

Artículo de Investigación

# Impacto y audiencia de la plataforma en línea de un diccionario del lenguaje emergente en las redes sociales

## Impact and audience of an online dictionary of the emerging language in social media

Antonio Rico-Sulayes<sup>1</sup>: Universidad de las Américas Puebla, México.  
[antonio.rico@udlap.mx](mailto:antonio.rico@udlap.mx)

Fecha de Recepción: 11-05-2024

Fecha de Aceptación: 01-07-2024

Fecha de Publicación: 03-07-2024

### Cómo citar el artículo (APA 7<sup>a</sup>):

Rico Sulayes, A. (Año). Impacto y audiencia de la plataforma en línea de un diccionario del lenguaje emergente en las redes sociales [Impact and audience of an online dictionary of the emerging language in social media]. *European Public & Social Innovation Review*, 9, 01-17.  
<https://doi.org/10.31637/epsir-2024-302>

### Resumen:

**Introducción:** Más allá de la dificultad de crear un diccionario con un contenido novedoso o valioso que le granjee usuarios, la divulgación en línea de este tipo de obra representa un reto en sí. Este artículo presenta los desafíos y logros en la construcción y divulgación en línea del ERES | Diccionario del Español Especializado en Redes Sociales. **Metodología:** Con un enfoque diferencial, capturando términos que no existen en los diccionarios generales, el ERES incluye términos de diversas áreas del conocimiento o del quehacer humano, así como palabras que se suelen etiquetar como informales, populares o groseras. Este diccionario cuenta con una plataforma en línea, alojada en un dominio y un servidor dedicados, y tiene interfases de escritorio y móvil. **Resultados:** Actualmente, el diccionario ofrece en línea 821 entradas, 1.082 significados y 1.368 ejemplos. Estos contenidos captan usuarios en todos los países de habla hispana y en varios otros países con numerosos hablantes del español como L1 y L2. **Conclusiones:** A pesar de tener una audiencia internacional y diversa, con los Estados Unidos como segundo país con mayor número de usuarios, el posicionamiento del diccionario en los motores de búsqueda sigue siendo un reto en el que hay que trabajar constantemente.

<sup>1</sup> Antonio Rico-Sulayes: Universidad de las Américas Puebla, México.

**Palabras clave:** diccionarios electrónicos; lexicografía digital; divulgación en línea; motores de búsqueda; audiencia internacional; comunicación digital; neologismos; lexicografía diferencial.

#### **Abstract:**

**Introduction:** Beyond the difficulty in producing a dictionary with novel and valuable content that attracts users, disseminating it online is a challenge in itself. This article shows the issues and achievements in the construction and online dissemination of ERES | Diccionario del Español Especializado en Redes Sociales. **Methodology:** Through a differential approach, including terms that are not contained in general dictionaries, ERES targets words in various disciplines, as well as words often labeled as colloquial, slang or vulgar. The dictionary has its own website, hosted in a dedicated server and domain, with desktop and mobile interfaces. **Results:** Currently, the dictionary has 821 entries, 1.082 senses, and 1.368 examples. The website gets users from all Spanish-speaking countries and other countries with numerous speakers of Spanish as an L1 or L2. **Conclusions:** Despite of its diverse, international audience – the US is the second country with the most users – positioning the dictionary in search engines remains a challenge that has to be constantly addressed.

**Keywords:** electronic dictionaries; digital lexicography; online dissemination; search engines; international audience; digital communication; neologisms; differential lexicography.

## **1. Introducción**

Partir del uso del lenguaje en las redes sociales para la creación de un diccionario no está exento de cuestionamientos. Algunas de las características inherentes del uso de la lengua en dicho contexto hacen que muchos hablantes rechacen este uso. Por un lado, la comunicación mediada por computadora, en general, se caracteriza por una constante innovación lingüística (Cameron y Panović, 2014). Motivada por diferentes factores, como la imitación de la lengua oral, los estudios lingüísticos muestran que los usuarios de los medios de comunicación digital han encontrado en ésta un laboratorio de experimentación lingüística. Por otro lado, otra característica de este tipo de comunicación es la relajación de las normas de escritura, que está bien documentada por los estudiosos de la lengua (Fernández de Molina, 2015). Aunque estas dos características pueden generar rechazo, la primera de ellas en particular hace del lenguaje de la comunicación digital una fuente de nuevos términos y conceptos que surgen con rapidez. La velocidad de su aparición (y muchas veces de su desaparición) es una de las razones por las que estos términos están ausentes en los grandes diccionarios, ya sea por la dificultad de capturarlos, ya sea por la aversión que generan entre los miembros de las instituciones que financian estas obras.

Las nuevas palabras, o neologismos, que surgen en las redes sociales pueden ser más o menos aceptadas por los hablantes, los expertos de la lengua o las instituciones, pero su presencia es insoslayable. Esto es cierto cuando menos para los usuarios de este medio de comunicación, los cuales incluyen a una gran proporción de la población mundial, y a una proporción aún mayor entre los hablantes del español (Adame, 2019). Además de lo anterior, si se tiene en cuenta que el español es una de las lenguas más utilizadas mundialmente en la red (Instituto Cervantes, 2022), la importancia de que el significado de estos términos pueda ser consultado en alguna fuente fiable por los hablantes resulta evidente.

Este artículo presenta el diseño de un recurso electrónico que atiende la codificación lingüística de los nuevos términos que aparecen en las redes sociales, de manera primordial, y en general, de los que no se encuentran en los grandes diccionarios. De manera más

concreta, este trabajo intenta contestar a la pregunta de si las redes sociales son una fuente de términos léxicos novedosos o valiosos con los que se pueda crear un diccionario que resulte útil para los usuarios en la red, logrando posicionarse en los motores de búsqueda más importantes.

Para intentar responder de manera afirmativa al cuestionamiento anterior, el diseño del diccionario aquí descrito, el *ERES | Diccionario del Español Especializado en Redes Sociales*, está guiado por el enfoque de la lexicografía diferencial y especializada. Partiendo de ese enfoque, se ha construido, en primer lugar, un corpus de datos con comunicaciones digitales para hacer una extracción paulatina de los términos antes descritos. Una vez identificados dichos términos, en segundo lugar, se ha generado una planta para el diccionario que define las reglas de selección y captura de los elementos léxicos. Un tercer paso ha sido la adaptación de una versión de la plataforma en línea para diccionarios electrónicos *Lexonomy* (2024a), financiada por la Comunidad Europea bajo el proyecto Elexis (2024), y su colocación en un servidor y un dominio dedicados. Finalmente se ha trabajado en el posicionamiento de dicha plataforma en los motores de búsqueda más importantes.

## 2. Metodología

Existen tres áreas primordiales en las que se ha tenido que trabajar al construir un diccionario en línea dedicado a las palabras que surgen en redes sociales e intentar posicionarlo en los principales motores de búsqueda en la red. La más importante de estas áreas ha sido el procedimiento que han seguido los colaboradores del diccionario para contribuir al mismo. Este procedimiento incluye la selección de temáticas para extraer masivamente comentarios en redes sociales, el procesamiento de dichos comentarios y la extracción de neologismos y terminología, todo esto con un enfoque de lexicografía diferencial. La primera parte de esta sección está dedicada a presentar este procedimiento.

Además, este apartado presenta las principales adaptaciones que se realizaron a la plataforma *Lexonomy* en 2021 para desplegar el diccionario en un servidor y un dominio dedicados. Finalmente, se discuten las acciones principales que se han realizado para responder a las exigencias de los motores de búsqueda más importantes con la intención de posicionar el diccionario en los mismos.

### 2.1. Un proyecto de lexicografía diferencial

El concepto de lexicografía diferencial tiene al menos dos acepciones, particularmente entre los creadores de diccionarios del español. El significado más común para el calificativo *diferencial* en la lexicografía del español tiene que ver con la elaboración de diccionarios para las variantes diatópicas (Zimmerman, 2018), es decir, de las variantes del español que se hablan en diferentes regiones. En esta acepción, la lexicografía diferencial se dedica a documentar el léxico de una variedad del español distinta a la europea peninsular. Este tipo de obra no está exenta de una visión política e ideológica. De acuerdo con Zimmerman, el riesgo más común que representan estos proyectos lexicográficos es el de promover o cimentar (aún de manera no intencional) la estigmatización de las variedades *diferentes*, en este caso primordialmente las latinoamericanas, respecto de aquella de la que se diferencian, en este caso la variedad hablada en España.

Frente al concepto de lexicografía diferencial, como se ha definido anteriormente, se contraponen el concepto de lexicografía integral, considerado por algunos como fundamental para describir la labor lexicográfica que faltaría en el español (Hernández, 2012). En un proyecto de lexicografía integral se documenta el español que se habla en una zona

geográfica de manera independiente a otras variedades de la lengua. En las lenguas pluricéntricas, en términos de Zimmerman (2018), donde no ha habido una visión monocéntrica donde una variedad regional es percibida como mejor que las demás, los diccionarios integrales han sido construidos de manera espontánea, sin siquiera requerir un término para describir y distinguir este tipo de obra. Algunas lenguas donde esto ha sucedido son el inglés, el alemán y el portugués. Sin embargo, el español con sus 19 variedades americanas (y cuando menos una africana en Guinea Ecuatorial), apenas cuenta con dos diccionarios integrales. Además, estos dos primeros diccionarios integrales, para Argentina (Plager, 2008) y México (Lara, 2010), no sólo son pocos, sino que tienen una primera edición de muy reciente publicación. Ahora bien, la distinción entre lexicografía diferencial e integral no es central a este estudio, pues aquí aplicaré el término según su otra definición. Sin embargo, he explicado esta concepción de lexicografía diferencial en detalle porque es la más comúnmente utilizada entre los creadores de diccionarios del español.

La segunda acepción de lexicografía diferencial no está motivada por una diferenciación diatópica, regional o nacional, sino de nivel de discurso (Zimmerman, 2018). Es decir, los diccionarios diferenciales también pueden dedicarse a documentar las variedades profesionales, técnicas o científicas, entre otras. Los productos resultantes de este enfoque son más comúnmente conocidos como diccionarios especiales o especializados. Ahora bien, para el proyecto descrito en este artículo, he preferido el término diferencial, y no especial o especializado, porque el primero también puede aplicarse a la documentación que el *ERES | Diccionario* hace de términos coloquiales, populares o groseros.

Además, otro elemento metodológico primordial que comparte el *ERES | Diccionario* con la lexicografía diferencial es que corrobora que los términos que incluye no se encuentren en ninguno de los diccionarios generales, o más propiamente hablando, integrales. De esta manera, el *ERES | Diccionario* incluye de manera exclusiva (o diferencial) términos que no se encuentran en el *Diccionario de la Lengua Española* de la Real Academia Española (2024b), comúnmente abreviado como *DRAE*, ni en el *Diccionario del Español de México* (*DEM*) (El Colegio de México, 2024), ambos en sus versiones en línea. Cabe señalar que aquí no consideramos al *Diccionario Integral del Español de la Argentina* (Plager, 2008), antes mencionado, porque éste no cuenta con una versión en línea de constante actualización, como los dos anteriores.

## 2.2. Recolección y procesamiento de datos

Tanto estudiosos, como practicantes de la lexicografía actual, coinciden en que un proyecto lexicográfico serio debería apoyarse en una recolección de datos y la utilización del corpus resultante (Atkins y Rundell, 2008; Zimmerman, 2018). Los colaboradores del *ERES | Diccionario*, después de seleccionar un área del conocimiento (como la política o la música) o un ámbito del quehacer humano (como el cine o la moda), en el cual se consideran medianamente expertos o con un conocimiento avanzado, se dedican a identificar fuentes en las redes sociales donde podrían recabarse datos relacionados con dicho tema. Cabe señalar que los colaboradores del proyecto del diccionario han sido primordialmente estudiantes avanzados o por graduarse de la Licenciatura de Idiomas de la Universidad de las Américas Puebla.

Según la red social de la que decidan estos colaboradores hacer la descarga automática de comentarios, las fuentes podrían ser videos con sus identificadores o sus direcciones de internet, usuarios destacados o con abundantes seguidores, y cuentas de alguna red social,

ya sea populares, ya sea institucionales, dedicadas a la temática en cuestión. Una vez identificadas dichas fuentes, a lo largo de los varios años que lleva el proyecto hemos utilizado diversos recursos para la descarga masiva de comentarios de redes sociales. Estos recursos han sido principalmente de dos tipos.

Por un lado, hemos utilizado servicios en línea que permiten la descarga automática de comentarios de ciertas redes sociales, como *YouTube Data Tools* (Rieder, 2015), *Vicinitas* (2024), *Ncapture for Chrome* (NVivo, 2019) y *Export Social Media Comments* (exportcomments.com, 2024). Estos servicios funcionan como una araña o robot, es decir, un programa de cómputo que descarga todos los comentarios asociados con una fuente de datos en línea, como las antes descritas. Entre los servicios en línea mencionados, algunos son especializados para alguna red social en particular. Por ejemplo, *YouTube Data Tools*, como su nombre lo indica, sirve para extraer comentarios de YouTube y *Vicinitas*, que estuvo en funcionamiento hasta 2023, extraía comentarios de Twitter, ahora X. Al contrario, *Ncapture for Chrome* es un servicio de descargas que funciona para Facebook, YouTube y X, y *Export Social Media Comments* permite la descarga de 9 redes sociales distintas, incluyendo todas las anteriormente mencionadas.

Este tipo de herramientas no están exentas de inconvenientes. En este sentido, *YouTube Data Tools* tuvo un funcionamiento intermitente durante el otoño de 2023, cuando el último grupo de colaboradores del diccionario estuvo trabajando en la próxima versión que está por lanzarse para la segunda mitad de 2024. Mientras que, con los cambios en la administración y reglas de operación de X, *Vicinitas* dejó de funcionar a mediados del 2023. Por su parte, *Ncapture for Chrome*, particularmente diferente a los anteriores al permitir la descarga de Facebook, produce datos que están acompañados de abundante código xml, el cual dificulta en demasía la producción de un corpus de texto con los mismos. Finalmente, hay que señalar que algunos de estos servicios requieren de un pago para la extracción de datos masivos. Este era el caso con *Vicinitas* que, hasta antes de su desaparición en 2023, permitía una extracción gratuita de máximo 2,000 comentarios producidos en las 2 semanas previas a la descarga. Actualmente, éste sigue siendo el caso de *Export Social Media Comments*, que es una herramienta muy atractiva por todas las redes sociales de las que puede descargar datos, pero que sólo extrae 100 comentarios a la vez sin ningún cargo, y que requiere de un pago para extraer comentarios por encima de esta cifra.

Ante las desventajas antes discutidas, el autor de este artículo también ha implementado algunas soluciones alternativas, escribiendo programas de cómputo de diseño simple que realizan esta función en YouTube y X. Para esto, he utilizado principalmente las librerías de Tweepy (2024) y de Google API Client (2024). Lamentablemente los cambios de política de X también han obstaculizado varias de las soluciones programáticas implementadas para nuestro proyecto en esa plataforma.

Por todo lo anterior, lo que parece ser el común denominador en la descarga automática de comentarios de redes sociales es la constante exploración de nuevas herramientas. Esto resulta ineludible ante los constantes cambios tanto de los servicios en línea, como de las políticas de las grandes tecnológicas que se encuentran detrás de estas valiosas fuentes de datos. Otro elemento que es común a esta tarea en la recolección de datos es el procesamiento mínimo que requieren los datos obtenidos para convertirlos en un corpus textual manipulable con las herramientas clásicas de la lingüística de corpus.

Todos los servicios de descarga automática de datos que se han mencionado con anterioridad, incluso las librerías para las soluciones programáticas implementadas en Python, recuperan muchos datos asociados con cada uno de los comentarios. En el caso de

los servicios en línea, los datos recuperados suelen descargarse en bases de datos en formatos compatibles con Excel u otras hojas de cálculo. Por tanto, es necesario hacer una selección de los datos relevantes y una importación a formatos de texto cuidando que no se degrade el contenido. La figura 1 a continuación muestra una base de datos recuperada por *YouTube Data Tools*.

**Figura 1.**

*Datos recuperados por un servicio de descarga automática de comentarios en YouTube*

A	B	C	D	E	F	G	H	I	J	K	L
id	replyCount	likeCount	publishedAt	authorName	text	authorChannelId	authorChannelUrl	isReply	isReplyTo	isReplyToName	
Ugw4e0xC	0	0	05/02/2018 19:31	@JohnRidesMx	Mami - Play Movil Project	UCxDpSV9EjEE9K-zcF	http://www.youtu	1	Ugw4e0xO5DI	@elzorromaicero20	
Ugw4tE8h	0	1	12/01/2018 05:59	@JohnRidesMx	No me la compré yo, sí	UCxDpSV9EjEE9K-zcF	http://www.youtu	1	Ugw4tE8h71g	@nazihirzamarripa	
Ugw6A8KV	0	0	12/01/2018 05:52	@JohnRidesMx	Sí- tienen una calidad ac	UCxDpSV9EjEE9K-zcF	http://www.youtu	1	Ugw6A8KW3g	@arturorangelrang	
UgwABt6C	0	0	08/05/2020 02:09	@alexgetgear	Misma duda	UCS84roTn15lk1J70fz	http://www.youtu	1	UgwABt6OpW	@victorherrer4790	
UgwARCW	0	0	12/01/2018 05:54	@JohnRidesMx	Gracias Carlos! un salud	UCxDpSV9EjEE9K-zcF	http://www.youtu	1	UgwARCWstH	@tiesto2889	
Ugw-bdQf	0	0	08/03/2018 19:03	@JohnRidesMx	Sí-, ya la vi, yo creo que p	UCxDpSV9EjEE9K-zcF	http://www.youtu	1	Ugw-bdQfAEC	@librepensadormx	
UgwC2so8	0	0	12/01/2018 05:53	@JohnRidesMx	Si tienes tiempo tío mism	UCxDpSV9EjEE9K-zcF	http://www.youtu	1	UgwC2so8t6F	@ismaelnar36	
UgwCtQVC	0	2	10/01/2018 04:38	@JohnRidesMx	Por estilo y precio esa es l	UCxDpSV9EjEE9K-zcF	http://www.youtu	1	UgwCtQVOgz	@leyaaz13	
UgwCtQVC	0	0	10/01/2018 05:31	@leyaaz13	Gracias por la opinión er	UC5uqfwh9Xfms-w4W	http://www.youtu	1	UgwCtQVOgz	@leyaaz13	
UgwCtQVC	0	0	03/02/2018 17:58	@juancv3700	leyaaz13 Tuve una discove	UCE3LPz1VXli4xw39a	http://www.youtu	1	UgwCtQVOgz	@leyaaz13	
UgwdmHL	0	1	10/01/2018 05:10	@JohnRidesMx	Por qué@ eliminaste tu i	UCxDpSV9EjEE9K-zcF	http://www.youtu	1	UgwdmHUp_	@mauwayne6937	
UgwdmHL	0	1	10/01/2018 05:28	@mauwayne6937	JohnRides es que ya soy n	UCHHZCjoitycc2LKgsi	http://www.youtu	1	UgwdmHUp_	@mauwayne6937	
UgwdmHL	0	0	11/01/2018 05:26	@ErickMG	Mau Wayne seran de 4k tu	UC4hDPxfWgbyOU58	http://www.youtu	1	UgwdmHUp_	@mauwayne6937	
UgwdmHL	0	0	11/01/2018 05:55	@mauwayne6937	erick daniel marito gonzal	UCHHZCjoitycc2LKgsi	http://www.youtu	1	UgwdmHUp_	@mauwayne6937	
UgwwebXUI	0	3	28/06/2019 23:52	@albertoarellano6397	Haz un versus con izuka 2	UC62Rv1QxUMoUoSR	http://www.youtu	1	UgwwebXUF6Q	@fernandoinfantefr	
UgwwebXUI	0	1	12/11/2019 05:36	@ottorhynotekuraeloy	Me gusta mucho más; el c	UCzBQ4PPH56n4KrNf	http://www.youtu	1	UgwwebXUF6Q	@fernandoinfantefr	
UgwENDQ	0	0	29/04/2020 03:20	@luiscamacho9885	Están buenas al menos n	UCPF3XnqFtuzmlSUH	http://www.youtu	1	UgwENDQTxh	@MrCrossdelena	
UgwGtjrm	0	0	30/07/2018 21:09	@JohnRidesMx	Sí-, ¡á;stima, es buena m	UCxDpSV9EjEE9K-zcF	http://www.youtu	1	UgwGtjrm5x	@lvanDelarge	
UgwGtc-N	0	0	05/02/2018 19:31	@JohnRidesMx	Estas son mejores, pero l	UCxDpSV9EjEE9K-zcF	http://www.youtu	1	UgwGtc-NeUn	@metalmoster001	
UghwA8r0	0	0	31/01/2018 18:13	@JohnRidesMx	Insurgentes sur, a un cost	UCxDpSV9EjEE9K-zcF	http://www.youtu	1	UghwA8r0h56	@TheRene88888	
UgwJqEtl	0	1	18/04/2019 01:13	@diegotgh	Saltate a una 300 o la vas	UCF_8rpnAipG6v3swf	http://www.youtu	1	UgwJqEtlbTl	@giuseppegarciari	

**Fuente:** Elaboración propia (2024) a partir de datos obtenidos en [https://ytdt.digitalmethods.net/mod\\_video\\_comments.php](https://ytdt.digitalmethods.net/mod_video_comments.php)

Como se puede observar en la Figura 1, la base de datos recuperada tiene once columnas, que van de la A a la K. La columna L no tiene contenido. De entre estas once columnas, tan sólo tres (las columnas D, E y F) son utilizadas para crear nuestro corpus. Esta misma figura muestra que el texto de los comentarios en sí, contenidos en la columna F, normalmente se degrada al desplegarlo en un editor de hojas de cálculo como Excel. Así, en la línea 3 de dicha columna, la palabra *compré* tiene la vocal acentuada degradada o desconfigurada para su correcta lectura. Lo mismo pasa con otros caracteres especiales o con diacríticos del español. En las líneas 4, 8, 10 y 18, de la misma columna F, se puede observar cómo se despliegan incorrectamente el resto de las vocales acentuadas, las cuales aparecen en las palabras *Sí, tú, opinión, y están*, respectivamente.

Lidiar con los conflictos en el manejo del texto recuperado requiere de una intervención cuidadosa por parte de los colaboradores, quienes deben ser entrenados para enfrentar estos problemas en la producción del corpus. En las soluciones más comunes está el uso de procesadores de texto que reconocen o permiten manipular los protocolos de codificación como UTF-8, entre otros, o el uso de pequeños programas de cómputo producidos por el autor de este artículo con la misma finalidad. La descripción detallada del preprocesamiento de los textos para la producción de nuestro corpus rebasa los objetivos de este trabajo. Sin embargo, es importante señalar que existen retos implícitos en la construcción de un corpus con comentarios producidos en diferentes redes sociales y recuperados con diversos métodos de descarga automática.

### 2.3. Extracción terminológica semiautomática

Una vez ensamblado el subcorpus de un área del conocimiento o quehacer humano escogida por un colaborador en particular, se hace una extracción terminológica semiautomática. Cabe

mencionar que los subcorpus construidos por los colaboradores del diccionario eventualmente se van sumando a un corpus general del proyecto. Para realizar la extracción terminológica, hemos estado utilizando la función de extracción de palabras clave que ofrece *AntConc* (2020).

Ahora bien, *AntConc* es un programa de cómputo popular en la lingüística de corpus, ya que permite manipular una colección de textos para realizar algunas acciones simples (búsquedas de palabras, obtención de sus frecuencias, producción de listas ordenadas de las palabras más frecuentes), así como algunas funciones más avanzadas (extracción de colocaciones, de palabras clave y de n-gramas o secuencias de palabras). En cuanto a la extracción de palabras clave, que nosotros utilizamos en este proyecto como si fuera un extractor terminológico, se puede considerar una función más bien semiautomática. Esto es porque *AntConc* cuenta con una interfase gráfica en la que el usuario, siguiendo una serie de pasos, puede producir una lista de palabras distintivas en su uso en un corpus especializado, llamado de análisis, respecto de un corpus general, llamado de referencia. Entre los pasos necesarios para dicha extracción, los principales incluyen justamente la carga de un corpus de referencia, de un corpus de análisis y de una lista de palabras vacías, así como la selección de un método estadístico de comparación y evaluación de términos.

En cuanto al fundamento para la extracción de dichas palabras, distintivas de un corpus frente al otro, *AntConc* utiliza uno de entre varios posibles modelos estadísticos. El modelo estadístico que nosotros seleccionamos es la *chi cuadrada* o *chi<sup>2</sup>*. Dicho método estadístico lo utilizamos, entre otras razones, porque representa un modelo estadístico de amplio uso y aceptación en los estudios de la lingüística aplicada, particularmente entre aquellos que recurren a las herramientas de la lingüística de corpus (Lindquist y Magnus, 2018). Cabe mencionar, sin embargo, que la popularidad de la *chi<sup>2</sup>* en la lingüística de corpus se deriva principalmente de uso como prueba de significancia estadística ante la distribución de las frecuencias de una variable dependiente con dos o más niveles, frente a una variable independiente con varios niveles también.

Para entender de manera simple este uso de la *chi<sup>2</sup>* podemos tomar como ejemplo dos elementos léxicos que se cree que están distribuidos de manera diferenciada entre dos niveles de discurso. En este caso la prueba de *chi<sup>2</sup>* sirve para determinar si cada uno de estos elementos se asocia de manera distintiva con un nivel de discurso en particular. Utilizando un caso concreto, si se tiene el binomio léxico *por favor* vs *xfa* (el segundo es un equivalente semántico del primero), se podrían extraer las frecuencias de ambos términos en un corpus de comunicaciones en el contexto laboral y en un corpus de comentarios en redes sociales. Con dichas frecuencias, asumiendo que el primer ítem léxico fuese más frecuente en el corpus más formal y el segundo en el corpus de comunicaciones digitales, la prueba de *chi<sup>2</sup>* podría decir si la distribución diferenciada es producto del azar o no y mostrar un nivel de certeza para dicha conclusión.

En cuanto a la aplicación de la prueba de la *chi<sup>2</sup>* en nuestro contexto, el constructo matemático es ligeramente distinto pues aquí se intenta evaluar la hipótesis de independencia entre un elemento lingüístico, en nuestro caso se trata de una palabra, y los corpus de referencia y análisis que se están comparando. De esta manera, si un elemento léxico se asocia más con uno de dichos corpus, se puede argüir que el tipo de lenguaje de dicho corpus hace más probable la ocurrencia del ítem léxico, y que este es un elemento característico del anterior (Manning, Raghavan, & Schütze, 2009). Partiendo de dicho razonamiento, la *chi<sup>2</sup>* puede producir una lista ordenada de los elementos léxicos que son característicos de un corpus especializado en contraste con su uso en un corpus de la lengua

general. Es de esta forma que la  $\chi^2$  puede ser utilizada como la base de un extractor de palabras clave, o en el caso del proyecto aquí descrito, de un extractor terminológico.

Ahora bien, ya hemos mencionado que los subcorpus especializados en los que se hace la extracción son el resultado del preprocesamiento de comentarios contenidos en múltiples descargas masivas que hacen los colaboradores en redes sociales. Sin embargo, falta aclarar cuál es nuestro corpus de referencia y qué tipo de contenidos lo hacen apropiado para la evaluación y extracción de terminología. Para este propósito utilizamos el *Corpus de Referencia del Español Actual (CREA)* de la Real Academia Española (2024a). Este corpus está compuesto con textos escritos y orales, producidos en 21 países, subdivididos en 5 medios o géneros, y en 10 temas generales con 124 subtemas. En este sentido, resulta evidente que el CREA fue diseñado con la intención de contar con una muestra representativa del español contemporáneo utilizado a nivel internacional. En cifras más específicas, nosotros utilizamos una lista global de palabras de este corpus que cuenta con 737,799 tipos y 152,558,294 tókenes producidos entre los años 1975 a 2004. Con esta lista de frecuencias y una del subcorpus construido por un colaborador, *AntConc* extrae una lista de palabras clave que dicho colaborador usa como punto de partida para identificar y proponer términos que puedan ser incluidos en nuestro diccionario.

#### 2.4. Una plataforma en línea para el ERES | Diccionario

Finalmente, el último aspecto de la metodología que falta por describir es la plataforma en línea que permite poner a disposición del público en general el diccionario cuyo diseño se ha descrito en los apartados anteriores. *Lexonomy* es una plataforma para escribir y publicar diccionarios en línea (Méchura, 2017). Es gratuita, no requiere instalaciones, ni conocimientos de programación por parte de los usuarios. Hay tres aspectos primordiales que el creador de un diccionario debe trabajar en la plataforma, pero que pueden llevarse a cabo a través de una muy amigable interfase en línea y que no requieren de conocimientos técnicos.

En primer lugar, se debe decidir cuál es la estructura de las entradas. Es decir, se debe determinar cuáles son los campos que potencialmente podrían tener las entradas del diccionario, tales como el lema, la información gramatical, etimológica, de pronunciación, la definición o los ejemplos. Para un proyecto en forma, la serie de reglas que determinan cómo se llenan estos campos es lo que constituye la planta (Real Academia Española, 1997) o guía de estilo del diccionario (Atkins y Rundell, 2008). Una vez que se han creado los campos de captura en *Lexonomy*, los colaboradores pueden comenzar a trabajar añadiendo y editando sus entradas. En segundo lugar, se necesitan decidir algunas características generales del diccionario, que van desde el orden de las entradas y las opciones de búsqueda, hasta el formato de presentación de los diferentes campos de cada entrada, es decir, tipos y opciones de letra, colores y puntuación, entre otros. Finalmente, algunas decisiones sobre la publicación no son menos importantes. Si el diccionario se publica en uno de los sitios donde se aloja *Lexonomy* (2024a; 2024b), las decisiones no son muchas. Las más importantes, como el nombre y la descripción del diccionario, así como la URL, que se convierte en el subdominio del diccionario mismo, pueden editarse en la plataforma de manera simple.

Ahora bien, como *Lexonomy* es un software de distribución libre, sin costo, siempre y cuando un tercero no cobre por su uso o redistribución, la plataforma puede instalarse en un servidor y dominio dedicados. El autor de este artículo decidió optar por esta última opción, pues, aunque para ello sí se requieren conocimientos técnicos, eso también permite la modificación del código html de las páginas que despliega la plataforma. Con esta opción, es posible responder a algunas de las exigencias de los motores de búsqueda más populares como Google Search Console (2024) o Bing Webmaster (2024). Los beneficios y los retos de

desplegar el ERES | Diccionario en un servidor y dominio dedicados se discutirán en las siguientes dos secciones, respectivamente.

### 3. Resultados

Para presentar los resultados del proceso de construcción del diccionario, descrito en la sección anterior, dos aspectos son los más importantes a discutir. Por un lado, esta sección presenta cuál es el contenido del diccionario que se ha logrado producir y cuáles son sus características en términos de cifras. En este sentido, se mostrará cuántas entradas, significados y ejemplos ha ido acumulando el diccionario durante los años que ha durado el proyecto. Por otro lado, al tiempo que se discuten las modificaciones más importantes implementadas en la plataforma *Lexonomy*, se presentarán el número de clics y de impresiones (o apariciones) en los motores de búsqueda más importantes.

En cuanto al contenido del diccionario, éste ha sido puesto a la disposición del público en tres etapas o entregas. La primera de ellas fue todavía en la plataforma original del proyecto de *Lexonomy* (2024a) y las siguientes dos en el servidor y dominio dedicados del diccionario (Rico-Sulayes, 2024). La Tabla 1, a continuación, muestra el crecimiento progresivo del diccionario en sus contenidos a través de las tres entregas que ha tenido.

**Tabla 1.**

*Contenidos del ERES | Diccionario en su plataforma en línea*

Entrega	Año	Entradas	Significados	Ejemplos
1ª	2021	357	492	621
2ª	2022	620	819	1.059
3ª	2023	821	1.082	1.368

**Fuente:** Elaboración propia (2024) a partir de datos obtenidos en <http://eresdiccionario.com/>

Como se puede observar en la Tabla 1, el contenido del diccionario en su primera entrega era de 357 entradas, 492 significados y 621 ejemplos. El mayor número de significados, frente al número de entradas se debe a que algunas entradas pueden tener más de un sentido. Además, según las reglas de la planta del diccionario, algunos sentidos exigen más de un ejemplo. Esto sucede, por mencionar un caso, cuando un lema puede ser usado como un sustantivo o un adjetivo, lo cual requiere de un ejemplo para cada uno de estos usos. Aunque los casos con esta exigencia no son tan comunes, cuando las entradas exhiben fenómenos morfológicos o gramaticales complejos, se suelen incluir varios ejemplos que muestran algunas de las excepciones en los patrones lingüísticos de dichas entradas. Esto sucede, por ejemplo, cuando una palabra tiene casos complejos de formación de accidentes de género (masculino y femenino) o número (singular y plural). Por mencionar un caso, el término *creepypasta* (género de historias de terror) puede producir las siguientes variaciones: *el creepypasta*, *la creepypasta*, *los creepypasta*, *las creepypasta*, *los creepypastas* y *las creepypastas*. Aunque este es un caso extremo, con 5 ejemplos, existen muchas otras entradas que incluyen, de manera similar, varios ejemplos para mostrar algunos modelos gramaticales atípicos en las entradas.

Otro dato importante que se puede observar en la Tabla 1 es el crecimiento progresivo de los contenidos del diccionario. Cabe señalar que las cifras de la primera entrega en el 2021 fueron el resultado del trabajo de dos generaciones previas de estudiantes que participaron

en el piloto del proyecto. Después, sin embargo, hemos continuado creciendo anualmente y de manera significativa. Así con dos entregas más, pasamos de 357 a 821 entradas, de 492 a 1.082 significados o sentidos, y de 621 a 1.368 ejemplos. También vale la pena comentar que con los contenidos de la primera entrega se publicó una primera edición en papel (Rico-Sulayes, 2022) y en el momento en que se escribe este artículo estamos trabajando en una segunda edición que incluye los contenidos más recientes.

En cuanto al impacto que se ha logrado en los usuarios del diccionario, la forma más importante de llegar a nuestro público es a través de los motores de búsqueda en la red. La necesidad de responder a las exigencias de dichos motores fue justo lo que motivó la migración del diccionario a un servidor y dominio dedicados.

El primer gran problema que nos presentaba estar bajo un subdominio de un servidor de *Lexonomy* es que Google no nos encontraba y no podíamos pedirle que nos indizara. Tampoco le podíamos pedir que indizara alguna página en concreto, menos aún todas. Para lograr lo anterior es necesario reclamar la propiedad de un dominio o un subdominio, y para esto se deben añadir etiquetas en el código de una página web. Entendiblemente, los administradores de *Lexonomy* no ofrecen una función tan avanzada que permita algo así por parte de sus usuarios. Una vez instalado el software de *Lexonomy* en un servidor propio pudimos ser los dueños del dominio y realizar las acciones anteriores junto con otras más que los motores requieren para mejorar la indización de sitios. Por ejemplo, Google requiere la inclusión de una etiqueta con la dirección canónica de cada página, y Bing pondera mucho la presencia de una etiqueta h1. El nivel técnico en la discusión de estas modificaciones supera los objetivos de este trabajo, pero únicamente en un servidor y dominio dedicados hemos podido responder a algunas de las exigencias de estos motores para posicionar mejor el diccionario.

En cuanto a las cifras que miden el impacto del diccionario, dos son particularmente importantes: el número de clics y de impresiones. El número de clics representa la cantidad de veces que alguien entra a una de las páginas del diccionario a través de un resultado mostrado por un motor de búsqueda, como Google o Bing. Aunque este dato suele ser el más importante para muchas páginas, en el caso de un diccionario, muchas veces el resumen que presenta una lista de resultados de un motor llega a contener la definición, junto con algunos otros datos, y con dicho resumen se resuelve la duda del usuario. Esto sucede en el ejemplo de la impresión de la palabra *churpio* en la Figura 2.

## Figura 2.

*Ejemplo de la impresión de la palabra churpio en el motor de búsqueda de Google*



**Fuente:** Elaboración propia (2024) con datos obtenidos en <https://www.google.com/>

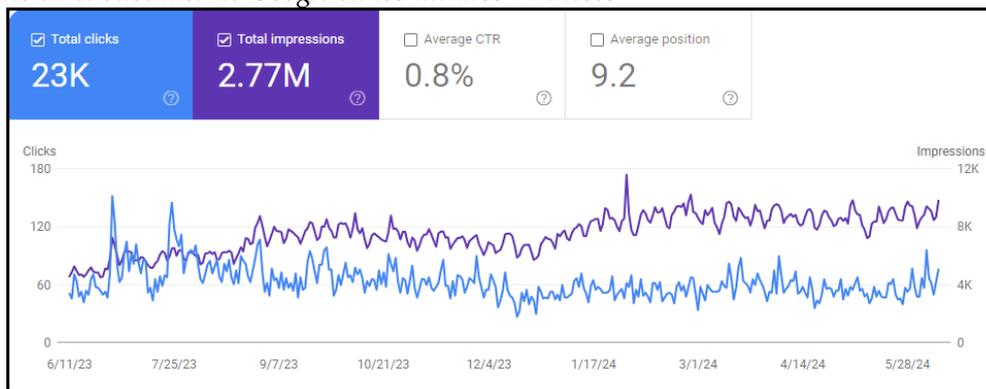
En cuanto a la Figura 2, aquí hay que señalar que el robot de Google, que constantemente indiza la red, puede cambiar sus impresiones de manera constante. Debido a que las impresiones pueden llegar a contener los significados del diccionario, es muy importante reportar el número de impresiones que produce el dominio del mismo. Si bien esto no

significa que ese número de dudas en los usuarios han sido resueltas, al menos representa qué tan frecuentemente el nombre del diccionario es visualizado por los internautas.

Ahora bien, la Figura 3 a continuación presenta el número de clics e impresiones producidas por el dominio eresdiccionario.com durante los 12 meses previos a la redacción de este artículo. Como esta figura muestra, nuestro dominio ha conseguido en un año poco más de 23 mil clics y casi 2,8 millones de impresiones.

**Figura 3.**

*Desempeño en el buscador de Google en los últimos 12 meses*



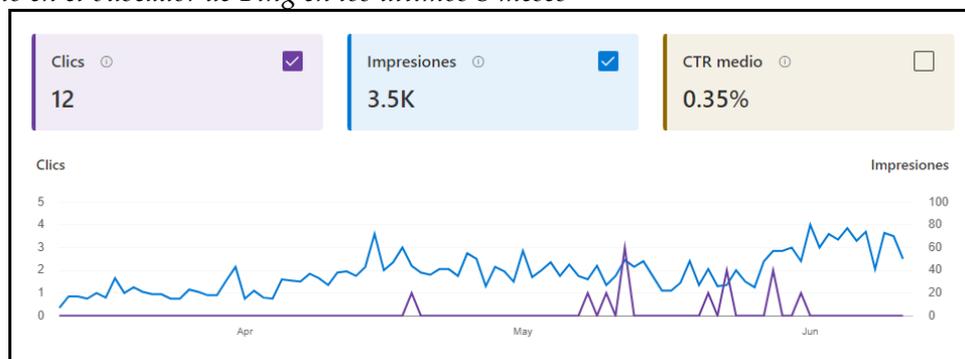
**Fuente:** Elaboración propia (2024) con datos y herramientas disponibles en <https://search.google.com/search-console/>.

De hecho, haciendo una proyección de los últimos tres meses, si el desempeño del diccionario se mantuviera, se deberían lograr producir más de 21 mil clics y más de 3,2 millones de impresiones, con un descenso de la primera cifra, pero con un aumento proporcionalmente mayor de la segunda, aunque ésta es tan sólo una proyección.

Los resultados obtenidos en el motor de búsqueda de Bing en los 3 meses previos al momento en que se escribe este artículo son mucho más discretos, como muestra la Figura 4. Sin embargo, esta misma figura muestra una clara tendencia a incrementar las impresiones y los clics en este lapso de tiempo, por lo que se espera que mejore el desempeño del diccionario en este buscador con el paso del tiempo.

**Figura 4.**

*Desempeño en el buscador de Bing en los últimos 3 meses*



**Fuente:** Elaboración propia (2024) con datos y herramientas disponibles en <https://www.bing.com/webmasters/>.

Las cifras mostradas en la Figura 4, comparadas con las de la Figura 3, tienen cuando menos dos posibles explicaciones. Por un lado, la interacción que se ha tenido con el motor de Bing es mucho más reciente, y por el otro, el número de búsquedas en internet que se hacen en Google comparado con Bing es dramáticamente superior. De hecho, según el servicio de monitoreo de buscadores en la red StatCounter (2024), Google posee actualmente más del 80% de las búsquedas en equipos de escritorio y el 95% de las búsquedas en equipos móviles, frente al 11% y sólo 0.6% de Bing, respectivamente. La diferencia es particularmente notoria en el caso de los equipos móviles, 95% frente a 0.6%, desde los cuales proceden el 83% de todos los clics que recibe nuestro dominio en el motor de Google. Este último dato parece explicar de manera simple por qué nuestros resultados son mucho mejores en la Figura 3. Ahora bien, en la siguiente sección se discuten dos aspectos más, también relacionados con el impacto del diccionario, el tráfico que tiene más allá de los motores de búsqueda y la distribución geográfica de sus usuarios.

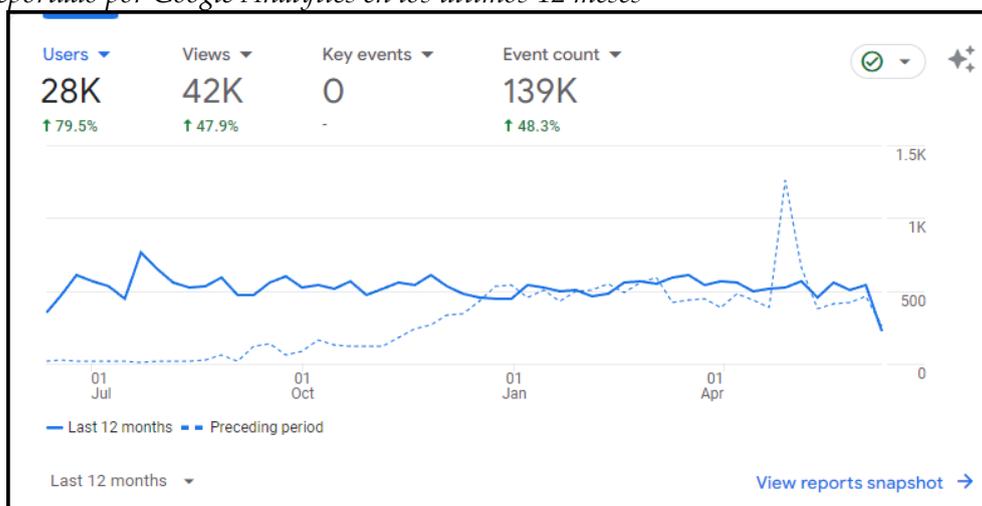
#### 4. Discusión

Más allá de los logros en la construcción del diccionario y de su posicionamiento y desempeño en los motores de búsqueda, de los que se habló en la sección anterior, nuestra plataforma ha tenido un número de visitas y eventos (o acciones por parte de los usuarios) que demuestra un éxito sólido entre los internautas. La distribución geográfica de este uso es, además, uno de los logros más importantes que ha conseguido la plataforma. Para presentar estos datos, particularmente relevantes para ponderar el impacto del diccionario, haré uso de las estadísticas del servicio de monitoreo de tráfico en la red de Google Analytics (2024).

En la Figura 5 a continuación se puede ver que el tráfico actual de nuestro diccionario rebasa por mucho los clics captados por medio de motores de búsqueda. En esta figura se puede ver que, en los 12 meses previos a la redacción de este artículo, durante el mismo lapso de tiempo mostrado en la Figura 3, el dominio del diccionario logró captar 28 mil usuarios y 42 mil vistas.

**Figura 5.**

*Tráfico reportado por Google Analytics en los últimos 12 meses*



**Fuente:** Elaboración propia (2024) con datos y herramientas disponibles en <https://analytics.google.com/>.

En primer lugar, la Figura 5 muestra más usuarios, 28 mil, frente a los 23 mil clics de la Figura 3. De hecho, esta última cifra es más bien comparable a las vistas de la Figura 5, que en este caso alcanzan la cifra de 42 mil. Esto quiere decir que, de las 42 mil páginas vistas en el sitio, tan sólo 23 mil fueron alcanzadas por un clic en los resultados de Google. Esto se podría deber a que quizá muchos usuarios realizan más de una consulta o si ya conocen nuestra página, llegan de manera directa a ella. En cuanto a la realización de varias consultas, la otra cifra que muestra la Figura 5 son los eventos. Éstos representan diversos tipos de acciones, como búsquedas dentro del mismo diccionario, clics a ligas internas que presentan variantes, sinónimos o antónimos de la misma palabra, desplazamientos con las barras del navegador, entre otras. Esta es una buena medida del interés que puede despertar el diccionario por el contenido que no es inmediatamente alcanzado con un primer clic. El total de eventos en el dominio a lo largo de los dos meses previos ya mencionados es de casi 140 mil. Esto significa que cada visitante del diccionario realiza un promedio de 5 acciones una vez que llega a él.

Finalmente, vale la pena decir algo respecto de la procedencia geográfica de los usuarios del sitio. Considerando sólo los últimos 12 meses, 19 países de habla hispana se encuentran entre los 22 con más usuarios. En orden, según el número de usuarios, esos países son: México, España, Colombia, Perú, Chile, Argentina, Ecuador, Venezuela, Guatemala, Bolivia, Puerto Rico, Costa Rica, República Dominicana, El Salvador, Honduras, Nicaragua, Panamá, Uruguay y Paraguay. Sólo 3 países se cuelan entre los 19 antes listados: Estados Unidos, que es el número 2 de toda la lista, Brasil en el puesto 19 e Italia justo antes de Paraguay. Sólo Cuba, en el puesto 29, es desplazado por varios países donde el español no es mayoritariamente la primera lengua. Cabe señalar que Filipinas aparece en el puesto 53 y la Guinea Ecuatorial en el 64, ambos países con tan sólo unos pocos usuarios. La Figura 6 a continuación muestra un mapa con un sombreado cuya intensidad representa la cantidad de usuarios.

**Figura 6.**

*Mapa representando el número de usuarios en los últimos 12 meses*



**Fuente:** Elaboración propia (2024) con datos y herramientas disponibles en <https://analytics.google.com/>

La Figura 6, además, muestra una lista ordenada de los 7 países con más usuarios, con su número de usuarios redondeado y la tendencia de crecimiento o disminución de los mismos. Como se puede observar en dicha figura, todos los 7 países listados presentan una tasa de crecimiento en los últimos 12 meses. Este mapa muestra también de manera visualmente sencilla el impacto del diccionario, el cual destaca en todo el territorio del continente americano y el occidente de Europa.

## 5. Conclusiones

Este artículo lanzó desde su inicio la pregunta de si las redes sociales podrían ser una fuente de términos léxicos novedosos o valiosos, a partir de los cuales se pudiera crear un diccionario que lograra un posicionamiento en los motores de búsqueda más importantes y cuyo contenido resultara útil para los internautas. Para poder lograr una respuesta afirmativa, se tuvo que diseñar un proyecto lexicográfico formal. En ese proyecto se diseñó un método de selección diferencial con la intención de no repetir el trabajo de los grandes diccionarios, primordialmente para que el contenido resultara valioso para usuarios y motores de búsqueda. Se partió de las recomendaciones de la lexicografía actual, produciendo un corpus actual y utilizando herramientas de corpus para intentar detectar la terminología importante de manera rápida. Y se trabajó en la mejora de una plataforma digital para desplegar los contenidos del diccionario. Los resultados parecen mostrar de manera fehaciente que se ha conseguido responder afirmativamente a la pregunta anterior.

Al mismo tiempo, es necesario reconocer que aún quedan muchos retos por resolver. Por ejemplo, el CTR o tasa de clics por impresiones es muy baja, 0.8% en Google (Figura 3) y 0.35% en Bing (Figura 4). Esto quiere decir que una muy pequeña proporción de quienes nos ven en la red entran al diccionario. Una posible explicación de esto se ha ofrecido en la sección de Resultados. Muchas veces el resumen de una página en la lista de resultados de un motor de búsqueda muestra toda la información necesaria para resolver la duda de un usuario, lo cual hace que se pierdan muchos clics. Para el autor de este artículo, director del proyecto del diccionario, el objetivo primordial de dicho proyecto es ayudar a los usuarios del internet en español resolviendo sus dudas léxicas. El servidor que personalmente pago con mi propio dinero tiene recursos muy limitados. Si una proporción considerable de las impresiones que produce el diccionario obtuviera una visita, probablemente el servidor se colapsaría. En este sentido, creo que el trabajo futuro más importante, además de continuar añadiendo contenidos valiosos al diccionario, es monitorear el comportamiento de los motores de búsqueda y de su indización de nuestro dominio para mantener un posicionamiento que haga que los usuarios de la red puedan resolver sus dudas llegando al diccionario directa o indirectamente.

## 6. Referencias

- Adame, A. (2019). *Redes sociales más usadas en el mundo hispano: tips para crecer tu presencia y alcance social*. <https://blog.hootsuite.com/es/redes-sociales-mas-usadas/>
- Anthony, L. (2020). *AntConc* (Versión 3.5.9). [Software]. Tokio, Japón: Waseda University. <https://www.laurenceanthony.net/software/antconc/>
- Atkins, B. T. y Rundell, M. (2008). *Oxford Guide to Practical Lexicography*. Oxford University.
- Bing Webmaster. (2024). *Microsoft Bing. Webmaster tools*. <https://www.bing.com/webmasters/about>

- Cameron, D. y Panović, I. (2014). *Working with Written Discourse*. SAGE.
- El Colegio de México (2024). *Diccionario del Español de México*. <https://dem.colmex.mx/>
- Elexis. (2024). *Opening up dictionaries, linguistic data and language tools for European communities*. <https://elex.is/>
- exportcomments.com (2024). *Export Social Media Comments*. <https://exportcomments.com/>
- Fernández de Molina Ortés, E. (2015). La interacción lingüística en las redes sociales. Aspectos ortográficos. En E. T. Montoro del Arco (ed.), *Estudios sobre ortografía del español* (pp. 39-51). Axac.
- Google Analytics (2024). *Google Marketing Platform. Analytics*. <https://marketingplatform.google.com/about/analytics/>
- Google API Client (2024). *Google API Python client library* (version 1.0.290.0). [Software]. <https://github.com/googleapis/google-api-python-client>
- Google Search Console (2024). *Google Search Console tools*. <https://search.google.com/search-console/about>
- Hernández, H. (2012). La lexicografía didáctica del español en Hispanoamérica: aspectos teóricos y descriptivos. En F. Rodríguez González (ed.), *Estudios de lingüística española: Homenaje a Manuel Seco* (pp. 269-296). Universidad de Alicante.
- Instituto Cervantes (2022). *El español: una lengua viva. Informe 2022*. [https://cvc.cervantes.es/lengua/anuario/anuario\\_22/informes\\_ic/p01.htm](https://cvc.cervantes.es/lengua/anuario/anuario_22/informes_ic/p01.htm)
- Lara, L. F. (dir.) (2010). *Diccionario del español de México*. El Colegio de México.
- Lexonomy (2024a). *Welcome to Lexonomy. Online, open-source platform for writing and publishing dictionaries*. <https://www.lexonomy.eu/>
- Lexonomy (2024b). *Lexonomy*. <https://lexonomy.elex.is/>
- Lindquist, H. y Magnus, L. (2018). *Corpus Linguistics and the Description of English* (2a ed.). Edinburgh University.
- Manning, C. D., Raghavan, P. y Schütze, H. (2009). *An Introduction to Information Retrieval*. Cambridge University.
- Měchura, M. (2017). Introducing Lexonomy: an open-source dictionary writing and publishing system. En I. Kosem, C. Tiberius, M. Jakubíček, J. Kallas, S. Krek, y V. Baisa (eds), *Electronic lexicography in the 21st century: Proceedings of eLex 2017 conference* (pp. 662-679). Lexical Computing.
- NVivo (2019). *NCapture for Chrome* (version 1.0.290.0). [Software]. <https://help-nv.qsrinternational.com/20/win/Content/ncapture/ncapture.htm>
- Plager, F. (coord.) (2008). *Diccionario integral del español de la Argentina*. Tinta Fresca - Voz Activa.

- Real Academia Española (1997). *Nueva Planta del Diccionario de la Real Academia Española*. Real Academia Española.
- Real Academia Española (2024a). *Corpus de referencia del español actual*. <http://corpus.rae.es/creanet.html>
- Real Academia Española (2024b). *Diccionario de la Lengua Española*. <https://dle.rae.es/>
- Rico-Sulayes, A. (Ed.). (2024). *ERES | Diccionario del Español Especializado en Redes Sociales*. <http://eresdiccionario.com/>
- Rico-Sulayes, A. (Dir.). (2022). *ERES | Diccionario del Español Especializado en Redes Sociales*. Aracne.
- Rieder, B. (2015). *YouTube Data Tools* (Version 1.42) [Software]. <https://ytdt.digitalmethods.net/>
- StatCounter (2024). *Search Engine Market Share Worldwide*. <https://gs.statcounter.com/search-engine-market-share/>
- Tweepy (2024). *Internet's Most Loved TwitterAPI Python Library* (Version 4.14.0) [Software]. <https://www.tweepy.org/>
- Vicinitas (2024). *Twitter analytics tool for tracking hashtags, keywords, accounts and websites*. [https://x.com/vicinitas\\_io/](https://x.com/vicinitas_io/)
- Zimmermann, K. (2018). Lexicografía diferencial y lexicografía integral. En M. Álvarez de la Granja, y E. González Seoane (eds.), *Léxico dialectal y lexicografía en la Iberoromania* (pp. 121-144). Iberoamericana Vervuert.

## CONTRIBUCIONES DE AUTORES/AS, FINANCIACIÓN Y AGRADECIMIENTOS

**Agradecimientos:** Hoy día el número de colaborades que ha habido en el proyecto del *ERES | Diccionario* hace prohibitiva una mención individual de cada uno de ellos. Se trata de los estudiantes de 5 generaciones (2019-2023) del curso avanzado LIO3042 Léxicos Especializados, de la licenciatura de Idiomas en la Universidad de las Américas Puebla. De entre ellos, más de una decena de estudiantes han colaborado en un proyecto terminal de servicio social, en el cual editamos las propuestas de los estudiantes del curso para darle la calidad de publicable a los contenidos que están disponibles en línea. Son todos estos jóvenes colaboradores los verdaderos autores del diccionario.

**AUTOR/ES:****Antonio Rico Sulayes:**

Universidad de las Américas Puebla.

Antonio Rico Sulayes es doctor en Lingüística Computacional por Georgetown University, en Washington, DC. Ha sido profesor de lingüística y computación en México, Colombia y los Estados Unidos. También ha trabajado como lingüista computacional para instituciones como la Organización Mundial de la Salud y para varios contratistas del Pentágono. Es autor de más de 40 artículos y de 3 libros de investigación. Ha dictado conferencias en una decena de países y es miembro fundador de la Asociación Mexicana de Procesamiento del Lenguaje Natural. Actualmente es profesor-investigador tiempo completo de la Universidad de las Américas Puebla, en México. Su investigación se enfoca en las tecnologías del lenguaje, la lingüística forense y los estudios léxicos.

[antonio.rico@udlap.mx](mailto:antonio.rico@udlap.mx)

**Índice H: 6**

**Orcid ID:** <https://orcid.org/0000-0003-0932-4733>

**Scopus ID:** <https://www.scopus.com/authid/detail.uri?authorId=51462128400>

**Google Scholar:** <https://scholar.google.com/citations?user=6HfpVBIAAAA>

**ResearchGate:** <https://www.researchgate.net/profile/Antonio-Rico-Sulayes/stats>