

## *La consulta en línea de frasemas y sus significados: un análisis de las búsquedas en un diccionario electrónico de neologismos*

---

### Online search of phrasemes and their meanings: analysis of queries in an electronic dictionary of neologisms

**Antonio Rico Sulayes**

Universidad de las Américas Puebla, Cholula, México

ORCID: <https://orcid.org/0000-0003-0932-4733>

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

#### RESUMEN

**Introducción:** Este artículo analiza las búsquedas en línea más populares que remiten a frasemas en un diccionario electrónico de neologismos. En particular, se analizan las estrategias de selección de un término de búsqueda que utilizan los hablantes para consultar el significado de expresiones pluriverbales.

**Métodos:** El análisis de los términos de búsqueda está guiado por la literatura, la cual ha explorado cuatro factores que podrían explicar su selección. Estos factores son la frecuencia del término, su categoría gramatical, su posición en el frasema y su extensión.

**Resultados:** Los resultados muestran que el factor con mayor poder explicativo en la selección de un término de búsqueda es la frecuencia. Esto coincide con estudios previos, confirmando la importancia de este factor con una fuente de datos sin precedentes en la literatura.

**Conclusiones:** Este análisis no sólo permite detectar cuál es el factor con el máximo poder explicativo, sino que, además, permite encontrar la aparición de nuevas estrategias de búsqueda de información fraseológica en los medios digitales. La frecuencia se confirma como el factor más relevante para predecir el término de búsqueda elegido por los usuarios. Los diccionarios electrónicos y los motores de búsqueda permiten nuevas estrategias de recuperación fraseológica no limitadas por la estructura univocal de los diccionarios impresos.

**PALABRAS CLAVE:** frasemas; neologismos; frecuencia; diccionario electrónico; término de búsqueda; Google

#### ABSTRACT

**Introduction:** This article examines the most frequent online search queries related to phrasemes in an electronic dictionary of neologisms. It specifically analyzes the strategies speakers use when selecting search terms to retrieve the meanings of multiword expressions.

**Methods:** The analysis of search terms is informed by the literature, which has identified four factors that may account for their selection: term frequency, grammatical category, position within the phraseme, and length.

**Results:** The findings indicate that frequency is the factor with the greatest explanatory power in the selection of search terms. This result is consistent with previous studies and further confirms the central role of frequency using a dataset unprecedented in the literature.

**Conclusions:** This analysis not only identifies the factor with the highest explanatory value, but also reveals the emergence of new strategies for retrieving phraseological information in digital environments. Frequency is confirmed as the most relevant predictor of users' choice of search terms. Electronic dictionaries and search engines, such as Google, enable new forms of phraseological retrieval that are not constrained by the single-word structure of printed dictionaries.

**KEYWORDS:** phrasemes; meaning; electronic dictionary; Search term; Google

## CONTRIBUCIÓN DE AUTORÍA

**Concepción y/o diseño de investigación:**

Antonio Rico Sulayes (100 %)

**Adquisición de datos:**

Antonio Rico Sulayes (100 %)

**Análisis e interpretación de datos:**

Antonio Rico Sulayes (100 %)

**Escritura y/o revisión del artículo:**

Antonio Rico Sulayes (100 %)

## INTRODUCCIÓN

Como primera aproximación, un frasema es una secuencia de palabras que producen un significado particular al aparecer juntas en cierto orden. Esta definición es controversial al menos por dos razones distintas. Por un lado, en torno al concepto de frasema se han utilizado diversos términos o etiquetas para denominar este fenómeno lingüístico. Dos términos importantes que han sido adoptados en diferentes comunidades académicas son el de *unidad fraseológica*, más común en la lingüística, y el de *expresión multiverbal*, más frecuente en el procesamiento del lenguaje natural (Corpas Pastor, 2013). La segunda fuente de controversia tiene que ver con qué tan específica es la definición misma de frasema. Diferentes autores hacen énfasis en aspectos diversos para definir este término, el presente artículo se apoyará en Castillo Carballo (1998), quien define un frasema como una secuencia de palabras con al menos dos características básicas, su fijación y su idiomática.

En cuanto a su fijación, los frasemas se consideran secuencias de palabras fijas porque han sido preestablecidas por los hablantes y son más o menos estables o inalterables. Así, por ejemplo, con el frasema verbal  *echar a perder* , más o menos equivalente con los verbos  *deteriorar* ,  *malograr*  o  *pervertir* , según su acepción (Real Academia Española, 2026b), uno puede decir que  *algo se echó a perder* , pero no se puede decir que  *a perder se echó algo* . En cuanto a la segunda característica antes mencionada, estas expresiones son idiomáticas porque su significado no resulta de la suma de los significados de las palabras que las

componen. Si algo *se echó a perder*, no *echó* ni *perdió* nada, y si uno no conoce el significado de esta expresión, no es posible deducirlo a partir de dichas palabras.

Dadas estas dos características esenciales, un frasema puede experimentar un proceso de lexicalización o especialización semántica, es decir, una consolidación del grupo de palabras en un significado unitario (Corpas Pastor, 1997). Por esta razón es que los frasemas son material primordial en la construcción de los diccionarios de una lengua, así como en la enseñanza y aprendizaje de la misma (Mel'čuk, 2001).

Ahora bien, el reconocimiento de los frasemas como material inequívoco de los diccionarios implica otros problemas teóricos y metodológicos que han sido objeto de estudio en la lingüística moderna. Por un lado, es necesario establecer cuál es la mejor manera de incluir dichos elementos multiverbales en un recurso lingüístico como el diccionario, cuyos materiales necesariamente se ordenan por una lista univocal (Lew y Tomczak, 2022). En la literatura especializada (Pamies, 2022), es posible encontrar análisis lexicográficos que muestran diccionarios con una inclusión de frasemas asistemática y potencialmente problemática para los usuarios. Por otro lado, para responder cuál es la mejor manera de introducir los frasemas en los diccionarios, sería ideal saber cómo es que los usuarios de una lengua intentan acercarse a un frasema cuyo significado pretenden consultar (Lew y Tomczak, 2022). Con este objetivo en mente, este trabajo ha analizado cómo los usuarios de diccionarios electrónicos llegan a los frasemas a través de los motores de internet. Para ello, se ha utilizado como fuente de datos un recurso que, después de una búsqueda exhaustiva, no parece haber sido explotado anteriormente en la literatura con este propósito: las listas de búsquedas más frecuentes de *Google Search Console* (2025).

Para llevar a cabo el tipo de análisis antes planteado, se utilizaron las 1,000 búsquedas más frecuentes en Google que remiten a un diccionario electrónico de neologismos, el *ERES | Diccionario del Español Especializado en Redes Sociales* (Rico-Sulayes, 2026a). Esta lista contiene las búsquedas más populares de dicho diccionario durante los 12 meses anteriores a la recolección de datos, que tuvo lugar a finales de abril de 2025. Cabe señalar que tan sólo estas mil búsquedas produjeron 2,080,477 impresiones, o apariciones de páginas de dicho diccionario, entre los resultados de Google vistos por usuarios en línea.

El análisis de esta lista de búsquedas ha estado guiado por las siguientes preguntas de investigación: ¿Qué proporción de las búsquedas del diccionario electrónico estudiado son búsquedas univocales o multiverbales? ¿Qué proporción de las búsquedas contienen elementos léxicos de entradas con frasemas en el diccionario? ¿De entre las anteriores, qué elementos léxicos de los frasemas están incluidos en dichas búsquedas? Y finalmente, ¿cómo estos elementos léxicos presentes en las búsquedas en línea de frasemas se corresponden con lo encontrado en estudios previos acerca del comportamiento de usuarios que buscan el significado de frasemas? Las respuestas a estas preguntas intentan mostrar las particularidades de la recuperación de significados de *unidades fraseológicas* en

medios electrónicos y cómo estas particularidades se corresponden o no con los hallazgos de estudios anteriores con otros enfoques metodológicos.

## METODOLOGÍA

La presente investigación se apoya en la literatura que previamente ha estudiado el comportamiento de los hablantes al momento de elegir un término para buscar un frasema en el diccionario. Como ya se dijo, los factores que se han considerado como posibles detonantes de esta elección han sido la frecuencia del término de búsqueda, su categoría gramatical, su posición en el frasema y su extensión (Lew y Tomczak, 2022). Particularmente, lo que se ha encontrado en la literatura es que, respecto de la frecuencia, el término de búsqueda preferido suele ser el de menor frecuencia, en cuanto a la categoría gramatical, parece haber un orden de prioridad (sustantivos, adjetivos, verbos, adverbios y el resto de las categorías), sobre la posición del término, las palabras iniciales parecen ser preferidas, y de la extensión, se ha hipotetizado que la palabra más larga debería tener mayor saliencia. Estos cuatro factores y sus hallazgos inspiran los pasos del análisis que se presentan a continuación.

Para analizar la lista de las 1,000 búsquedas más populares del *ERES | Diccionario* en *Google Search Console* se siguieron los siguientes pasos. En primer lugar, se decidió hacer una primera clasificación de las búsquedas como univerbales y pluriverbales. Como se aprecia en las filas con la posición 1 y 4 de la Tabla 2, *churpio* y *churpio significado* representan una búsqueda univerval y una pluriverval, respectivamente. Aquí debe señalarse que una búsqueda pluriverval o con varias palabras (como *churpio significado*) no tiene nada que ver, al menos no necesariamente, con un frasema. De igual manera, una búsqueda univerval (como *churpio*) podría remitir tanto a un frasema, como a una entrada univerval.

En segundo lugar, se determinó si la búsqueda incluía un término que pudiese llevar a una entrada con un frasema en el diccionario. Si este era el caso, la búsqueda se clasificó como fraseológica y, como tercer paso, se analizó según los cuatro factores antes mencionados. Es decir, se analizó si el término era la palabra más rara o menos común, si su categoría gramatical respondía al orden de prioridad de la literatura, si la palabra era la primera y si era la más larga. Finalmente, si una búsqueda pluriverval contenía varias palabras del frasema, no se realizó el análisis anterior, pues los usuarios no seleccionaron un término en concreto. Sin embargo, este tipo de búsqueda se contabilizó como una búsqueda fraseológica directa.

Cabe señalar que una búsqueda clasificada como fraseológica, es decir, que incluye un término de búsqueda que remite a un frasema del diccionario, podría ser vista como ambigua en el sentido de que podría buscar o bien a la palabra entrada, o bien a alguno de sus frasemas. Este sería el caso sin duda en un diccionario general, que tiene como objetivo capturar todas las palabras de la lengua. Sin embargo, en un diccionario de neologismos, esta situación no es tan problemática, ya que, si el término de entrada de un frasema es

una palabra común, esta no suele tener una definición en este tipo de diccionario. La siguiente sección presenta los resultados de la clasificación de las búsquedas según los pasos descritos anteriormente.

## **RECUPERACIÓN DE FRASEMAS POR LOS HABLANTES DE LA LENGUA**

Lew y Tomczak (2022) presentan probablemente el estudio más completo y actual acerca del comportamiento de los hablantes en el momento en que reconocen e intentan recuperar el significado de un frasema. Con base en una revisión bibliográfica exhaustiva, estos autores listan cuatro factores como los que han sido consistentemente utilizados para predecir la palabra específica que los hablantes seleccionan cuando intentan consultar el significado de un frasema en el diccionario. Como ya se ha dicho antes, los diccionarios se encuentran ordenados por una lista de palabras individuales, así que al momento de buscar un frasema, los usuarios del diccionario se ven forzados a elegir un término de búsqueda. Los factores que han sido explorados en la literatura para predecir la elección de dicho término de búsqueda son la frecuencia del término, su categoría gramatical, su posición en el frasema y su extensión.

Acerca del primer factor antes listado, la frecuencia de la palabra seleccionada al buscar el significado de un frasema es explorada independientemente por Bogaards (1990; 1991; 1992) y Lew (2012). Los tres estudios de Bogaards (1990; 1991; 1992) exploran la importancia de la frecuencia del ítem léxico de búsqueda y concluyen que la palabra de menor frecuencia en el frasema es la más comúnmente seleccionada para realizar una búsqueda de la expresión multiverbal. Lew (2012) también explora la frecuencia del término de búsqueda y la compara con otros factores, como la categoría gramatical y la posición de dicho término. Este estudio confirma los hallazgos de las tres investigaciones de Bogaards antes mencionadas y concluye que es el ítem léxico menos frecuente es el que más utilizan los usuarios de los diccionarios al momento de buscar el significado de un frasema. En la misma línea, Tono (1987) argumenta que los usuarios tienden a preferir los ítems léxicos menos familiares.

Entre los estudios que exploran la importancia de la categoría gramatical en la selección del ítem léxico de búsqueda, está Béjoint (1981). En dicha investigación, el autor concluye que los usuarios prefieren utilizar al sustantivo núcleo, cuando buscan un sustantivo compuesto, y optan por el verbo en los frasemas verbales, aunque si estos últimos contienen un sustantivo, este llega a competir en las preferencias de los usuarios con el verbo central. Otro trabajo de investigación que se enfoca en la categoría gramatical del término de búsqueda es Tono (1987). Este estudio concluye que las palabras de contenido son preferidas a las palabras función. En Bogaards (1990), el autor también explora la importancia de las categorías gramaticales y concluye que los usuarios prefieren en primer lugar a los sustantivos, en segundo a los adjetivos y finalmente a los verbos. Aunque Lew (2012) encuentra que la frecuencia juega un rol más importante que la categoría

gramatical, este estudio concluye que las categorías más relevantes para seleccionar el término de búsqueda obedecen el siguiente orden de prioridad: en primer lugar, los sustantivos, en segundo, los adjetivos, en tercero, los verbos, en cuarto, los adverbios y finalmente, las demás categorías.

En cuanto a la posición del término de búsqueda en el frasema, Lew (2012) concluye que no tiene prioridad entre las preferencias de los usuarios. Bogaards (1990) encuentra que los términos iniciales son importantes en la búsqueda de frasesmas en francés, pero estos suelen ser los núcleos sintácticos, al menos de los frasesmas estudiados en esa lengua en este estudio, así que esa preferencia podría tener un origen más bien gramatical según el autor. Esto último aplica también para el estudio de Béjoint (1981), que observa la preferencia de los sustantivos núcleo que aparecen al inicio de las expresiones analizadas por este otro autor.

Finalmente, el único estudio que explora la importancia de la longitud del término de búsqueda es Lew y Tomczak (2022). De hecho, este estudio evalúa los cuatro factores listados al inicio de esta sección: la frecuencia, la categoría gramatical, la posición del término de búsqueda y su extensión. La conclusión de este estudio es que la frecuencia y la posición del término de búsqueda son los dos factores más importantes en las estrategias de los usuarios que intentan acceder a los frasesmas, representando la gran mayoría de las selecciones de los participantes en el estudio. De esta manera, el estudio concluye que la extensión del término de búsqueda tiene una importancia marginal en la consulta de frasesmas.

## LA INCLUSIÓN DE FRASEMAS EN LOS DICCIONARIOS ELECTRÓNICOS

Como se dijo al inicio de este artículo, las características de los frasesmas los hacen material de inclusión en el diccionario de una lengua (Mel'čuk, 2001). En este sentido, la idiomática, o falta de transparencia en su significado, hacen que los hablantes que no conocen un frasema en particular no puedan deducir su significado a partir de las palabras que lo componen. Además, un *frasema lexicalizado* se ha consolidado semánticamente de manera tal que representa un significado unitario y, por tanto, se puede definir en el diccionario al igual que se define cualquier palabra individual de la lengua (Corpas Pastor, 1997). Por lo anterior, los diccionarios impresos han incluido frasesmas de manera constante a través del tiempo (Ermakova et. al., 2022) y los problemas de dicha inclusión han sido abordados por los estudiosos de la lengua (Dalpanagioti, 2021; Pamies, 2022). Sin embargo, este estudio aborda la recuperación de frasesmas en un diccionario electrónico, por lo que esta sección se enfoca más bien en algunas de las particularidades de la inclusión de frasesmas en medios digitales desde la perspectiva de quienes trabajan con este tipo de soporte.

Los temas de discusión más comunes entre los creadores de bases de datos de frasesmas suelen ser las ventajas de los medios electrónicos para almacenar estas colecciones de datos y la manera de optimizar dicho almacenamiento y su recuperación. Ermakova et al.

(2022) presentan los proyectos del Centro de Lexicografía Digital de la Lengua Alemana y del Diccionario Digital de la Lengua Alemana. De acuerdo con estos autores, ambos proyectos comparten como uno de sus objetivos primordiales la inclusión de expresiones multiverbales idiomáticas o frasemas en sus plataformas de búsqueda. La discusión de Ermakova et al. (2022) se enfoca principalmente en el reto que implica la alta variabilidad de muchas expresiones multiverbales del alemán. Ante este problema, los autores proponen la creación de múltiples entradas que capturan las diversas formas de un frasema de manera separada y la construcción de una entrada central con características específicas que, según argumentan, permiten el vínculo más eficiente entre todas las demás formas.

En su trabajo de investigación y desarrollo, Osenova y Simov (2024) presentan un proyecto de inclusión de frasemas de la lengua búlgara en diferentes recursos electrónicos: un banco de árboles para el entrenamiento de etiquetadores gramaticales, un diccionario con información gramatical y un diccionario del tipo *wordnet*, es decir, una base de datos léxica con información semántica de sus entradas, como sinónimos y antónimos. Estos autores abordan el ya mencionado problema de la lematización de los frasemas. Es decir, el conflicto al momento de decidir cuál es el término de recuperación de un frasema para incluirlo en un diccionario ordenado por una lista de elementos univerbales. Osenova y Simov presentan una propuesta de entrada que busca ser común a los diccionarios antes mencionados, así como a otros recursos léxicos digitales que los autores han trabajado en el pasado. En la delimitación de dicha propuesta, los autores definen los campos que una entrada cualquiera debería tener, incluyendo su información paradigmática. En el caso de un frasema, los autores proponen la inclusión de elementos adicionales en dicha información paradigmática para con ellos responder a la complejidad del contenido del frasema a nivel de su flexibilidad morfológica, sintáctica y semántica.

En su construcción de una base de datos multilingüe, Mogorrón (2012) presenta un recurso digital que almacena *unidades fraseológicas verbales* primordialmente. Este recurso, enfocado sobre todo a la traducción y la enseñanza de lenguas, incluye al español, que es la lengua en que se enfoca el presente estudio. Antecediendo los comentarios de Ermakova et al. (2022), Mogorrón (2012) enfatiza las ventajas del soporte digital, el cual permite la captura de múltiples entradas para las diversas formas de un mismo frasema, así como la recuperación de las relaciones entre las variantes de un frasema central.

## **LOS FRASEMAS EN EL ERES | DICcionario**

El ERES | *Diccionario del Español Especializado en Redes Sociales*, cuya lista de búsquedas más populares en Google es el objeto de análisis de este artículo, también contempla la inclusión de frasemas en sus entradas. La última edición impresa del diccionario incluye la planta de construcción y las reglas de inclusión de expresiones multiverbales (Rico Sulayes, 2025b). En dichas reglas, se explica que este diccionario sigue a la última edición

del diccionario de la Real Academia Española (2026b) y que selecciona como término de entrada del frasema al primer sustantivo, verbo, adjetivo, adverbio, pronombre, preposición o conjunción, en ese orden de prioridad. Este orden está parcialmente apoyado por la literatura (Bogaards, 1990; Lew, 2012), aunque según las investigaciones citadas, el adjetivo debería ir antes del verbo y después del adverbio no importa el orden. Además, como ya se dijo también, la frecuencia ha sido reconocida como el criterio más importante para la selección del término de búsqueda o entrada (Bogaards, 1990; 1991; 1992; Lew, 2012; Lew y Tomczak, 2022). Este es un punto de mejora que el diccionario está comenzando a implementar, así como la política de capturar un frasema bajo varios términos y con diversas formas, que es una de las ventajas de los soportes digitales señalada por los creadores de diccionarios electrónicos de o con frasesmas (Dalpanagioti, 2021; Ermakova et al., 2022; Mogorrón, 2012). Sin embargo, este diccionario ha apostado por la recuperación de sus frasesmas a través de dos recursos ajenos al equipo lexicográfico que trabaja en el desarrollo de sus contenidos, la plataforma de publicación del diccionario y su indexación por los motores de búsqueda.

Por un lado, este diccionario está disponible en la red por medio de un servidor dedicado que contiene una instalación del software de acceso abierto *Lexonomy* (2025). Esta es una plataforma para la edición y publicación de diccionarios, cuyo desarrollo inicial estuvo fondeado por la Unión Europea (Měchura, 2017). Haciendo uso de las funciones de dicha plataforma, el *ERES | Diccionario* utiliza la opción de consulta de *Lexonomy* para hacer búsquedas simultáneamente sobre cualquier elemento del lema de la entrada y del frasema. Por otro lado, el autor de esta investigación ha modificado el código de la plataforma de *Lexonomy* para interactuar de la manera más conveniente posible con los requerimientos del motor de indexación de Google. La apuesta por mejorar la indexación de Google tuvo como principal objetivo que este motor de búsqueda se encargara de hacerle llegar a los usuarios la información que mejor resolviera sus dudas, incluyendo cualquier contenido fraseológico. El análisis de la lista de búsquedas del *Google Search Console* (2025) aquí presentado es un primer intento para analizar si este trabajo ha sido fructífero.

## DATOS

En esta investigación se analizan las búsquedas más populares en *Google Search Console* para el *ERES | Diccionario del Español Especializado en Redes Sociales* (Rico Sulayes, 2026a). Este es un diccionario de neologismos y términos especializados que aparecen en la comunicación digital. Sus criterios de construcción han sido descritos con detalle en Rico Sulayes (2024). Aquí vale la pena mencionar al menos dos aspectos de los contenidos del diccionario: los criterios de selección y el alcance de los mismos.

Por un lado, este diccionario tiene un enfoque diferencial, es decir, incluye términos que no aparecen en otros diccionarios, particularmente en los grandes diccionarios del español de confección independiente y disponibles en línea. De manera específica, el diccionario

incluye términos que no aparecen en el *Diccionario de la Lengua Española* (Real Academia Española, 2026b), ni en el *Diccionario del Español de México* (El Colegio de México, 2025). Por otro lado, el diccionario ha estado disponible en la red desde el año 2021 y ha incrementado sus contenidos de manera constante, con una nueva entrega cada año desde su aparición. A continuación, la Tabla 1 muestra el desarrollo de los contenidos del diccionario. Esta tabla proporciona cifras para algunos de los elementos más importantes del diccionario, incluyendo sus frasemas.

**Tabla 1.** Contenidos del *ERES | Diccionario* en su plataforma en línea

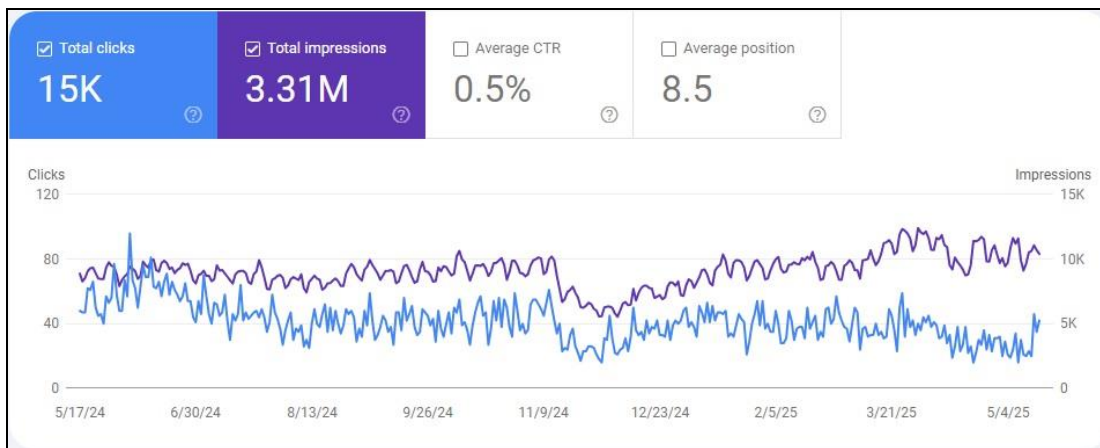
ENTREGA	AÑO	ENTRADAS	SIGNIFICADOS	EJEMPLOS	FRASEMAS
1 <sup>a</sup>	2021	357	492	621	38
2 <sup>a</sup>	2022	620	819	1,059	55
3 <sup>a</sup>	2023	821	1,082	1,368	106
4 <sup>a</sup>	2024	959	1,245	1,541	117

**Fuente:** Elaboración propia

Como muestra la tabla anterior, el diccionario ha pasado de 357 entradas en 2021, a 959 en su versión más actual, puesta a disposición del público en la segunda mitad de 2024. Frente al número de entradas, el número de significados es mayor, 492 en 2021 y 1,245 en 2024. Esto obedece a que muchas entradas ofrecen más de un sentido. Además, el diccionario se ha fijado el propósito de ser ejemplificado, o más bien, documentado. Es decir, el diccionario ofrece ejemplos reales de uso que documentan cómo los hablantes utilizan el término de una entrada. Esta característica se debe en parte al hecho de que el diccionario se construye por medio de una extracción semiautomática de términos en un corpus de comentarios en redes sociales, el cual va produciéndose año con año. Por tanto, los ejemplos fueron una forma más de aprovechar la construcción paralela del corpus del que se extraen los datos. De esta manera, el diccionario comenzó con 621 ejemplos en 2021, y cuenta con 1,541 en su versión actual. Finalmente, un dato importante para la presente investigación es el número de frasemas. Como se puede ver en la Tabla 1, el diccionario ha tenido desde sus inicios la intención de codificar y documentar neologismos y términos especializados que sean expresiones multiverbales. Esta intención se ha mantenido a través de los años que lleva activo el proyecto y el diccionario ha pasado de tener 38 frasemas en su primera versión, a 117 en 2024.

En cuanto a la presencia del *ERES | Diccionario* en la red, particularmente en el motor de Google, la Figura 1 muestra dos datos importantes, el número de clics que recibe el dominio dedicado del diccionario y el número de apariciones entre los resultados de Google que los usuarios de este motor llegan a visualizar. Este último número representa lo que los administradores de motores de búsquedas llaman impresiones. La gráfica de la Figura 1 muestra estos datos para los 12 meses anteriores a la redacción de este artículo.

**Figura 1.** Desempeño del diccionario en el buscador de Google en los últimos 12 meses



**Fuente:** Elaboración propia

Según muestra la figura anterior, el diccionario ha recibido 15 mil clics y poco más de 3.3 millones de impresiones en el último año. Este número concreta la predicción del autor de este trabajo, quien había proyectado que eventualmente el diccionario alcanzaría anualmente «más de 3,2 millones de impresiones» (Rico Sulayes, 2024). Otro elemento a destacar de entre los números presentados en la Figura 1 es la importancia de las impresiones, que pueden ser vistas incluso como más relevantes que los clics. Para los administradores de muchas páginas de internet obtener un clic a través de un motor de búsqueda es el resultado más deseado. Sin embargo, para el *ERES | Diccionario* lo más importante es resolver las dudas de los hablantes o aprendices del español. En este sentido, las impresiones pueden ser tan o más útiles que los clics. Esto se debe a que, dada la particularidad de los contenidos del diccionario, en muchas ocasiones los motores de búsqueda copian la mayor parte de la información relevante de una entrada y la hacen parte de la impresión de la misma en los resultados. Esto sucede, por ejemplo, con la palabra *churpio*, cuya impresión aparece en la Figura 2.

**Figura 2.** Impresión de la palabra *churpio* en el motor de búsqueda de Google



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

La figura anterior muestra que la necesidad de conocer la definición de *churpio* es perfectamente cubierta por la impresión en Google de su entrada en el diccionario. Incluso, cabe señalar que las impresiones en dicho motor de búsqueda se han refinado con

el tiempo y actualmente la impresión presenta la definición resaltada en negritas, algo que no sucedía hace un año (Rico Sulayes, 2024). Por lo anterior, las impresiones representan en muchos casos la solución de la duda que el *ERES | Diccionario* busca allanar con sus materiales.

Ahora bien, teniendo en cuenta las características del diccionario y de su indexación, en este trabajo se analizaron las 1,000 búsquedas que obtuvieron el mayor número de clics e impresiones durante el periodo comprendido entre el 30 de abril de 2024 y el 30 de abril de 2025. El conjunto de estas búsquedas produjo un total de 11,278 clics y 2,080,477 impresiones, como se muestra en la fila inferior de la Tabla 2 a continuación. Esta tabla muestra, además, las 10 búsquedas más frecuentes del diccionario con algunos datos adicionales reportados por *Google Search Console*.

**Tabla 2.** Las 10 búsquedas más populares del diccionario en el motor de Google

POSICIÓN	BÚSQUEDA	CLICS	IMPRESIONES	CTR
1	Churpio	1,350	39,955	0.0338
2	Yoyopo	686	14,331	0.0479
3	Castroso	621	60,282	0.0103
4	churpio significado	416	23,611	0.0176
5	Funable	324	97,179	0.0033
6	yoyopo que es	271	3,954	0.0685
7	Papulince	168	58,014	0.0029
8	que es el yoyopo	151	1,969	0.0767
9	Nmms	149	17,369	0.0086
10	que es yoyopo	147	2,361	0.0623
11 ... 1000	...	6,995	1,761,452	...
Total	...	11,278	2,080,477	...

**Fuente:** Elaboración propia

Entre los datos que muestra la tabla anterior, la posición de popularidad de una búsqueda aparece en la primera columna de la izquierda. La búsqueda en la posición 1, *churpio*, es la que más clics generó en el periodo de tiempo para el que se extrajeron los datos. Además de las búsquedas mismas, incluidas en la segunda columna, se presentan los clics y las impresiones en la tercera y la cuarta columna de izquierda a derecha, respectivamente. Finalmente, en la quinta y última columna aparece el CTR (del inglés *clickthrough rate*). Este representa la ratio de clics entre el número de impresiones. Así, por ejemplo, la división de los 1,350 clics recibidos por la página de la palabra *churpio* entre las 39,955 impresiones que produjo esta búsqueda, dan el CTR de 0.0338. El total de las 1,000 búsquedas más populares ordenadas de esta manera fueron clasificadas y analizadas según se explica en la siguiente sección.

## ANÁLISIS Y RESULTADOS

Como se indicó anteriormente, la investigación aquí presentada estuvo guiada por cuatro preguntas. Ya que las últimas dos preguntas comparten un mismo fin, este trabajo ha tenido como propósito encontrar tres datos principales: (1) la proporción de búsquedas univerbales frente a las búsquedas multiverbales en el diccionario electrónico estudiado, (2) la proporción de las búsquedas que contienen realmente elementos léxicos de entradas con frasemas en el diccionario (muchas búsquedas multipalabra no se refieren a un frasema y algunas búsquedas univerbales podrían buscar una entrada fraseológica), y (3) de entre las búsquedas que se refieren a frasemas, qué elementos léxicos se han utilizado en dichas búsquedas y cómo se corresponden estos elementos con los criterios de selección que, según los estudios previos, los usuarios emplean al buscar el significado de frasemas.

Para presentar el primer dato de la lista anterior, un primer paso en el análisis fue clasificar rápidamente las búsquedas en univerbales, como *churpio*, o pluriverbales, como *churpio significado*. Con una búsqueda a través del sublenguaje de patrones de las expresiones regulares se pudieron contabilizar rápidamente estos dos tipos de búsquedas. La fórmula utilizada para contar las coincidencias univerbales fue  $^{\wedge}\backslash w+\$$  (refinada en  $^{\wedge}\backslash S+\$$ ) y para contar las coincidencias pluriverbales la fórmula fue  $^{\wedge}\backslash w+(\backslash w+ ?)^+$ .

La primera fórmula, en sus dos variantes, comienza con un metacarácter (o carácter con significado especial en este sublenguaje),  $^{\wedge}$ , que le pide a un procesador de texto profesional, colocarse al inicio de un campo donde hay una secuencia de caracteres cualquiera. Luego el metacarácter  $\backslash w$  o  $\backslash S$ , según las dos versiones, le pide al procesador identificar cualquier carácter propio de un elemento lingüístico (todas las letras mayúsculas y las minúsculas, los números y por razones históricas que rebasan la presente discusión, el guion bajo). El siguiente metacarácter,  $+$ , solicita que haya cuando menos uno de los caracteres anteriores y cualquier número de otros caracteres de este rango de opciones, y finalmente, el metacarácter  $\$$  indica que después de eso, allí debe acabar el campo en cuestión. Como no hay lugar para que exista un espacio intermedio, la fórmula sólo captura expresiones lingüísticas univerbales o palabras individuales ocupando un campo completo. La fórmula contabilizó un total de 125 búsquedas univerbales.

La segunda fórmula utilizada,  $^{\wedge}\backslash w+(\backslash w+ ?)^+$ , comienza con los primeros tres metacaracteres de la fórmula anterior en su versión inicial,  $^{\wedge}\backslash w+$ , y con ello detecta la aparición de una primera palabra, pero después solicita la aparición forzosa de un espacio en blanco y cuando menos una combinación de aquello que se encuentra entre los paréntesis seguidos del símbolo  $+$ ,  $\backslash w+$ , lo cual representa la fórmula de otra palabra como ya se ha explicado. Esta fórmula detectó un total de 825 búsquedas pluriverbales.

Este conteo rápido por medio del sublenguaje de las expresiones regulares fue además corroborado por un etiquetador humano, para llegar a la clasificación de las búsquedas univerbales y pluriverbales, con 125 y 825, respectivamente. Con estas cifras, se puede apreciar que la gran mayoría de las búsquedas, de entre las 1,000 más populares, son

pluriverbales, con un 82.5% del total de las mismas. La Tabla 3 muestra las cinco búsquedas univerbales y pluriverbales más populares, con su posición original en la lista ordenada de *Google Search Console*.

**Tabla 3.** Las 5 búsquedas univerbales y pluriverbales más populares

UNIVERBAL		PLURIVERBAL	
POSICIÓN	BÚSQUEDA	POSICIÓN	BÚSQUEDA
1	Churpio	3	churpio significado
2	Yoyopo	5	ptm significado
4	Castroso	8	castroso significado
6	Churpia	9	yoyopo que es
7	Funable	11	que es el yoyopo

**Fuente:** Elaboración propia

En la tabla anterior se puede apreciar que, al menos entre las búsquedas más populares, las univerbales tienen posiciones más altas; sin embargo, la posibilidad de crear búsquedas pluriverbales más diversas hace que estas últimas sean muy superiores en su proporción frente a las búsquedas univerbales. También a partir de la tabla anterior, es posible ver que los usuarios de internet utilizan diferentes estrategias para llegar a la definición de una palabra en el diccionario, sobre todo en las búsquedas pluriverbales. Para ello, incluyen otras palabras en sus búsquedas como *significado* y *qué es*. Otras expresiones de este tipo que son frecuentemente utilizadas son *qué significa*, *significado de*, *qué es ser [adjetivo]* y *define*.

Aunque la gran mayoría de las búsquedas más populares en Google son pluriverbales, eso no significa que todas ellas remitan a un frasema. Por esto, en un segundo paso, se analizaron todas las búsquedas para determinar si contenían un término de búsqueda que les permitiera a los usuarios llegar a una entrada con un frasema. Presentar las cinco o diez búsquedas que remiten a un frasema y que más clics tienen sería demasiado repetitivo, pues la mayoría de las búsquedas más populares que remiten a un frasema son ligeras variaciones que remiten a una sola entrada en el diccionario, *yoyopo*. Esta entrada cuenta con el frasema, *beso en el yoyopo*, «expresión utilizada en forma de despedida» (Rico Sulayes, 2026a). Más interesante resulta mencionar que se encontraron 59 búsquedas que remiten a entradas fraseológicas del diccionario y en las que los usuarios seleccionaron un término de búsqueda. Estas 59 búsquedas remiten potencialmente a 13 frasesmas distintos. La Tabla 4 a continuación presenta las 11 búsquedas con más clics que podrían remitir a esos 13 frasesmas del diccionario. Esta tabla muestra la posición de la búsqueda según su número de clics, la búsqueda, la entrada en el diccionario en que se encuentra el frasema y el frasema mismo al que se remite o se podría remitir.

**Tabla 4.** Las búsquedas más populares que remiten a un frasema a través de un término seleccionado

POSICIÓN	BÚSQUEDA	ENTRADA EN EL ERES	FRASEMA EN EL ERES
2	Yoyopo	Yoyopo	beso en el yoyopo
40	Peyoyo	Peyoyo	beso en el peyoyo
85	Pinshi	Pinshi	ah su pinshi madre
117	Verguiza	verguiza	en verguiza
247	que es prro	Prro	ah prro
282	Gracias	Gracias	dar las gracias
334	ñam que significa	ñam ñam	ñam ñam
544	Pex	Pex	mal pex, qué pex, ni pex
617	Mamu	Mamar	se mamu
623	que significa ps	Ps	ps bueno
942	que es laik	auto like	auto like

Fuente: Elaboración propia

Como muestra la Tabla 4, las 11 búsquedas listadas remiten a 13 frasemas porque la búsqueda con la posición 544 podría remitir a tres frasemas distintos. Algunas observaciones que vale la pena hacer es que el diccionario ha ido implementando la entrada de frasemas por la palabra menos común, como sucede en las entradas para *yoyopo*, *peyoyo* y *pinshi*. También habría que notar que el diccionario le ha dado la entrada a *gracias*, que se origina en un error ortográfico, pero que según el equipo lexicográfico parece estar ganando nuevos matices de significado, sobre todo estilístico. También cabe señalar que algunas entradas en el diccionario ya contienen en su lema el frasema completo, como *ñam ñam* y *auto like*. Esto se debe a que en estos casos el primer unigrama, o secuencia de palabras, no tiene estatus de palabra independiente en la lengua. En el caso de *auto*, este debería ser un prefijo derivativo con enclisis. Las entradas listadas están capturadas así, no porque el equipo desarrollador ignore las normas de la lengua, sino porque su objetivo es describir el uso de la misma por quienes utilizan las redes sociales, y no prescribir cómo deberían hacerlo.

Ahora bien, en el tercer paso, se hizo un análisis de las 59 búsquedas que remiten a un frasema y que seleccionan un término de búsqueda según los cuatro factores que podrían explicar el criterio de selección de dicho término. La Tabla 5 muestra los conteos totales de dicho análisis.

**Tabla 5.** Categorización de términos de búsqueda según factores de selección

TÉRMINOS SELECCIONADOS	FACTORES DE SELECCIÓN			
	MENOS FRECUENTE	1A CATEGORÍA DEL ORDEN CANÓNICO	POSICIÓN INICIAL	MÁS LARGO
59	59	22	12	41

Fuente: Elaboración propia

Los resultados de la clasificación en la Tabla 5 muestran que el factor de selección que más poder explicativo tiene es el de la frecuencia, ya que, en el cien por cien de los casos analizados, el término de búsqueda seleccionado es efectivamente el menos frecuente. Las frecuencias de las palabras en las búsquedas fueron verificadas en el *Corpus del Español del Siglo XXI (CORPES)* (Real Academia Española, 2026a), aunque la rareza de varios de los términos de los neologismos hacía obvia la menor frecuencia que deberían exhibir. En cuanto al segundo factor, tan sólo 22 casos seleccionaban la primera categoría gramatical en el orden canónico que se ha argumentado en los estudios previos. Esto representa apenas el 37.2 % de los casos. La posición inicial es el factor menos útil para explicar la selección del término de búsqueda, con apenas 12 casos que representan el 20.3 %. Finalmente, la extensión del término parece ser el segundo factor más importante, con 41 casos que representan el 69.4 %. Sin embargo, existe la coincidencia de que 23 búsquedas son variaciones de los neologismos *yoyoyo* y *peyoyo*, términos relativamente largos. Si le restamos las 21 repeticiones de estos términos, este factor tendría apenas 20 casos que representan el 33.8 %. Por lo anterior, se concluye que el factor más consistente que explica la selección del término de búsqueda es la frecuencia, con el término de menor frecuencia como el preferido por los usuarios.

## CONCLUSIONES

Los resultados de la clasificación presentados en la sección anterior muestran que un mismo término de búsqueda puede ser explicado por diferentes factores al mismo tiempo. Esto se debe a que el análisis de los diferentes factores es explicativo y no causativo. Es decir, el hecho de que un factor explique una selección no significa que esa sea la causa que produjo esa selección. Ya en la revisión bibliográfica se había hecho notar que algunas explicaciones que los autores habían dado en sus análisis, como que la selección estaba basada en una categoría gramatical, se podían también explicar por un factor de posición inicial, debido a que en una lengua en concreto dicha categoría aparece al inicio de una secuencia en particular. Este era el caso de algunos argumentos en los trabajos de Béjoint (1981) y Bogaards (1990) con la lengua francesa. A pesar de este inconveniente, el análisis recién presentado permite mostrar cuáles son los factores que comúnmente justifican la selección de un término de búsqueda y que pueden ofrecer un argumento más simple y elegante para explicar los procesos utilizados por los usuarios de diccionarios. Por todo lo anterior, la frecuencia, que permite la explicación del 100% de los casos frente a los otros factores que aplican a una proporción mucho menor (entre el 37 % y el 20 %), aparece indiscutiblemente como el factor con mayor poder explicativo.

También vale la pena mencionar que además de las 59 búsquedas que seleccionaron un término para potencialmente llegar a un frasma, en las 1,000 búsquedas analizadas se encontraron 38 búsquedas que buscaron los frasemas de manera directa. Esta es justamente una de las ventajas de los diccionarios electrónicos, que sus plataformas de

publicación y los motores de búsqueda que los indizan permiten búsquedas no univocales. Estas búsquedas pluriverbales y fraseológicas al mismo tiempo tienen efectos bastante útiles, pues estas 38 búsquedas producen 6,427 impresiones, que representan casi una quinta parte de lo que producen las 59 búsquedas con un término seleccionado, con 33,943 impresiones. Esto muestra una tendencia en los usuarios de medios digitales de adoptar otras estrategias de búsqueda que no están limitadas por las características de los medios impresos.

Finalmente, es necesario reconocer también una limitante de este estudio. De las búsquedas con un término seleccionado, aunque muchas sí incluyen indicios de que están buscando un significado, algunas podrían no estar buscando el significado del término o de su fraseo. Sin embargo, los grandes números con que permite trabajar el *Google Search Console* justifican la posible falta de precisión respecto a algunas de las acciones en los usuarios que dan origen a la lista que aquí se ha analizado.

## REFERENCIAS

- BÉJOINT, H. (1981). The foreign student's use of monolingual English dictionaries: A study of language needs and reference skills. *Applied Linguistics* 2 (3), 207–222. <https://doi.org/10.1093/applin/ii.3.207>
- BOGAARDS, P. (1990). Oú cherche-t-on dans le dictionnaire? *International Journal of Lexicography* 3 (2), 79–102. <https://doi.org/10.1093/ijl/3.2.79>
- BOGAARDS, P. (1991). Word frequency in the search strategies of French dictionary users. *Lexicographica* 7, 202–212. <http://dx.doi.org/10.1515/9783110244113.202>
- BOGAARDS, P. (1992). French dictionary users and word frequency (pp. 51–61). En TOMMOLA, K. H., VARANTOLA, K., SALAMI-TONONEN, T., SCHOPP, J. (EDS.), *EURALEX '92 Proceedings*. Tampere, Finland: University of Tampere.
- CASTILLO CARBALLO, M. A. (1998). El concepto de unidad fraseológica. *Revista de Lexicografía* 4, 67–79.
- CORPAS PASTOR, G. (1997). *Manual de fraseología española*. Madrid: Gredos.
- CORPAS PASTOR, G. (2013). Detección, descripción y contraste de las unidades fraseológicas mediante tecnologías lingüísticas (pp. 335–374). En OLZA MORENO, I., Y MANERO RICHARD, E. (COORDS.), *Fraseopragmática*. Alemania: Frank & Timme.
- DALPANAGIOTI, T. (2021). A Cognitive perspective on the representation of MWEs in Electronic Learner's dictionaries (pp. 177–199). En KOSEM, I., CUKR, M., JAKUBÍČEK, M., KALLAS, J., KREK, S., TIBERIUS, C. (EDS.), *Electronic lexicography in the 21st century. Proceedings of the eLex 2021 conference*. Brno, Czechia: Lexical Computing.
- EL COLEGIO DE MÉXICO. (2025). *Diccionario del Español de México*. <https://dem.colmex.mx/>
- ERMAKOVA, M., GEYKEN, A., LEMNITZER, L., Y ROLL, B. (2022). Integration of multi-word expressions into the Digital Dictionary of German Language (DWDS): Towards a lexicographic representation of phraseological variation (pp. 851–860). En KLOSA-KÜCKELHAUS, A., ENGELBERG, S., MÖHRS, C., STORJOHANN, P. (EDS.), *Dictionaries and*

- Society. *Proceedings of the XX EURALEX International Congress*. Mannheim, Alemania: IDS-Verlag. <https://doi.org/10.14618/phpy-6r66>
- GOOGLE SEARCH CONSOLE. (2025). *Google Search Console tools*. <https://search.google.com/search-console/about>
- LEW, R. (2012). The role of syntactic class, frequency, and word order in looking up English multi-word expressions. *Lexikos* 22, 243–260.
- LEXONOMY. (2025). *Welcome to Lexonomy. Online, open-source platform for writing and publishing dictionaries*. <https://www.lexonomy.eu/>
- MĚCHURA, M. (2017). Introducing Lexonomy: an open-source dictionary writing and publishing system (pp. 662–679). En KOSEM, I., TIBERIUS, C., JAKUBÍČEK, M., KALLAS, J., KREK S., BAISA V. (EDS.), *Electronic lexicography in the 21st century: Proceedings of eLex 2017 conference*. Brno, Czechia: Lexical Computing.
- MEL'CUK, I. (2001). Fraseología y diccionario en la lingüística moderna (pp. 267–310). En UZCANGA VIVAR, I., LLAMAS POMBO, E., PÉREZ VELASCO, J. M. (EDS.), *Presencia y renovación de la lingüística francesa*. Salamanca: Universidad de Salamanca.
- MOGORRÓN HUERTA, P. (2012). Explotación informática de una base de datos multilingüe de unidades fraseológicas (pp. 63–80). En GONZÁLEZ REY, M. I. (ED.), *Unidades Fraseológicas y TIC*. Las Rozas: España: CERSA.
- OSENOVA, P., Y SIMOV, K. (2024). Representation of multiword expressions in the Bulgarian integrated lexicon for language technology (pp. 117–146). En GIOULLI, V., BARBU MITITELU, V. (EDS.), *Multiword expressions in lexical resources: Linguistic, lexicographic, and computational perspectives*. Berlín, Alemania: Language Science Press.
- REAL ACADEMIA ESPAÑOLA. (2026a). *Corpus del Español del Siglo XXI (CORPES)*. <https://www.rae.es/corpes/>
- REAL ACADEMIA ESPAÑOLA. (2026b). *Diccionario de la lengua española (23ª ed.)*. <https://dle.rae.es/>
- RICO SULAYES, A. (Director) (2026a). *ERES | Diccionario del Español Especializado en Redes Sociales*. <http://eresdiccionario.com/>
- RICO SULAYES, A. (Director) (2025b). *ERES | Diccionario del Español Especializado en Redes Sociales (2ª ed.)*. Roma, Italia: Aracne.
- RICO SULAYES, A. (2024). Impacto y audiencia de la plataforma en línea de un diccionario del lenguaje emergente en las redes sociales. *European Public & Social Innovation Review* 9, 1–17. <https://doi.org/10.31637/epsir-2024-302>
- TONO, Y. (1987). *Which word do you look up first? A study of dictionary reference skills*. Tesis de Maestría. Tokio, Japón: Universidad de Tokio Gakugei.

## DATOS DEL AUTOR

**Antonio Rico Sulayes, (1975, México)**. Doctor en Lingüística Computacional por Georgetown University en Washington, DC. Ha sido profesor de Lingüística y Computación en México,

Colombia y los Estados Unidos. Es autor de alrededor de 50 artículos y de 4 libros. Actualmente es profesor tiempo completo de la Universidad de las Américas Puebla.

**CÓMO CITAR ESTE ARTÍCULO:** RICO, A. (2026). La consulta en línea de frases y sus significados: un análisis de las búsquedas en un diccionario electrónico de neologismos. *Islas*, 68(213): e1634. <https://islas.uclv.edu.cu/index.php/islas/article/view/1634>



Este texto se distribuye bajo una licencia Creative Commons  
Reconocimiento-NoComercial 4.0 Licencia Internacional.

ISSN: 0042-1547 (papel) ISSN: 1997-6720 (digital)

<http://islas.uclv.edu.cu>