con un significante y un significado nuevos. Para el estudio de los neologismos seleccionados, nos hemos en los criterios de clasificación propuestos por Cabré (2006). Esta propuesta resulta ser la más adecuada en nuestro análisis, ya que permite dar cuenta de la variedad de los datos con los que vamos a trabajar. La selección de los neologismos objeto de estudio responde, además, a los parámetros ofrecidos por la autora, que plantea los signos de inestabilidad formal (morfológica, gráfica, fonética) o semántica como indicativos de que una unidad léxica es nueva. Esta clasificación de los neologismos estudiados nos ayudará a comprobar qué tipo de neologismos aparecen con mayor frecuencia en espacios como Twitter o en el corpus académico CORPES XXI.

### 3. LAS REDES SOCIALES COMO SOPORTE

En los últimos años, la interacción entre la informática y la lingüística ha estado cada vez más presente en numerosas investigaciones. Del resultado de su combinación, surge la denominada lingüística de corpus (LC), un área de la lingüística encargada de llevar a cabo investigaciones con datos obtenidos a partir de corpus (Martín Peris *et al*, 2008). En este sentido, la LC es un sistema o procedimiento cuyo soporte es el corpus, un conjunto de textos, ya sean orales o escritos, obtenidos en condiciones naturales, característicos de una lengua o una variedad, que se almacenan en formato electrónico y se codifican para ser analizados científicamente *a posteriori* (Rojo, 2021). Anteriormente, Sinclair(1991) defiende que un corpus es un conjunto de textos procedentes de lenguaje natural, seleccionados con el fin de caracterizar un estado o variedad lingüística. Basándonos en estas definiciones, cabe preguntarse si las redes sociales constituyen un corpus textual o se consideran meras fuentes o soportes de acceso a la información. Para encontrar una respuesta, debemos saber que, en términos generales, dado que tienden a mostrar usos reales de la lengua, los corpus deben cumplir tres requisitos para que así sean considerados: diseño previo, recuperación de datos y codificación. El diseño es el rasgo diferenciador. Un corpus textual posee un diseño previo que le hace seleccionar el tipo de documentación requerida según sus objetivos, algo de lo que carece un agregador de textos o un simple archivo. Otro rasgo diferenciador es la calidad de sus textos, orales o escritos, que va relacionada

con su codificación y anotación. En algunos corpus, por ejemplo, la falta de diseño, y su falta de codificación textual, hace que se seleccionen y clasifiquen datos que, en la posterior recuperación, pueden perjudicar e invalidar la investigación. Por ejemplo, si consultamos la forma *basuraleza* en el Corpus del español de Mark Davies, descubrimos que ejemplos procedentes de *Soria Noticias*, *Diario Ferrol*, *EL PAÍS*, *EFE*, *Madridpress.com*, entre otros, aparecen clasificados como documentación de US (the United States), cuando en realidad la prensa procede de ES (España). El papel que juega la selección y la codificación del material es fundamental para dar calidad a lo que posteriormente se analizará científicamente.

En este sentido, los conocidos como macrocorpus de lengua digital, formados en su mayoría por redes sociales como Twitter, Instagram o Facebook, carecen de diseño previo en el sentido más estricto. La recuperación de datos en ellos, así como la codificación de la información, tiene limitaciones evidentes. Para que se dé una recuperación selectiva de información, debe existir previamente una buena codificación de textos. La codificación es, precisamente, lo que hace posible el análisis científico de sus rasgos generales y la recuperación selectiva de la información total que contiene el corpus. Es cierto que existen, en la actualidad, diversos APIs (interfaz de programación de aplicaciones) que facilitan el acceso a esta información, ofreciendo al usuario, bajo distintos parámetros previamente seleccionados por el programa, la posibilidad de recuperar datos específicos. Sin embargo, se dan numerosas restricciones de uso y acceso, así como considerables problemas de anotación. Por tanto, si queremos que nuestro análisis sea más fino y ofrezca resultados fiables, no bastará con hacer una descarga automatizada, tendremos que anotar los datos y, sobre ellos, trabajar.

Teniendo esto presente, sabemos que los datos extraídos de las redes sociales no se someten a una codificación, lo que hace imposible una recuperación selectiva de estos. En cambio, como hemos señalado anteriormente, sí ofrece información general, que, según nuestro objetivo, puede aportarnos datos relevantes para nuestro estudio. Por su parte, los corpus permiten mayor fiabilidad en los datos consultados y, en consecuencia, en los resultados. De la misma manera, ofrecen amplias muestras representativas de textos que ayudan a establecer ciertas tendencias lingüísticas.

Los principales estudios dedicados al análisis de redes sociales como recurso para el estudio de la lengua coinciden en que, por su fácil accesibilidad a la información, este tipo de fuentes son un recurso valioso para el análisis lingüístico. Los datos extraídos de las diferentes plataformas son gratuitos, públicos y, además, están digitalizados, por lo que su recopilación es, teóricamente, tarea sencilla. Sin embargo, cuando realizamos una consulta, observamos que los datos extraídos presentan gran variedad, lo que complica la recuperación de un conjunto de documentos estructurados y libres de «ruido» (Martínez Santiago *et al.*, 2001). Por ejemplo, el acceso limitado a datos sociológicos como el sexo, la edad, etc. impiden tener información más detallada al respecto —un inconveniente sobre todo en las investigaciones sociolingüísticas—. Por este motivo, es importante tener en cuenta el tipo de texto que vamos a encontrar presente en las redes sociales y no solo eso, también la aparición de imágenes, gifs,

emoticonos, vídeos, imágenes, etc. Con esto, podemos ver que construir un corpus de textos procedentes de redes sociales como Twitter, Instagram o Facebook no es tarea fácil. Debemos ser conscientes de que las redes sociales nos pueden servir como herramientas para atestiguar un uso determinado, pero no como un corpus lingüístico en sí mismo. El proceso de codificación de los textos, a partir de las consideraciones anteriormente vistas, ayuda a crear un corpus, teniendo siempre presente un diseño y un objetivo. Si nuestra finalidad es conocer si en redes sociales como Twitter, Instagram o Facebook se atestigua un uso real de una voz, estas plataformas serán útiles; en cambio, si lo que queremos es trabajar con datos ingentes extraídos de ellas para sacar conclusiones, deberemos ser conscientes de que, en muchos casos, la recolección de datos será controvertida, lo que impedirá afirmar ciertas generalizaciones de procesos lingüísticos. Para ofrecer mayor certeza en nuestra hipótesis, deberemos acudir a corpus lingüísticos que incluyan una cantidad considerable de este tipo de fuentes documentales digitales, como el Corpus MEsA 2.0, que incluye exclusivamente materiales digitales procedentes de foros, WhatsApp, YouTube, Facebook, etc. Asimismo, el Corpus del Español del Siglo XXI (CORPES XXI) incluye dentro de sus materiales, un tipo específico de texto procedente de redes sociales: Twitter, Blogs, Instagram... En este corpus, la codificación, anotación y desambiguación permite la recuperación de fenómenos lingüísticos, léxicos o gramaticales, gracias a la acotación selectiva de la información.

Como se ha señalado, Internet en general, y las redes sociales en particular, son medios en los que plasmar nuestra naturaleza lingüística, es decir, los rasgos gramaticales o léxicos que nos representan. Tal y como explica Crystal (2002) en Internet encontramos texto *linear*, es decir, unidimensional, propio de la lengua escrita, que permite la inteligibilidad mediante espacios, división textual, etc.; y otro tipo de texto *non-linear* no ordenado de forma tradicional y que puede leerse de forma multidimensional. Esto quiere decir que con la aparición de Internet y de la Web, el usuario ve modificada su percepción textual, acomodándose a la nueva realidad. Lo mismo ocurre con la lengua. Las herramientas de comunicación digital han creado nuevos paradigmas de interacción comunicativa. Crystal defiende la existencia de un *ciberhabla* y explica que Internet es el encargado de registrar nuevas áreas de expresión adaptadas a los nuevos contextos digitales.

En realidad, lo que propone este autor no es la existencia de una nueva lengua propia de Internet sino una adaptación al medio. Pero este suceso no se da con la llegada de Internet. Diariamente, dependiendo del contexto en el que nos encontremos, decidimos qué tipo de lengua utilizar con el fin de adecuarnos a la situación que vivimos, a veces coloquial, otras más formal. En definitiva, este nuevo modelo de comunicación da acceso a un tipo de híbrido lingüístico entre diferentes registros, sin que esta combinación suponga la aparición de un nuevo tipo de lengua.

En el caso de las redes sociales, se ha debatido mucho sobre el tipo de comunicación que se da en ellas. Se ha llegado a hablar incluso de una lengua específica de este canal. Sánchez (2012) señala que este tipo de lenguaje es breve y exacto y tiende a la

globalidad y neutralidad. La autora afirma que en Internet se emplea un español neutro y cercano que facilita la comprensión y la interacción del usuario con el resto de usuarios. La finalidad del español neutro es entender, pero también ser entendido. Otros estudios enfocados en redes sociales son los presentados por Vela Delfa y Cantamutto (2020), que abordan los textos del discurso digital. En ellos se diferencian subclases como mensaje, publicación o comentario. De ello se concluye que dependiendo del tipo de red social que empleemos, la lengua parece tener diferentes propiedades, por ejemplo, basadas en su grado de interactividad. Así lo explican también Cantamutto y Vela Delfa (2016: 317):

«En la pantalla del móvil, de la computadora o en las pestañas de los navegadores, confluyen muchas plataformas de interacción que serán seleccionadas en función de qué (y, por consiguiente, de cómo) quiere comunicar quien acciona la comunicación. (...) Las comunidades discursivas configuran y reconfiguran los medios en su constante necesidad de situarlos en relación con las opciones emergentes en un entorno tan dinámico y cambiante como la comunicación actual».

Destacan también las investigaciones de De Benito Moreno (2021), que presentan la utilidad de Twitter como herramienta para el personal investigador gracias a la facilidad de acceso a los datos, teniendo en cuenta las dificultades que, como hemos visto anteriormente, se presentan en el camino. Asimismo, De Benito Moreno estudia rasgos distintivos en el habla de Twitter, pareciendo ser este un tipo de lenguaje no estandarizado. Por otra parte, cabe señalar la apreciación que hace al sugerir la creación de un corpus cerrado basado en Twitter con el que trabajar. Esta idea es importante, pues confirma lo que venimos señalando: Twitter, tal y como se utiliza en los estudios actuales, es una herramienta, no un corpus. Un corpus seleccionado, codificado y anotado de, por ejemplo, Twitter, abriría las puertas a la investigación de fenómenos lingüísticos desde diferentes perspectivas. Desde otra área, Boyd et al. (2010) apoyan el uso de Twitter como herramienta de análisis y sostienen que la transmisión de información a través de esta plataforma contribuye a que exista un ambiente conversacional en el que las interacciones tienen rasgos propios del diálogo público en un contexto conversacional conjunto. Respecto a la brevedad que caracteriza a Twitter, una investigación muy interesante es la realizada por Mworira (2015), que explica cómo influye la capacidad de síntesis del usuario a la hora de crear voces nuevas que se acomoden al tipo de canal, teniendo que acotar el mensaje para que sea preciso y conciso, dando cabida a creaciones neológicas.

Naturalmente, el tipo de lenguaje que usamos en este medio presenta características propias de la lengua oral, como son su espontaneidad y brevedad, así como de la lengua escrita, por su normatividad, enfocada en la ortografía y la gramática. Sin embargo, Koch y Oesterreicher (2000) señalan que no existe una división estricta entre lo oral y lo escrito, sino un *continuum* que va desde la distancia hasta la inmediatez comunicativa. Ambos apuntan al corto período de tiempo que tiene el usuario para

pensar en el mensaje, rasgos típicamente orales; y a la premeditación o poca espontaneidad en algunos casos, características más propias de la escritura.

Antes de nada, para tratar de entender por qué la lengua en redes funciona como lo hace, debemos distinguir dos clases de usuarios de Twitter e Instagram. En estos canales comunicativos encontramos, por un lado, cuentas institucionales o divulgativas, que ofrecen contenido formal; y, por otro lado, en un tono desenfadado y alejado de la normatividad, cuentas personales que, por lo general, no dan importancia a la escritura. Estas últimas reflejan una comunicación parecida a la que encontrábamos en los antiguos SMS: breves, inmediatos y sintéticos. Esta distinción es relevante, pues existe un condicionamiento social en el habla de las redes sociales que vendría a dirigir la naturaleza del mismo. En los últimos años, se ha incrementado la corrección social a causa de las faltas de ortografía. Cada vez más, lo que escribimos en Twitter o Instagram se convierte en la primera impresión que damos a alguien sobre nosotros. Estudios sociolingüísticos como los ofrecidos por Méndez y Linares (2021) concluyen que en la aplicación Tinder, tener o no faltas de ortografía puede ser determinante para escoger una cita. El 75 % de los hombres considera como negativo tener faltas de ortografía, mientras que en mujeres esta cifra se incrementa a un 86,6 %. Es, por tanto, la lengua y la forma de expresarnos y comunicarnos nuestra carta de presentación, con lo que ganaremos aceptación y respeto por parte de la comunidad. En cualquier caso, podemos determinar que el tipo de lengua presente en redes sociales como Twitter, dependerá del tipo de perfil del usuario y su tipo de registro y forma de expresar. Así, no encontramos el mismo tipo de registro en perfiles públicos con muchos seguidores — como pueden ser cuentas de periódicos, ministerios, instituciones, periodistas, divulgadores, etc. — que en cuentas de pocos usuarios, generalmente, de edad temprana. Cabe señalar que, con esa afirmación, no tratamos de crear una generalización, pues somos conscientes de que lo que determina nuestra ortografía en diferentes ámbitos tiene relación con diversos factores (edad, sexo, nivel educativo, finalidad en el uso de redes, etc.). Naturalmente, un estudio sociolingüístico orientado en redes sociales y ortografía podría ayudarnos a obtener más información sobre el comportamiento del usuario sobre este tema en este canal de comunicación.

## 4. ANÁLISIS

La presente investigación tiene como principales objetivos determinar la relevancia de las redes sociales para el análisis de formaciones neológicas y establecer la representatividad de la red social Twitter comparada con el CORPES XXI, con el fin de descubrir qué herramienta resulta más útil para el estudio de neologismos. Por ello, asumiremos una metodología cuantitativa que ofrezca, gracias al estudio estadístico, la frecuencia de aparición de los neologismos seleccionados en ambos repositorios.

Gracias a estos datos, descubriremos qué herramienta resulta más útil para el estudio de neologismos y en qué espacio es más habitual encontrar este tipo de formaciones.

Para cumplir con los objetivos de la presente investigación se utilizarán diferentes fuentes de datos. En primer lugar, se ha procedido a la selección de diferentes neologismos formales. Por su formación, la búsqueda de neologismos formales es más sencilla que otro tipo de neologismos, como puedan ser los neologismos sintácticos o los neologismos semánticos, que alteran características gramaticales o el significado de la unidad. Los neologismos formales seleccionados para nuestro trabajo han sido *tiktoker*, *cogobernanza*, *presintomático*, *chinofobia*, *galacticida*, *estatuilla*, *pyme*, *machirulo* y *deco*. Todos ellos constituyen una evidencia clara de la creatividad léxica y de la productividad de los procesos morfológicos.

En segundo lugar, se ha extraído la frecuencia de uso de los neologismos tanto en el Corpus del Español del Siglo XXI (CORPES XXI) como en la red social Twitter. Esta última resulta ser la red social más atractiva de analizar por dos motivos fundamentales: la accesibilidad, gracias a que la mayoría de los usuarios tienen cuentas públicas en esta red social; y el API, que permite realizar búsquedas automatizadas gratuitamente desde programas externos, en nuestro caso Rstudio, una aplicación web que permite, a través del lenguaje de programación R, el tratamiento de ingentes cantidades de datos y estadísticas. Por otro lado, el CORPES XXI es un corpus lingüístico de referencia que integra textos, orales y escritos, comprendidos entre los años 2001-2021. La versión con la que nosotros vamos a trabajar es la 0.94, que cuenta con más de 350 millones de formas ortográficas. Para nuestro análisis hemos elegido el CORPES XXI basándonos en las características del corpus mismo. Los textos que incorpora se seleccionan de acuerdo con unos parámetros específicos. Asimismo, la codificación y anotación que posee el corpus permite la recuperación de sus datos desde cualquiera de esos parámetros, lo que convierte al CORPES XXI en una herramienta muy útil para la investigación lingüística. En contraste, comprobamos que su actualización anual dificulta la búsqueda y la observación de algunos neologismos. El corpus carece del carácter dinámico que sí tiene una red social como Twitter. Un hecho que, como veremos con los siguientes gráficos, pone de manifiesto la importancia de utilizar herramientas como las redes sociales en este tipo de investigaciones con el fin de observar la evolución de la palabra, aunque somos conscientes de las carencias que una herramienta de estas características tiene.

Como se ha mencionado anteriormente, para la selección de los neologismos formales en este estudio se ha tenido en cuenta la formación. Asimismo, ha sido relevante para la selección la frecuencia de uso por parte de los hablantes, es decir, se han seleccionado formas con cierta continuidad para trabajar con datos significativos que ayudaran a establecer las tendencias. La clasificación de los neologismos escogidos para la investigación se ha basado en la ofrecida por Cabré (2006) para clasificar los neologismos formales: sufijación, prefijación, sufijación y prefijación, composición, composición culta, siglación, acronimia, abreviación y lexicalización. Aunque esta autora señala también las formaciones sintácticas o de sintagmación, hemos descartado

para este estudio su incorporación por ser ambas formaciones más complejas en la búsqueda, al alterar características gramaticales o el significado de la unidad, cuestiones que en este tipo de repositorios no se distinguen. Tras la selección de los neologismos escogidos para el estudio, se ha procedido a identificar las unidades teniendo en cuenta diferente información: fecha, categoría, tipo de formación, definición, red social y contexto. Teniendo en cuenta la tipología de Cabré y los datos recogidos en el CORPES XXI y en Twitter, mostramos a continuación las características de cada voz, así como el número total de casos documentados en ambos espacios:

- A. Neologismos creados por sufijación. Este tipo de voces están creadas a partir de un sufijo añadido a una palabra base. Ejemplos: *beticismo*, *surfear*, *chavismo*, *viralizar*, *instagrameable*, *gasolinazo*, *brexitero*, *ruidazo*, *enchufable*, *bacheletismo*, *frentista*, etc. Es evidente que en el español actual existen mayor número de creaciones formadas a partir de la sufijación. Por su riqueza léxica y su facilidad de creación e, incluso, por su transparencia, este mecanismo es el preferido por los hablantes, que crean nuevas palabras a través de la sufijación sin gran esfuerzo.

Voz	tiktoker
Fecha	10 mayo de 2022
Categoría	Sustantivo masculino o femenino
Tipo de formación	Sufijación
Definición	Persona que crea y publica contenidos en la red social TikTok®.
Red social	Twitter
Contexto	«ojo que se descuidan un segundo y yo ya m dedico a ser tiktoker full time»

Periodo	CORPES XXI	TWITTER
2006 <sup>1</sup> – 2010	0 casos	4 casos
2011 – 2015	0 casos	10 casos
2016 – 2020	1 casos	2.185.946 casos
2021 <sup>2</sup> – 2022	1 casos	4.108.846 casos

Tabla 1. Frecuencia de uso de tiktoker en CORPES y Twitter

- B. Neologismos creados por prefijación. Este tipo de neologismos se forman añadiendo un prefijo a una palabra base. Ejemplos: *autoexamen*, *semipresencial*, *recen-tralizar*, *resaltante*, *sobreprotección*, etc. Al contrario de lo ocurre con las formas

1. La búsqueda de Twitter comienza el 23 de marzo de 2006, fecha en la que se publicó el primer tuit.
2. La última búsqueda data del 16 de junio de 2022.

creadas gracias a la adición de un sufijo, las palabras por prefijación se consideran de menor productividad en español. Este tipo de componente es tan versátil que, incluso, muchas investigaciones afirman la inexistencia de los prefijos, considerándolos preposiciones o prefijoides. Como apunta Martín García (2020), establecer un listado o clasificación es complicado, pues existen en español elementos que comparten propiedades con los elementos compositivos y con los prefijos. Y, por ende, no pueden encasillarse exclusivamente en uno de los dos grupos. Si bien es cierto que también Martín García (2020) establece algunas pautas que ayudan a la identificación de estos elementos: modificación de la base, posibilidad de ser argumentos del núcleo, combinación con sufijos, significado léxico específico, etc. Por su parte, el Observatori de Neologia (OBNEO) clasifica elementos como *ante-*, *fuera-*, *pro-*, *bajo-*, *mal-*, *sin-*, *bien-*, *menos-*, *sobre-*, *casi-*, *no-*, *tras-*, *contra-*, *plus-*, *ultra*, como prefijos, otorgando a las palabras resultantes la condición de palabras derivadas y no compuestas. Ejemplos: *bienpensante*, *ultrarápido*, *malinterpretación*, *sinvivir*, etc. Sin duda, el contenido poco preciso de los prefijos favorece su comportamiento ambivalente.

Voz	cogobernanza
Fecha	09 de mayo de 2022
Categoría	Sustantivo femenino
Tipo de formación	Prefijación
Definición	Gobierno basado en la cooperación y el acuerdo de diferentes poderes.
Red social	Twitter
Contexto	«Jajajajajaja A esa le quedan dos telediarios y lo sabes. Estoy muuuuuuy tranquila. Que se hayan introducido modificaciones ideológicas es algo político pero vendrán los rodillos jurídicos como vinieron con los Estados de Alarma, el cierre del Parlamento, la cogobernanza...»

Periodo	CORPES XXI	TWITTER
2006 – 2010	0 casos	2 casos
2011 – 2015	0 casos	29 casos
2016 – 2020	3 casos	55.170 casos
2021 – 2022	0 casos	50.868 casos

Tabla 2. Frecuencia de uso de cogobernanza en CORPES y Twitter

- C. Neologismos creados por sufijación o prefijación. Estas voces se han creado por prefijación y sufijación. En estos casos, la interferencia entre ambos impide deter-

minar qué proceso ha sido el último en realizarse, por tanto, podríamos incluirlos bajo prefijación o sufijación indistintamente. Ejemplos: *enloquecer*, *desvainar*, etc.

Voz	presintomático
Fecha	08 de febrero de 2022
Categoría	Sustantivo femenino y masculino Adjetivo
Tipo de formación	Prefijación o sufijación
Definición	Persona que padece una enfermedad sin síntomas. Que es anterior al momento en que se presentan síntomas de una enfermedad. Que padece una enfermedad que todavía no presenta síntomas.
Red social	Twitter
Contexto	«Sin síntomas no contagias, no es un lentivirus como el VIH, otra cosa es presintomático (un poco de garraspera en la garganta, moquillos, tos, febrícula). Tu en 2019 faltabas al trabajo por una gripe invisible??»

Periodo	CORPES XXI	TWITTER
2006 – 2010	1 caso	1 caso
2011 – 2015	1 caso	120 casos
2016 – 2020	3 casos	3.384 casos
2021 – 2022	0 casos	2.689 casos

Tabla 3. Frecuencia de uso de *presintomático* en CORPES y Twitter

- D. Neologismos creados por composición. Estos neologismos son creados a partir de dos raíces. Ejemplos: *cochebomba*, *coronavirus*, *rojinegro*, *fotomulta*, *googleadicto*, etc.

Voz	chinofobia
Fecha	10 de mayo de 2022
Categoría	Sustantivo femenino
Tipo de formación	Composición
Definición	Aversión a China o a las personas de nacionalidad china.
Red social	Twitter
Contexto	«Una pena la mala experiencia pero la verdad Xiaomi es buena marca, su único pecado es ser chino y ya sabemos como esta la chinofobia y rusofobia alborotada últimamente»

Periodo	CORPES XXI	TWITTER
2006 – 2010	0 casos	82 casos
2011 – 2015	0 casos	1.109 casos
2016 – 2020	0 casos	3.335 casos
2021 – 2022	0 casos	1.107 casos

Tabla 4. Frecuencia de uso de chinofobia en CORPES y Twitter

- E. Neologismos creados por composición culta. Estas voces están formadas a partir de distintos procesos: forma prefijada culta + forma sufijada culta, forma prefijada culta + raíz o raíz + forma sufijada culta. Ejemplos: *eurófilo*, *simpaticoide*, *filonazi*, *ciberactivista*, etc.

Voz	galacticida
Fecha	17 de enero de 2020
Categoría	Adjetivo
Tipo de formación	Composición culta
Definición	Dicho de un árbitro: que es perjudicial para el equipo del Real Madrid.
Red social	Twitter
Contexto	«Un árbitro, por el hecho de hacer bien su trabajo, es bautizado como el 'Galacticida'. Con semejantes PRESIONES... ¿Cómo va a existir imparcialidad? ¿Cómo se van a atrever a pitarle neutralmente al Madrid? Siempre a su favor, y así se quitan de líos»

Periodo	CORPES XXI	TWITTER
2006 – 2010	0 casos	3 casos
2011 – 2015	0 casos	18 casos
2016 – 2020	0 casos	18 casos
2021 – 2022	0 casos	0 casos

Tabla 5. Frecuencia de uso de galacticida en CORPES y Twitter

- F. Neologismos creados por lexicalización. Formaciones creadas a partir de la lexicalización de una forma flexiva, normalmente de una forma del paradigma verbal. Ejemplo: *removida*, *encuadernado*, *hormigonado*.

Voz	estatuilla
Fecha	10 de mayo
Categoría	Sustantivo femenino

Tipo de formación	Lexicalización
Definición	Premio, trofeo o galardón
Red social	Twitter
Contexto	«Olvidate de la estatuilla, el señor es un gran actor, y un productor exitoso, puede tener la casa que quiera, pero en este sexenio, es pecado ser exitoso, quieren todos mediocres»

Periodo	CORPES XXI	TWITTER
2006 – 2010	278 casos	4.814 casos
2011 – 2015	181 casos	185.491 casos
2016 – 2020	103 casos	318.675 casos
2021 – 2022	3 casos	90.295 casos

Tabla 6. Frecuencia de uso de estatuilla en CORPES y Twitter

- G. Neologismos creados por siglación. Estas palabras están formadas por las letras principales que corresponden a la sigla de la que procede. Ejemplos: *un dvd* (DVD), *el mype* (Montajes y Proyectos Electrotécnicos), *el dedeté* (Dicloro Difenil Tricloroetano), *el pepe* (Partido Popular), etc.

Voz	pyme
Fecha	10 de mayo de 2022
Categoría	Sustantivo femenino
Tipo de formación	Siglación
Definición	Pequeña o mediana empresa con un número reducido de trabajadores y con un volumen de ingresos bajo.
Red social	Twitter
Contexto	«Literal que tengo una pyme pero gasto como si tuviera una multinacional 🤔»

Periodo	CORPES XXI	TWITTER
2006 – 2010	289 casos	48.106 casos
2011 – 2015	122 casos	1.189.297 casos
2016 – 2020	129 casos	2.514.348 casos
2021 – 2022	1 caso	1.327.433 casos

Tabla 7. Frecuencia de uso de pyme en CORPES y Twitter

- H. Neologismos creados por acronimia. Estos neologismos están formados por la combinación de segmentos de palabras que forman una unidad. Ejemplos: *espan-english* (*español y english*), *infodemia* (*información y epidemia*), *turismática* (*informática aplicada al turismo*).

Voz	machirulo,a
Fecha	10 de mayo de 2022
Categoría	Sustantivo masculino o femenino Adjetivo
Tipo de formación	Acronimia
Definición	Persona, generalmente hombre, con actitudes machistas e irrespetuosas. Propio de los machirulos o relativo a ellos.
Red social	Twitter
Contexto	«He puesto una foto mia en el feed de Instagram y eso ha sido argumento suficiente para que un machirulo me empiece a hablar»

Periodo	CORPES XXI	TWITTER
2006 – 2010	0 casos	14 casos
2011 – 2015	1 caso	15.661 casos
2016 – 2020	4 casos	1.086.296 casos
2021 – 2022	0 casos	281.961 casos

Tabla 8. Frecuencia de uso de machirulo en CORPES y Twitter

- I. Neologismos creados por abreviación. Formados por el acortamiento de una palabra o de un sintagma. Ejemplo: *protá* (*protagonista*), *cel* (*celular*), *eco* (*ecológico*), *neocon* (*neoconservador*).

Voz	deco
Fecha	10 de mayo de 2022
Categoría	Sustantivo femenino Adjetivo
Tipo de formación	Abreviación
Definición	Decoración de un espacio. Relativo a la decoración.
Red social	Twitter
Contexto	«Del aguamarina al azul verdoso pasando por el azul Tiffany o el verdigris, el color turquesa aporta a la deco serenidad y frescura. ¡Nos encanta!»

Periodo	CORPES XXI	TWITTER
2006 – 2010	0 casos	570.905 casos
2011 – 2015	0 casos	9.776.633 casos
2016 – 2020	0 casos	12.202.609 casos
2021 – 2022	0 casos	9.182.073 casos

Tabla 9. Frecuencia de uso de deco en CORPES y Twitter

Como vemos, *a priori*, las redes sociales son un valioso y potente recurso para la investigación lingüística. Voces como *tiktoker*, *machirulo* o *galacticida* son un tipo de formaciones neológicas casi exclusivas de espacios de habla coloquial, con un tono más relajado que favorezca su aparición. Así lo vemos en el número de apariciones en Twitter, donde los casos se multiplican. Teóricamente, gracias a esta información, descubrimos que palabras como las aquí propuestas, lejos de la improductividad que presentan los resultados del CORPES XXI, se utilizan con frecuencia. En cambio, antes de expresar tal afirmación, debemos reflexionar sobre diferentes cuestiones.

En primer lugar, tras la recuperación de datos, apreciamos que estos nueve neologismos se documentan en Twitter con mayor frecuencia que en el CORPES XXI. Para establecer el volumen de palabras en Twitter, tendríamos que conocer el número total de formas, dato que, por su continuo crecimiento, no está disponible. Sin embargo, podemos hacer un ejercicio para estimar la relación entre las dimensiones de Twitter y el CORPES XXI. Tomaremos como referencia los diez sustantivos más frecuentes que aparecen en el CORPES XXI (*año*, *vez*, *día*, *parte*, *país*, *tiempo*, *vida*, *persona*, *mujer*, *momento*) y utilizaremos su frecuencia, tanto en el CORPES XXI como en Twitter, como estimación de la proporcionalidad de casos en ambos repositorios.

	CORPES XXI	TWITTER	Cociente T/C
2006 – 2010	753.843 casos	261.779.010 casos	347,26 casos
2011 – 2015	575.708 casos	6.058.909.401 casos	10.524,28 casos
2016 – 2020	363.944 casos	7.286.573.909 casos	20.021,14 casos
2021 – 2022	7.542 casos	3.008.087.239 casos	398.844,77 casos

Tabla 10. Frecuencia absoluta de las formas en CORPES y Twitter

Si comparamos el cociente entre el total de las ocurrencias en Twitter y el total de las ocurrencias en el CORPES XXI de estos diez sustantivos, vemos, por ejemplo, que durante el periodo 2006-2010, hubo 347 ocurrencias en Twitter por cada ocurrencia en el CORPES XXI. Vemos a continuación lo que ocurre con el caso de *pyme*:

	CORPES XXI	TWITTER	T/C
2006 – 2010	289 casos	48.106 casos	166,46
2011 – 2015	122 casos	1.189.297 casos	9.748,34
2016 – 2020	129 casos	2.514.348 casos	19.491,07
2021 – 2022	1 caso	1.327.433 casos	1.327.433,00

Tabla 11. Frecuencia absoluta de pyme en Twitter y CORPES

Teniendo en cuenta la Tabla 10 —nuestra referencia de comparación de la dimensión de Twitter y CORPES—, vemos cómo *pyme* tenía, en el periodo 2006-2010, 166 ocurrencias en Twitter por cada ocurrencia en el CORPES XXI. Este valor es menor que el valor que tomamos como referencia (347,26). En cambio, en los periodos posteriores (2011-2015 y 2016-2020), los respectivos valores se aproximan más a los valores de la Tabla 10. Sin embargo, en el último periodo, que abarca 2021-2022, *pyme* está sobrerrepresentado, teniendo en cuenta los datos de referencia.

Con todo, queremos mostrar que el hecho de que palabras de reciente creación como las aquí estudiadas (*tiktokker*, *pyme*, *machirulo*, etc.) aparezcan representadas en la red social Twitter con mayor frecuencia que en el CORPES XXI no nos aporta un dato relevante sobre la preferencia de uno u otro repositorio. Naturalmente, la frecuencia de aparición siempre será mayor en Twitter pues, como hemos apuntado, estimamos por cada forma de CORPES XXI 347,26 formas en Twitter (2006-2010), 10524,28 formas (2011-2015), 20021,14 formas (2016-2020) y 398844,77 formas (2021-2022).

En segundo lugar, respecto al diseño previo y la calidad de sus textos, que va relacionada con su codificación y anotación, Twitter presenta un claro obstáculo: la anotación. Como hemos visto anteriormente, la forma *deco* no siempre se emplea como abreviación de decoración; en otras ocasiones forma parte de *art deco*, un movimiento de diseño que se popularizó a partir de 1920 hasta 1939. Naturalmente, la consulta en Twitter nos devuelve todas las apariciones de la voz *deco*, sin que sepamos cuándo podría pertenecer a este movimiento o cuándo se refiere al acortamiento. En otros corpus lingüísticos como el CORPES XXI, donde se permite la búsqueda restringiendo el elemento anterior, o posterior, de la palabra que buscamos, sí podríamos acotar la búsqueda para que el CORPES XXI no devuelva aquellos casos donde aparezca antepuesta a *deco* la voz *art*. Sin duda, hacer este tipo de búsquedas resulta clave para las investigaciones lingüísticas, sobre todo para aquellas que trabajan con datos y análisis cuantitativos. La ausencia de estas características impide realizar una búsqueda acotada que puede acabar incorporando datos que carecen de interés para nuestro estudio.

En tercer lugar, apuntando a la selección de material, insistimos en la importancia del estudio previo a la recopilación de textos. El vigor con el que se selecciona documentación en el CORPES XXI hace que un autor de Argentina que publica en un periódico de España, o cuyo tuit tiene localización de España, aparezca clasificado en

Río de la Plata y no en España. Esto ocurre gracias a un estudio de investigación por parte de los lingüistas y filólogos que trabajan en este corpus, quienes se encargan de estudiar el origen del autor, así como las características del texto para su inclusión en el corpus. Esta labor no se lleva a cabo en Twitter, donde la extracción de información de datos en bruto nos devuelve todo tipo de contenido sin previa selección. Así, por ejemplo, el autor argentino al que hacíamos referencia anteriormente podrá escribir un tuit desde España, que la API no diferenciará su origen, lo que puede dificultar los resultados de las búsquedas lingüísticas en cuestiones de variación dialectal.

En cuarto, y último lugar, cabe señalar la aparición de retuits dentro de los tuits seleccionados por Twitter. De esta manera, podríamos llegar a encontrar un número elevado de casos que, en realidad, pertenecen a un solo texto. Sí es interesante, en este aspecto, acotar la búsqueda de datos para recuperar exclusivamente los tuits creados originalmente, sin que contabilicen los retuiteados o citados.

## 5. CONCLUSIONES

Son muchas y obvias las ventajas de la incorporación de redes sociales como herramienta de estudio para la investigación lingüística. Su carácter dinámico hace que los resultados estén cambiando constantemente. Si bien es cierto que este tipo de recursos permite la recopilación de información en un periodo de tiempo determinado o, incluso, en un área geográfica específica, creemos que la falta de anotación y la cantidad de resultados, en algunos casos inabarcable, dificultan y ralentizan el análisis lingüístico. La anotación, por ejemplo, nos permitiría recuperar información de una palabra que pueda presentarse en dos categorías gramaticales diferentes. Algo que, con el API de Twitter, es imposible lograr. Si la palabra en cuestión presenta miles, o millones, de casos, el hecho de no distinguir su categoría gramatical dificultará su análisis. De esta manera, deberíamos, previo al análisis, estudiar exhaustivamente caso por caso para comprobar si la información recuperada es válida para nuestro estudio. Un hecho que dificulta nuestro trabajo.

Asimismo, creemos primordial establecer el objetivo de cada búsqueda, y análisis, para determinar qué corpus o herramienta resulta ser de interés. Si el objetivo es estudiar formaciones neológicas, podemos servirnos de herramientas, o corpus lingüísticos, que nos permitan acceder a este tipo de información (Twitter y CORPES, de forma complementaria). En cambio, si nuestro objetivo está enfocado en la oralidad, sabemos que las redes sociales no son una herramienta útil, al contrario que el CORPES u otros corpus lingüísticos como CORILGA, COSER o ESLORA, que sí presentan textos orales. Lo que pretendemos evidenciar es la importancia de completar la investigación con diferentes herramientas o corpus lingüísticos, siempre prestando atención a la finalidad. En la actualidad, el repertorio de corpus y herramientas que permiten la recuperación de información es amplio y especializado, por lo que podremos usar

un corpus que se ajuste a nuestros criterios, y que permitirá obtener resultados más veraces y, en consecuencia, reforzar la investigación empírica.

## REFERENCIAS BIBLIOGRÁFICAS

- ÁLVAREZ DE MIRANDA, P. (2009). «Neología y pérdida léxica», en E. de Miguel (ed.), *Panorama de la lexicología*. Barcelona: Ariel, págs. 133-158.
- BOYD, D., GOLDBER, S. & LOTAN, G. «Tweet, Tweet, Retweet: Conversational Aspects of Retweeting on Twitter», 2010 43rd *Hawaii International Conference on System Sciences*, 2010, pp. 1-10. Recuperado de doi: 10.1109/HICSS.2010.412.
- CABRÉ, M. T. (2006). La clasificación de neologismos: una tarea compleja. *Alfa Revista de Lingüística*, 50(2): 229-250. Recuperado de [https://www.researchgate.net/publication/26623352\\_La\\_clasificacion\\_de\\_neologismos\\_una\\_tarea\\_compleja](https://www.researchgate.net/publication/26623352_La_clasificacion_de_neologismos_una_tarea_compleja)
- CABRÉ, M. T. (2004). De los diccionarios a los bancos de conocimiento: nuevas herramientas para el traductor. *Las palabras del traductor: actas del II Congreso «El Español, Lengua de Traducción»* / coord. por Luis González, Pollux Hernández, 2004, págs. 29-55
- CANTAMUTTO, L. y VELA DELFA, C. (2016). «El discurso digital como objeto de estudio: de la descripción de interfaces a la definición de propiedades». *Aposta* 69: 296-323.
- CANTAMUTTO, L. y VELA DELFA, C. (2020). Mensajes, publicaciones, comentarios y otros textos breves de la comunicación digital. *Tonos digital*, 38, 1-27.
- CASADO VELARDE, M. (2015). *La innovación léxica en el español actual*. Síntesis: Madrid.
- CRYSTAL, D. (2001), *El Lenguaje e Internet*, trad. Pedro de Tena, Madrid: Cambridge University.
- DAVIES, M. Corpus del Español. <http://www.corpusdelespanol.org/> [Consulta: 15/06/2022].
- DE BENITO MORENO, C. (2021). «'The Spanish of the Internet': Is That a Thing? Discursive and Morphosyntactic Innovations in Computer Mediated Communication». En *English and Spanish in Interaction*, eds. D. Pérez, M. Hundt, J. Kabatek y D. Schreier, 258-286. Cambridge: Cambridge University Press.
- DÍAZ HORMIGO, M. T. (2015). Neología aplicada y lexicografía: para la (necesaria) actualización de las entradas de los elementos de formación de palabras en diccionarios generales. *Revista de lingüística y lenguas aplicadas*, Nº. 10, págs. 12-20.
- GONZÁLEZ, F. A. (2018). Big data y corpus lingüísticos para el estudio de la densidad léxica. *Skopos* 9, 107-122. Recuperado de <https://helvia.uco.es/xmlui/handle/10396/19125>.
- GUERRERO RAMOS, G. (1995). *Neologismos en el español actual*. Madrid: Arco/Libros.
- GUERRERO RAMOS, G. (2006). «Uso ¿o abuso? de los neologismos en la prensa». *Corcillvm / Estudios de traducción, lingüística y filología dedicados a Valentín García Yebra*. Eds. Consuelo Gonzalo y Pollux Hernández. Madrid: Arco/Libros, 871-88.
- HÜTT, H. H. (2012). Las redes sociales: una nueva herramienta de difusión. *Reflexiones*, 91 (2). 121-128. Recuperado de <https://www.redalyc.org/pdf/729/72923962008.pdf>
- KILGARRIFF, A. & GREFFENSTETTE, G. (2003). «Introduction to the special issue on the web as corpus», *Computational Linguistics*, 29 (3), 333-348.
- LAVALE-ORTIZ, R. M. (2019). Bases para la fundamentación teórica de la neología y el neologismo: la memoria, la atención y la categorización. *Círculo de Lingüística Aplicada a la Comunicación*, 80, 201-226.

- MARTÍN GARCÍA, J. (2020). La periferia izquierda de la palabra: prefijos y elementos compositivos. *Nueva revista de filología hispánica*, LXVIII, 3, 523–549.
- MATORÉ, G. (1952). «Le néologisme: naissance et diffusion», en *he Francais Moderne*, 1, págs. 87-92.
- MWORIA, M. H. (2015). Use of english neologisms in social media: a case of twitter language in kenya. Trabajo de grado, Máster en Humanidades. University of Nairobi. Recuperado de [http://erepository.uonbi.ac.ke/bitstream/handle/11295/93589/Mworia\\_Use%20Of%20English%20Neologisms%20In%20Social%20Media%20A%20Case%20Of%20Twitter%20Language%20In%20Kenya.pdf?sequence=3&isAllowed=y](http://erepository.uonbi.ac.ke/bitstream/handle/11295/93589/Mworia_Use%20Of%20English%20Neologisms%20In%20Social%20Media%20A%20Case%20Of%20Twitter%20Language%20In%20Kenya.pdf?sequence=3&isAllowed=y)
- Observatori de Neologia, OBNEO. Metodología del trabajo en neología: criterios, materiales y procesos. Papers de l'IULA. Sèrie Monografies 9. Universitat Pompeu Fabra: Barcelona, 2004.
- POTTIER NAVARRO, H. (1979). «La néologie en espagnol contemporain», en *Les langues néolatines*, 229-230, págs. 148-172.
- Real Academia Española. Banco de datos (CORPES XXI). Corpus del Español del Siglo XXI (CORPES). <http://www.rae.es>. [Consulta: 15/06/2022]
- Real Academia Española. (2014). *Diccionario de la lengua española*. Recuperado de <https://dle.rae.es/?w=diccionario>
- ROJO, G. (2021). *Introducción a la lingüística de corpus en español*. London: Routledge.
- ROUSSEAU, J. & AUGER, P. (1978). *Methodologie de la recherche Québec: Éditeur officiel du Québec*.
- SÁNCHEZ, M. (2012). Las redes sociales, ¿nuevos soportes para el estudio de la lengua? Publicado en *Cum corde et in nova grammatica. Estudios ofrecidos a Guillermo Rojo*, Universidad Santiago de Compostela: Servizo de Publicacións e Intercambio Científico.
- SÁNCHEZ MANZANARES, C. (2013). Valor neológico y criterios lexicográficos para la sanción y censura de neologismos en el diccionario general. *Sintagma: Revista de lingüística*, Vol. 25, págs. 111-125.
- SINCLAIR, J. (1991). *Corpus, concordance, collocation*. Oxford: Oxford University Press.
- VEGA MORENO, É. (2010). Las creaciones neológicas en el lenguaje de la publicidad. Cádiz: Universidad de Cádiz.