

Las secuencias formulaicas en la enseñanza del español como LE/L2. La frecuencia léxica como criterio de nivelación

NARCISO CONTRERAS IZQUIERDO

Universidad de Jaén

FERMÍN MARTOS ELICHE

Universidad de Granada

Recibido: 19 octubre 2019 / Aceptado: 3 abril 2020

ISSN paper edition: 1697-7467, ISSN digital edition: 2695-8244

RESUMEN: Durante los últimos años, la competencia léxica se ha ido erigiendo como un elemento central del desarrollo de la competencia comunicativa en la enseñanza de lenguas. De entre los distintos tipos de unidades que integran el componente léxico, destacan las denominadas *secuencias formulaicas* (SF), unidades léxicas pluriverbales que se almacenan y recuperan de la memoria como un todo en el momento de su uso, y cuyo dominio mejora la fluidez en la interacción comunicativa de los aprendientes. No obstante, para el docente puede resultar complejo decidir qué SF trabajar en cada nivel de enseñanza/aprendizaje, habiéndose recurrido tradicionalmente a la frecuencia como criterio para su nivelación. Es por esto por lo que el principal objetivo de este trabajo consiste en el análisis, a través de distintos corpus del español, de la frecuencia de las SF presentes en el *Plan Curricular del Instituto Cervantes* (PCIC) para comprobar la posible correlación entre la asignación de los niveles establecidos en el *Marco Común Europeo de Referencia para las Lenguas* (MCER) y la frecuencia de uso entre los hablantes nativos. Sobre esta base, proponemos orientaciones para la nivelación de dichas expresiones con el fin de guiar al profesor de español en su selección, contribuyendo así a la mejora de su enseñanza en el aula de ELE/EL2. Para ello, se ha analizado en diferentes corpus la frecuencia de 3260 SF extraídas del PCIC para confirmar su asignación a cada nivel, desde A1 a C2.

Palabras clave: competencia léxica, secuencias formulaicas, uso de corpus lingüísticos, enseñanza de español LE/L2, niveles de enseñanza/aprendizaje, frecuencia léxica.

The formulaic sequences (FS) in the teaching of Spanish as FL / 2L. The lexical frequency as a levelling criterion.

ABSTRACT: In recent years, lexical competence has emerged as a central element in the development of communicative competence in language teaching. Among the different types of units that make up the lexical component, the so-called *formula sequences* (FS) stand out, pluriverbal lexical units that are stored and recovered from memory as a whole at the time of use, and whose domain improves fluidity in the communicative interaction of the learners. However, it can be complex for the teacher to decide which FS to work at each level of teaching / learning, having traditionally used frequency as a criterion for levelling. This is why the main objective of this work is the analysis, through different Spanish corpus, of the frequency of the FS present in the *Curricular Plan of the Instituto Cervantes* (PCIC) to verify the possible correlation between the allocation of levels established in the *Common*

European Framework of Reference for Languages (CEFR) and the frequency of use among native speakers. On this basis, we propose guidelines for the levelling of these expressions in order to guide the Spanish teacher in their selection, thus contributing to the improvement of their teaching in the SFL / 2L classroom. For this, the frequency of 3260 FS extracted from the PCIC has been analysed in different corpus to confirm its allocation at each level, from A1 to C2.

Keywords: lexical competence, formulaic sequences, use of linguistic corpus, teaching Spanish FL/2L teaching, lexical frequency.

1. INTRODUCCIÓN

En el marco de la consideración de la competencia léxica como elemento central del desarrollo de la competencia comunicativa en la enseñanza y aprendizaje de lenguas extranjeras, en los últimos tiempos es evidente la importancia concedida a las denominadas *secuencias formulaicas* (SF), unidades léxicas pluriverbales, combinaciones de palabras que se almacenan y recuperan de la memoria como un todo en el momento de su uso (Wray, 2002, p. 9), y que los nativos emplean en su discurso de forma natural como recurso expresivo en situaciones comunicativas determinadas.

Teniendo en cuenta que estas SF pueden llegar a constituir más de la mitad del discurso nativo (Erman y Warren, 2000), su dominio contribuye decisivamente en el desarrollo de la competencia comunicativa de los aprendientes, mejorando su fluidez en la interacción, confirmando así los planteamientos del *enfoque léxico*, desarrollado por Lewis (1993, 1997).

En el caso del español como lengua extranjera o segunda (ELE/EL2), estas SF aparecen recogidas en el *Plan Curricular del Instituto Cervantes* (PCIC), concretamente en el apartado de “Nociones específicas”, aunque los exponentes allí recogidos forman parte de series abiertas, sin ánimo de exhaustividad, y se presentan para dar una idea del tipo de unidades léxicas que el hablante debe conocer en cada uno de los niveles de referencia (A1-C2). Respecto de la selección de estas expresiones, los criterios son la frecuencia y rentabilidad comunicativa, pero siempre partiendo de la apreciación intuitiva basada en la experiencia docente. Sin embargo, como ya evidenciaba Penadés (1999), y como sigue sucediendo en la actualidad, no contamos con estudios que nos permitan determinar con mayor precisión qué SF enseñar en los diferentes niveles de aprendizaje.

En un trabajo anterior (Martos y Contreras, 2018), comprobamos que dicha apreciación intuitiva es bastante precisa para la adscripción de las SF a los niveles en el PCIC, por lo que en este trabajo nos planteamos como objetivo el análisis, tanto en *Google* como en diversos corpus de español disponibles en la actualidad, de la frecuencia de SF presentes en el PCIC, desde el nivel A1 al C2, para comprobar si dicha asignación de nivel se ajusta a la frecuencia de uso entre los hablantes nativos.

El empleo de la frecuencia como criterio para medir de un modo relativamente objetivo la rentabilidad del léxico y determinar qué vocabulario debemos enseñar a los estudiantes de lenguas extranjeras cuenta con una larga tradición, e igualmente este mismo parámetro ha sido un factor esencial en la selección de las unidades destinadas para el aprendizaje en el ámbito de los estudios fraseológicos (Alvar, 2005; García Salido, 2017; García y Alonso, 2018; Nation, 2001; Sinclair y Renouf, 1985). No obstante, somos conscientes de que por sí sola la frecuencia resulta insuficiente a la hora de determinar qué léxico enseñar, por lo que

sería necesario tener en cuenta otros aspectos como la dispersión en los corpus de consulta, la rentabilidad y productividad, la opacidad o transparencia, los factores culturales, los intereses de los alumnos, las propias producciones de los estudiantes, o incluso las intuiciones de los hablantes nativos (Benigno, Kraiff, Grossmann y Velez, 2016; García Salido, 2017; García y Alonso, 2018; Gómez Molina, 2004; McGee, 2008; Nation, 2001).

2. MARCO TEÓRICO

2.1. Competencia léxica y secuencias formulaicas

El nivel léxico y el desarrollo de la competencia léxica han centrado la atención de la lingüística y de la adquisición de lenguas en los últimos años, ya que el léxico se considera un componente central en el discurso, de ahí la necesidad de implementar nuevos avances metodológicos para su enseñanza y aprendizaje.

Dicho nivel léxico está formado por diversos tipos de unidades: palabras simples y compuestas, fórmulas rutinarias o expresiones institucionalizadas, colocaciones, compuestos sintagmáticos, locuciones idiomáticas o modismos (Gómez Molina, 2004), y esta diversidad ya aparece recogida tanto en el *Marco común europeo de referencia* (MCER) como en el PCIC, puesto que ambos documentos manejan, desde una perspectiva nocional, un concepto amplio de unidad léxica que permite dar cuenta de la dimensión combinatoria del léxico, por lo que se incluyen, además de lexías simples, toda una serie de unidades léxicas pluriverbales. No obstante, autores como Velázquez (2018, p. 22) critican el tratamiento superficial y asistemático de estas unidades pluriverbales en el PCIC, ya que se basa en el principio cuantitativo-acumulativo en detrimento del criterio de la frecuencia de uso y de la funcionalidad comunicativa.

Una de las denominaciones más extendidas para estas expresiones es la de “unidades fraseológicas”, que según la definición ya clásica de Corpas (1996), son:

unidades léxicas formadas por más de dos palabras gráficas en su límite inferior, cuyo límite superior se sitúa en el nivel de la oración compuesta. Dichas unidades se caracterizan por su alta frecuencia de uso, y de coaparición de sus elementos integrantes; por su institucionalización, entendida en términos de fijación y especialización semántica; por su idiomatidad y variación potenciales; así como por el grado en el cual se dan todos estos aspectos en los distintos tipos (p. 20).

Igualmente, en el ámbito de la enseñanza de lenguas se ha extendido el uso de “secuencias formulaicas” (SF), que para Wray (2002, p. 9) son una combinación de palabras, continua o discontinua, que es o parece ser prefabricada, esto es, que se almacena y recupera de la memoria como un todo en el momento de su uso, en lugar de ser un enunciado construido palabra por palabra a partir del conocimiento que se tiene de la gramática.

El establecimiento de estas combinaciones de palabras prefabricadas se basa en la distinción que Sinclair (1991, p. 110 y ss.) establece entre “principio de la elección libre” (*open choice principle*), que consiste en la capacidad analítica, basada en reglas, de seleccionar palabras individuales, y el “principio de la locución” (*principle of idiom*), de carácter

holístico, que complementa al anterior con recursos ya preestablecidos y almacenados en la mente. Ambos procesos actúan de forma complementaria, y diversos estudios revelan que se asocian a las circunstancias de uso, en función de si las condiciones permiten planear o no el discurso (Moreno Teva, 2012, pp. 40-41; Pérez Serrano, 2017, p. 32).

Por lo que se refiere a las características básicas de las SF, por lo general se alude al carácter poliléxico, la idiomática, la fijación y la independencia, si bien estas propiedades pueden estar sometidas a gradación y no darse en todos los casos (Alvarado Ortega, 2010; Moreno Teva, 2012; Penadés, 1999, 2017; Ruiz Gurillo, 1997; Velázquez, 2018; Corpas, 1996; Zuloaga, 1980).

Igualmente, es importante destacar que son secuencias de palabras que los nativos sienten como la manera preferida y natural de expresar ideas y propósitos, aunque haya otras formas expresivas también posibles (Penadés, 2017; Sánchez Rufat, 2011, 2017). De este modo, el dominio de estas unidades, que pueden llegar a constituir la mitad del discurso nativo, aunque existen amplias diferencias en estas estimaciones (Erman y Warren, 2000; Moreno Teva, 2012, p. 20 y 55), mejora la competencia comunicativa de los aprendientes, pues aporta fluidez, precisión y mayor expresividad al discurso, tanto en la comprensión como en la producción (Martín, 2012, p. 61 y ss.; Moreno Teva, 2012, p. 36 y ss.; Pérez Serrano, 2017, p. 25; Szyndler, 2015, p. 203; Timofeeva, 2013, p. 326 y ss.).

Otra característica significativa para la explicación de estas unidades reside en la trascendencia de los aspectos sociales, discursivos y expresivos, pues las SF incluyen información de tipo sociocultural, reflejando aspectos de la idiosincrasia, la historia, los hábitos y costumbres, así como la manera de pensar y conceptualizar el mundo por parte de las comunidades de habla que las emplean (Moreno Teva, 2012, p. 40; Szyndler, 2015, pp. 203-4; Timofeeva, 2013). Por lo tanto, su desconocimiento o el mal uso puede provocar un discurso “no idiomático” (Lennon, 1998, p. 12), es decir, forzado, ensayado, superficial, que no se ajusta a lo esperado, con el riesgo de romper convenciones de cortesía, llevar a malentendidos y perjudicar el proceso de comunicación (Alvarado Ortega, 2010, p. 47 y ss.; Velázquez, 2018, p. 32).

Por otro lado, se han propuesto numerosas y diversas clasificaciones para estas unidades, pero quizás la más general y extendida sea la de Corpas (1996), quien, partiendo de una concepción más amplia de la fraseología en la que se incluyen todas aquellas combinaciones formadas al menos por dos palabras y cuyo límite superior se sitúa en la oración compuesta, las divide en colocaciones (“correo electrónico”, “paliar el dolor”, “tarjeta sanitaria”), locuciones (“de cuando en cuando”, “mover cielo y tierra”, “sano y salvo”) y enunciados fraseológicos (como citas, refranes y fórmulas rutinarias: “ojo por ojo y diente por diente”, “nos vamos a ir yendo”). Otras clasificaciones clásicas son las Casares (1950), Zuloaga (1980), Ruiz Gurillo (1997, 2000), Erman y Warren (2000), aunque estimamos que la de Corpas (1996) es muy útil desde el punto de vista didáctico, debido a su mayor amplitud y porque la división en tres esferas que esta autora realiza facilita la práctica de estas expresiones.

No obstante, somos conscientes de que asumir dicha clasificación también supone, como veremos más adelante, afrontar la amplia heterogeneidad sintáctica de las unidades léxicas que consideramos SF, algo que representa una dificultad añadida para la obtención de los datos para el estudio. A pesar de ello, estimamos que la rentabilidad comunicativa de la enseñanza y aprendizaje de las SF justifica sobradamente esta decisión.

2.2. Las secuencias formulaicas en el PCIC

Parece evidente, por tanto, la necesidad de integrar estas unidades pluriverbales en la enseñanza de lenguas, aunque la cuestión de cuándo introducirlas suscita cierta controversia, puesto que encontramos opiniones a favor de incluirlas en los niveles intermedio y avanzado y otras que defienden su presencia en todos los niveles. Igualmente, existe un amplio consenso en las dificultades para su aprendizaje debido a sus características formales, semánticas y pragmáticas (Olimpio, 2006, p. 6; Penadés, 2017, pp. 335-6; Timofeeva, 2013, pp. 324-5; Wray, 2000, p. 468).

En un marco más general, el interés en un mejor desarrollo de la competencia comunicativa de los alumnos está en la base del enfoque léxico, desarrollado por Michael Lewis (1993, 1997), para quien “la lengua consiste en léxico gramaticalizado, no en gramática lexicalizada” (Lewis 1993, p. 51). Este autor propone, entre otros postulados, potenciar el aprendizaje de *chunks*, bloques prefabricados de palabras, segmentos discursivos empleados con mucha frecuencia por los hablantes nativos en sus interacciones, y que se corresponderían en líneas generales a las SF.

En este sentido, en los últimos años se viene hablando de “fraseodidáctica” (Crida y Alessandro, 2019; Ettinger, 2008; González Rey, 2012; López Vázquez, 2011, Szyndler, 2015), cuyo objetivo es la descripción de los fenómenos relacionados con las expresiones fraseológicas, así como su enseñanza en lengua materna y en lengua extranjera, atendiendo tanto a la comprensión como a la expresión de fraseologismos (López Vázquez, 2011, p. 533).

No obstante, aunque desde la década de los años 90 del siglo pasado este tema empieza a adquirir importancia y ha aumentado considerablemente la atención que se le dedica, en el ámbito de ELE todavía carecemos de investigaciones que aborden aspectos teóricos y prácticos esenciales en la adquisición y producción de la fraseología (Aguilar Ruiz, 2013; González Rey, 2012; Olimpio, 2006, p. 129 y ss.; Serradilla, 2014), y así “se transfiere al profesor la tarea de investigador, dado que, para llevar a cabo ciertas propuestas didácticas, le corresponde determinar las unidades fraseológicas más frecuentes o rentables, y definir ciertos parámetros discursivos y pragmáticos de estas unidades, tareas que se encuentran en estado embrionario en la fraseología del español” (Olimpio, 2006, pp. 141-2).

2.3. Secuencias formulaicas, corpus lingüísticos y frecuencia léxica

En palabras de Sinclair (2004, p. 16) “a corpus is a collection of pieces of language text in electronic form, selected according to external criteria to represent, as far as possible, a language or language variety as a source of data for linguistic research”¹, aunque no obstante es necesario aclarar que un corpus no constituye la totalidad de la lengua, sino que recoge una muestra representativa de la misma (Cruz Piñol, 2017, p. 35).

En términos generales, podríamos apuntar que la lingüística de corpus (LC), dentro de la lingüística aplicada, se centra en la creación, la explotación, el análisis de textos orales y escritos de un corpus electrónico, así como en sus aplicaciones a otros campos como la elaboración de gramáticas y diccionarios, y también su empleo en disciplinas y fenómenos

¹ “Un corpus es una colección de fragmentos de textos en formato electrónico, seleccionados de acuerdo con criterios externos para representar, en la medida de lo posible, un idioma o una variedad de idiomas como fuente de datos para la investigación lingüística” [la traducción es nuestra].

como la dialectología, la disponibilidad léxica, la traducción, la sociolingüística, etc., (Cruz Piñol, 2017, pp. 77 y ss.).

Como señala Bolaños (2015, pp. 34-5), la LC se inscribe en el mismo paradigma de las disciplinas lingüísticas que, con el giro de la lingüística en los años sesenta del siglo pasado, se dedican a estudiar las diferentes manifestaciones del uso del lenguaje en contextos reales de interacción comunicativa con un énfasis en lo comunicativo y pragmático y con un enfoque lingüístico funcional, frente a los estudios teóricos del sistema de la tradición lingüística anterior.

Aunque en otros ámbitos como el anglosajón este campo evidencia un amplio desarrollo desde hace tiempo, el interés creciente por los corpus lingüísticos se manifiesta ahora como una novedad en ELE, debido principalmente a que son un recurso relativamente reciente en nuestro idioma, y también a que los profesores no suelen poseer formación en LC (Villayandre y Llanos 2017, p. 58). A pesar de este retraso, encontramos obras como la de Cruz Piñol (2017), que explora adecuadamente este surgimiento desde la óptica de ELE, presentando conceptos básicos de la LC, así como diversas aplicaciones a la enseñanza del español, todo ello acompañado por actividades para poner en práctica lo expuesto en la obra. Igualmente, varios autores, incluso ya con perspectiva histórica, y desde la lingüística teórica, analizan los corpus de español a los que podemos acceder como profesionales (Bolaños, 2015; Briz y Albelda, 2009; Cruz Piñol, 2014, 2017; Pérez Serrano, 2017; Rojo 2016).

Desde este punto de vista de la enseñanza y aprendizaje, los corpus constituyen excelentes recursos, tanto para los docentes como para los discentes, que propician la práctica de aspectos pragmáticos, discursivos y culturales, puesto que contribuyen a la contextualización de la lengua que se trabaja en clase, potenciando la interacción comunicativa y la negociación del significado en el aula. Del mismo modo, posibilitan que el alumno esté expuesto a muestras lingüísticas de hablantes nativos de diversas características sociales y dialectales. En cuanto a las SF, propician el desarrollo de la competencia léxica mediante la detección de patrones de palabras (*chunks*) que suelen aparecer juntas (Albelda, 2011; Bolaños, 2015; Buyse, 2017, Cruz Piñol, 2017, Villayandre y Llanos, 2017, pp. 53-55).

En segundo lugar, los corpus son igualmente una valiosa herramienta de información estadística sobre frecuencia de uso de las unidades léxicas, algo que podría ayudar al profesor a establecer prioridades en la enseñanza y seleccionar y gradar, así, los contenidos que son más habituales en la lengua, posibilitando una elección más objetiva y menos intuitiva. En el caso del español, el criterio de la frecuencia para la selección de exponentes ya es asumido en el PCIC, aunque parte de la apreciación intuitiva basada en la experiencia docente, y lo completa con la rentabilidad comunicativa, es decir, que se trate de exponentes que sirvan para realizar las funciones comunicativas propias de cada nivel (PCIC, Nociones específicas, Introducción).

La aplicación de estudios de frecuencia como fuente de información lingüística en general (Cruz Piñol, 2017, p. 25) y a la enseñanza del vocabulario de una L2/LE en particular, tiene una historia relativamente larga (Ferrando, 2012; García Salido, 2017; García y Alonso, 2018, pp. 159-61; Nation, 2001, p. 329) y, precisamente el desarrollo de la LC ha contribuido al renovado interés por el vocabulario que se observa en las últimas décadas, que viene acompañado por la vuelta a los estudios de frecuencia léxica. Por ejemplo, en el ámbito hispánico Alvar Ezquerro (2005) propone que sea la frecuencia lo que determine cuál es el vocabulario que con más urgencia necesita un aprendiz de español, aunque este mismo autor llame la atención sobre el hecho de que “la información sobre la frecuencia de uso no puede ser un valor rígido al que debemos sujetarnos, pues hay otros factores que influyen en el aprendizaje de las palabras” (Alvar Ezquerro, 2005, p. 19).

Es por esto por lo que, aunque la frecuencia ha sido un criterio esencial para la selección del vocabulario fundamental que debe enseñarse, por sí sola puede no ser suficiente, pues por un lado algunas palabras útiles presentan niveles muy bajos de frecuencia y, por otro, las palabras más frecuentes suelen ser las de contenido gramatical. Por tanto, parece necesario combinarla con más parámetros para la selección, como el de la disponibilidad léxica (Tomé Cornejo, 2015 & 2016), si bien los estudios en este ámbito se centran en la palabra, por lo que debería desarrollarse un modelo de encuesta que tuviera en cuenta también las unidades léxicas pluriverbales para determinar qué SF enseñar en cada nivel (Ferrando, 2012, p. 361). Además de la frecuencia, en los estudios sobre este tema se proponen otros criterios con respecto a la selección del léxico, entre otros, la dispersión en el corpus que se consulte (Alvar Ezquerro, 2005), su productividad y los intereses de los aprendices (Gómez Molina, 2004; Pérez Serrano, 2017, p. 56), las intuiciones de los hablantes (Benigno, Kraiff, Grossmann y Velez, 2016; McGee, 2008) o las producciones de los propios aprendices para determinar cuál es el vocabulario propio de los diferentes niveles distinguidos por el MCER (García Salido, 2017. p. 32).

3. METODOLOGÍA

Sobre la base de un trabajo anterior (Martos y Contreras, 2018), el presente análisis parte de la hipótesis de que tanto Internet como los corpus lingüísticos pueden emplearse como herramienta para el cálculo de la frecuencia de uso de las SF en el discurso nativo, y que dicha frecuencia podría, junto a otros criterios, contribuir a que el profesor pueda establecer el nivel en el que deben enseñarse estas unidades léxicas en el aula de ELE.

Para indagar en dicha hipótesis, se ha analizado en diferentes corpus la frecuencia de 3260 SF extraídas del PCIC, estudiando el grado de correlación de su frecuencia con los niveles establecidos en el MCER, desde A1 a C2.

3.1. Selección de los corpus

En primer lugar, siguiendo a Lindstromberg y Boers (2008), recurrimos al motor de búsqueda de *Google* para analizar la frecuencia de estas expresiones en la red. En cuanto al uso de este buscador como corpus, autores como Buyse (2017), Cruz Piñol (2017), González Fernández (2017) o Sinclair (2004) comentan las ventajas y desventajas del empleo de la red como corpus, destacando entre los inconvenientes su continuo crecimiento, con lo que la información cambia constantemente, o que el buscador no haya sido diseñado desde una perspectiva lingüística, por lo que no está lematizado. No obstante, todos los autores reconocen que parece indiscutible que la red ofrece un repertorio de textos y datos infinitamente mayor que cualquier otro corpus existente, por lo que su consideración como un gran repertorio textual está ya asumida por la mayoría de los investigadores.

En el ámbito hispánico destaca el *Corpus de Referencia del Español Actual* (CREA) de la Real Academia Española de la Lengua², compuesto por un amplio repertorio de textos de diversa procedencia, tanto escritos (libros, revistas, periódicos) como orales, procedentes

² <http://corpus.rae.es/creanet.html>. Puede ampliarse la información sobre el corpus en el manual de consulta, accesible en http://corpus.rae.es/ayuda_c.htm

en su mayoría de la radio y la televisión, desde 1975 hasta 2004, y que comprende algo más de 160 millones de formas.

La RAE también trabaja actualmente en el *Corpus del Español del Siglo XXI* (CORPES XXI)³ que, al igual que el CREA, está formado por textos escritos y orales procedentes de España, América, Filipinas y Guinea Ecuatorial, y que en su versión 0.91 (diciembre de 2018) cuenta con más de 285000 documentos que suman alrededor de 286 millones de formas.

Finalmente, otra herramienta importante es el *Corpus del Español* (CdE) de Mark Davies⁴, que consta de dos partes: la primera es el corpus original y más pequeño, que permite buscar cambios históricos y variación de géneros (incluye más de cien millones de palabras procedentes de más de veinte mil textos del español de los siglos XIII al XX), mientras que la segunda es el corpus nuevo, mucho más amplio, y que permite buscar variaciones dialectales. Este segundo contiene casi 5500 millones de palabras de páginas web de veintidós países de habla hispana, y permite hacer búsquedas en textos en español muy recientes, por lo que será el que utilicemos para nuestro estudio.

3.2 Selección de las SF en el PCIC

Por lo que se refiere a las SF seleccionadas para nuestro estudio, las hemos extraído de los catálogos de Nociones Específicas del PCIC (2006), desde el nivel A1 al C2, aunque hemos unido las pertenecientes a los niveles A1 (49) y A2 (69) para que sean estadísticamente representativas. De este modo, nuestro catálogo está formado por 3260 SF (A1-A2 = 118; B1 = 373; B2 = 913; C1 = 952; C2 = 904). Para su selección, hemos adoptado un criterio amplio, incluyendo todos los tipos de SF (colocaciones, locuciones y enunciados fraseológicos) y atendiendo no tanto a sus propiedades semánticas de opacidad o no composicionalidad, como a la rentabilidad comunicativa para el aprendiente. Nuestra decisión se basa en el hecho de que, como señalan diversos autores (Martínez, 2013, p. 187 y ss.), la no composicionalidad de estas unidades léxicas, a pesar de ser un criterio que debe considerarse para complementar a otros, es subjetivo, puesto que la opacidad puede depender de factores individuales de los aprendientes, y no solo de las características de las propias SF, y por lo tanto su interpretación y comprensión puede ser diferente en cada aprendiente. Así, una expresión aparentemente transparente puede resultar opaca por diversas razones para ciertos estudiantes.

De acuerdo con estos criterios, hemos seleccionado SF de diversa tipología y características como “café cortado”, “hijo natural”, “pronóstico reservado”, “reproducción asistida”, “echar un trago”, “entrado en carnes”, “ir hecho un adefesio”, “navegación de recreo”, “pagar a tocateja”, “a bombo y platillo”, “estar como un niño con zapatos nuevos”, “hecha la ley, hecha la trampa”, “ojo por ojo y diente por diente”.

3.3. Obtención y análisis de datos

La recolección de los datos se realizó entre julio y octubre de 2019, y debemos comentar algunos aspectos que inciden en los datos sobre la frecuencia de las SF analizadas. En primer lugar, hemos tenido en cuenta, siempre que ha sido posible, la variación mor-

³ <http://www.rae.es/recursos/banco-de-datos/corpes-xxi>. Para más información sobre la conformación y codificación del corpus véase <https://bit.ly/33Lhx4M>

⁴ <http://www.corpusdelespanol.org>

fológica, tanto de género y número en los sustantivos y adjetivos, como la flexión verbal. Para ello, en las consultas utilizamos las reglas de etiquetado morfológico y lematización que permiten los propios corpus.

En este sentido, en el caso concreto de *Google*, que no está lematizado, hemos contabilizado la variación de género y número realizando búsquedas individuales para singulares, plurales, masculinos, femeninos y sus combinaciones en ejemplos como “entrado en carnes”, “gato callejero”, “hermano gemelo”, “hijo único”, “ministro de defensa”, “niño adoptado”, “ponerse rojo de ira”, etc., pero no ha sido posible, por la ausencia de lematización, atender a la flexión verbal, por lo que en el caso de las SF cuya base es un verbo, hemos empleado el infinitivo como lema.

Igualmente, debido a la amplia heterogeneidad sintáctica de las unidades léxicas que consideramos SF, a las dimensiones de este estudio, y en ocasiones a las propias posibilidades de búsqueda que ofrecen los corpus seleccionados, no se han atendido configuraciones sintácticas concretas como, por ejemplo, en el caso en el que los componentes de la SF no forman una secuencia continua (esto es, cuando entre dichos componentes aparecen modificadores u otros complementos: “mar [un poco, bastante, muy...] picado”, “pagar [siempre, a veces...] a escote”, “relación [un tanto, a veces, bien, de lo más...] tormentosa”), así como en los posibles cambios de orden sintáctico de los constituyentes de la SF (por ejemplo, la anteposición en ciertas colocaciones: “actitud extraña/extraña actitud”, “amigo íntimo/íntimo amigo”, “barba poblada/poblada barba”, “carácter fuerte/fuerte carácter”, “fisonomía extraña/extraña fisonomía”, “sensación amarga/amarga sensación”, “trabajo ímprobo/ímprobo trabajo”). Por todo ello, deben tenerse en cuenta estas consideraciones para la interpretación de los datos ofrecidos en este trabajo.

Finalmente, para el tratamiento estadístico hemos utilizado el programa *IBM SPSS Statistics* (versión 24), el cual nos ha permitido realizar tanto la descripción estadística descriptiva de los datos recabados como la representación gráfica de los mismos.

4. PRESENTACIÓN DE RESULTADOS: LA FRECUENCIA DE LAS SF EN EL PCIC

Como hemos venido señalando, la frecuencia ha sido tomada tradicionalmente como un criterio nivelador para la selección del léxico en ELE, y tal y como señalan García y Alonso (2018, p. 162), si las expresiones comunicativamente más rentables se aprenden en niveles más bajos, debería darse una correlación negativa entre frecuencia léxica y nivel, es decir, que a mayor frecuencia en los corpus, el nivel sería menor y viceversa.

Por tanto, realizaremos en primer lugar una descripción general de los datos extraídos de los corpus analizados para comprobar si se produce dicha correlación negativa entre la frecuencia de las SF analizadas y el nivel asignado en el PCIC, pasando a continuación al estudio descriptivo y comparativo de dicha frecuencia y su distribución en los corpus estudiados, teniendo siempre en cuenta el nivel asignado a las mismas en el PCIC, lo que entendemos que puede arrojar conclusiones interesantes para la nivelación y selección de las SF para su enseñanza y aprendizaje en ELE.

Comenzando por el número de SF en cada nivel, y como podemos apreciar en la Tabla 1, estas cifras (A1-A2=118, B1=373, B2=913, C1=952, C2=904) parecen confirmar las indicaciones del MCER y del PCIC, que recomiendan posponer la enseñanza de SF para niveles altos, aunque dicha recomendación podría no ser adecuada teniendo en cuenta la rentabilidad comunicativa de estas unidades léxicas en los niveles iniciales.

Tabla 1. Estadística descriptiva. Datos generales.

	A1-A2 (n=118)	B1 (n=373)	B2 (n=913)	C1 (n=952)	C2 (n=904)
GOOGLE					
media	26341796,61	13801417,69	6616574,93	3607177,44	1952648,73
desv. est.	62994255,45	31141526,62	17767531,96	12762003,91	9383841,71
mediana	7388000	3244000	1260000	451500	137000
mín.	168000	12100	4	119	0
1 ^{er} cuart.	2446500	761000	364250	85625	21900
3 ^{er} cuart.	19400000	13739000	5990000	2568250	840950
máx.	498000000	316000000	276000000	295000000	217000000
CREA					
media	288,73	152,60	123,58	71,35	54,86
desv. est.	744,6	538,17	312,70	145,67	205,54
mediana	59	39	35	21	11
mín.	0	0	0	0	0
1 ^{er} cuart.	18	13	9	5	2
3 ^{er} cuart.	205,75	91	106	74,75	38,25
máx.	5649	7998	5427	2078	4274
CORPES					
media	658,65	400,41	260,98	153,57	82,22
desv. est.	1326,40	1141,89	791,54	353,91	315,32
mediana	174	83	63	40	15,5
mín.	0	0	0	0	0
1 ^{er} cuart.	51,25	25	16	9	2
3 ^{er} cuart.	600,75	278	210	149,5	63
máx.	7293	13687	17281	3776	6106
CdE					
Media	18777,28	9007,52	7159,28	4711,68	2818,73
desv. est.	59224,42	32437,44	33007,67	20675,79	14253,77
Mediana	2244,5	1177	1041	488	201
mín.	1	0	0	0	0
1 ^{er} cuart.	517	288	195	92,5	31
3 ^{er} cuart.	6506	4642	4333	2376,5	1017
máx.	476454	435498	801302	365368	247772

Para confirmar estadísticamente dicha correlación negativa entre el nivel asignado en el PCIC a las SF de nuestra muestra y su frecuencia en los corpus manejados, hemos calculado el coeficiente τ_b de Kendall, ya que dicho coeficiente establece la correlación entre dos rangos (en nuestro caso, el nivel y la frecuencia). Esta expectativa se confirma con una correlación negativa en todos los corpus: *Google* ($\tau_b = -0,342$), CREA ($\tau_b = -0,192$), CORPES ($\tau_b = -0,243$) y CdE ($\tau_b = -0,208$), y aunque el valor de la correlación sea bajo, sí que es estadísticamente muy significativo ($p < 0,0001$). Además, dicho grado de correlación podría deberse, como comentaremos con más detalle a continuación, al tipo de distribución de las frecuencias en cada grupo.

Por tanto, podemos corroborar lo que ya se apunta tanto en un trabajo anterior (Autor, 2018) como en García y Alonso (2018), esto es, que la nivelación realizada por el PCIC, a pesar de basarse en un criterio intuitivo, es bastante fiable, puesto que a medida que desciende la frecuencia de las SF, más alto es el nivel asignado.

En este sentido, ya Lindstromberg y Boers (2008) realizaron una estimación semejante tomando como base la frecuencia en *Google*. Para estos autores, una SF cuya busca se situara en tres millones de resultados podía considerarse como frecuente; si arrojaba un resultado de 230000 se encontraba en el nivel medio de frecuencia y, por último, si contaba con 400, sería una frecuencia rara. Basándose en esta clasificación, estos autores señalan que la mayoría de SF se encuentra en los dos últimos grupos, por lo que son las que deberían tratarse principalmente en el aula. El resto, con un índice de frecuencia muy alto, posiblemente se adquieran fuera de clase al escucharlas o leerlas frecuentemente, así que quizá no fuera rentable trabajarlas en clase.

Dejando de lado esta última consideración didáctica, los cálculos de Lindstromberg y Boers (2008) nos parecen muy interesantes como primera referencia a la hora de nivelar una SF por macroniveles, estimando como frecuentes las SF de los niveles A1-A2, de un nivel medio de frecuencia las propias de B1-B2, y por último las de C1-C2, que presentarían una frecuencia baja, tal y como se aprecia en la Figura 1.

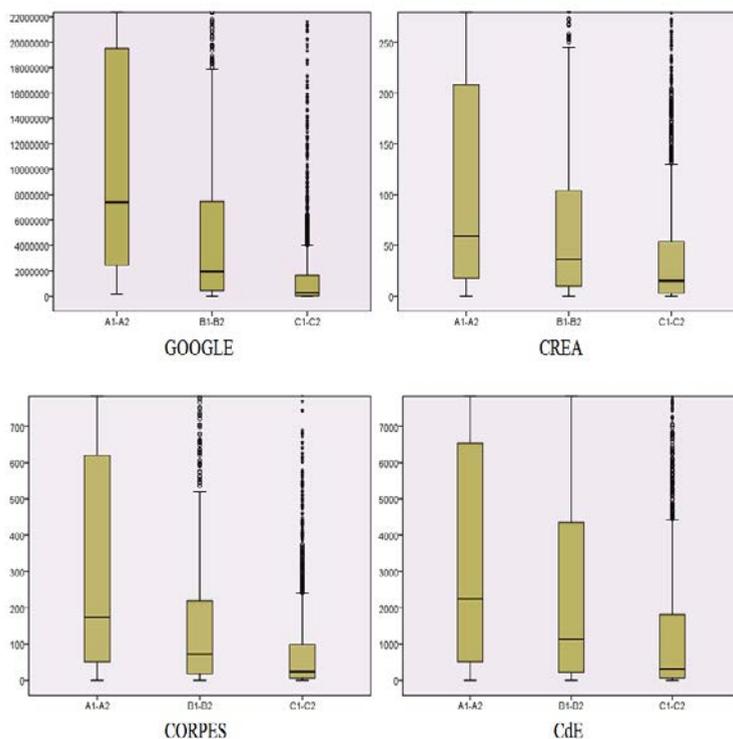


Figura 1. Distribución de frecuencias por nivel y por corpus (macroniveles).

Con este propósito, esto es, para ofrecer un punto de referencia, que no de corte, para la nivelación de una SF dada, presentamos en la Tabla 2 los datos extraídos de nuestra muestra tomando como eje central la mediana de cada grupo, puesto que este dato estadístico parece más fiable en distribuciones con tendencia paretiana, aunque también ofrecemos el dato de la media para poder comparar las diferencias entre ambas cifras. Dicha tendencia paretiana, como se observa en las cifras de la Tabla 1 y en los gráficos de la Figura 1, se produce cuando las frecuencias no presentan una distribución simétrica (que se da cuando la media, moda y mediana coinciden), sino que existe una gran asimetría, es decir, que el 20% de las observaciones arrojan un valor superior al 80% restante de la muestra. Esto confirma la ley de Zipf (1935), que establece que, en cualquier conjunto analizado, un reducido número de unidades léxicas presenta altos índices de frecuencia, mientras que un gran número son poco frecuentes.

Tabla 2. Frecuencias macroniveles.

	GOOGLE		CREA		CORPES		CdE	
	mediana	media	mediana	Media	mediana	media	mediana	media
A1-A2 (n=118)	7388000	26341796,6	59	288,73	174	658,65	2244,5	18777,28
B1-B2 (n=1286)	1932000	8711920,01	36	131,99	72	301,45	1125	7695,35
C1-C2 (n=1856)	271000	2801307,86	15	63,31	24,5	118,71	310,5	3798,16

Como se puede observar a simple vista, existen grandes diferencias entre los datos de *Google* ofrecidos por Lindstromberg y Boers (2008) y los arrojados por nuestro estudio, hecho que debemos atribuir sin duda al inmenso crecimiento de la información en Internet desde la publicación de dicho estudio⁵.

No obstante, estimamos que estos datos pueden servir de punto de partida y de referencia general para la nivelación de las SF, siendo necesario complementarlos con otros criterios como la dispersión, la rentabilidad comunicativa, la opacidad o los intereses de los alumnos para una nivelación más precisa de las SF que el profesor debe seleccionar para su trabajo en el aula.

5. CONCLUSIONES

A la luz de lo expuesto hasta aquí, podemos volver a confirmar, como ya hicimos en Martos y Contreras (2018), y como igualmente confirman otros análisis como el realizado por García y Alonso (2018), que la nivelación realizada por el PCIC de las SF, a pesar de basarse en un criterio intuitivo, es bastante fiable.

No obstante, a pesar de que tanto el MCER como el PCIC recomiendan posponer la presentación de estas unidades léxicas a niveles altos, esto podría no ser adecuado, especialmente cuando la expresión presenta una frecuencia alta, por lo que diversos autores (Martínez, 2013) proponen priorizar la enseñanza de las expresiones pluriverbales a la vez frecuentes y poco transparentes. Por tanto, estimamos que las SF deberían incluirse con mayor profusión en los repertorios de los niveles iniciales, debido a su alta rentabilidad comunicativa, avalada por su alta frecuencia, por lo que su aprendizaje contribuiría a la mejora tanto de la fluidez como de la comprensión de los aprendientes.

Por lo que se refiere a los resultados de nuestro estudio, estimamos haber contribuido a confirmar la pertinencia de utilizar la frecuencia para orientar en la nivelación de las SF.

⁵ Según World Wide Web Size (<https://www.worldwidewebsize.com/>) e *Internet Live Stats* (<https://www.internetlivestats.com/>) solo en cinco años, entre 2008 y 2013, la cantidad de información creció un 900%, alcanzando en 2016 la cifra de un Zettabyte (1 zetabyte = dos billones de gigabytes), y se espera que para 2021 alcancemos los 3,3 Zettabytes.

En este sentido, entendemos igualmente que *Google* puede ser empleado como herramienta para este mismo objetivo, pues los contenidos disponibles en la red constituyen una fuente de información adecuada para dicho fin.

No obstante, debido a la especial distribución de las frecuencias en los distintos niveles, que como hemos visto confirma la ley de Zipf (1935), se hace necesario sumar la frecuencia a otros criterios, pues tal y como apuntaba Alvar Ezquerro (2005, p. 19), “la información sobre la frecuencia de uso no puede ser un valor rígido al que debemos sujetarnos, pues hay otros factores que influyen en el aprendizaje de las palabras”, entre otros la dispersión en el corpus en el que se consulte, la rentabilidad y productividad, la opacidad o transparencia, los factores culturales, los intereses de los alumnos, las propias producciones de los estudiantes, o incluso las intuiciones de los hablantes nativos.

El análisis de esos otros posibles criterios de dispersión constituye una de las evidentes limitaciones de este trabajo, y que puede servir de base para futuras investigaciones. En ellas, igualmente, habría que contabilizar las ocurrencias de configuraciones sintácticas concretas de las SF, así como los casos de posibles cambios en el orden sintáctico de sus constituyentes.

Finalmente, los rápidos y continuos cambios que se producen en el volumen y la tipología de información en los corpus, siendo *Google* el máximo exponente de esta vertiginosa transformación, podrían imposibilitar la reproducción de los resultados, algo totalmente necesario a la hora de confirmar las conclusiones de cualquier investigación. No obstante, como matiza González (2017, p. 132), este hecho “no es sino un reflejo de la naturaleza dinámica del lenguaje. Si la realidad cambia, los resultados de los trabajos deben reflejar esta evolución, lo que constituirá no un error en las conclusiones de trabajos anteriores, sino nuevos resultados”. Además, como también señala este mismo autor, las herramientas tecnológicas con las que contamos actualmente pueden reducir este problema, ya que permiten almacenar la información en bases de datos y recuperarla en cualquier momento para su comparación con los nuevos datos.

6. REFERENCIAS BIBLIOGRÁFICAS

- Aguilar Ruiz, M. J. (2013). Notas sobre las posibilidades de aprendizaje de español mediante unidades fraseológicas. *MarcoELE*, 17, disponible en: <https://bit.ly/338dCOg>
- Albelda Marco, M. (2011). Rentabilidad de los corpus discursivos en la didáctica de lenguas extranjeras. En J. de Santiago, H. Bongaerts, J. J. Sánchez y M. Seseña (coords), *Del texto a la lengua: la aplicación de los textos a la enseñanza-aprendizaje del español L2-LE* (pp. 83-95). Salamanca: ASELE, disponible en: <https://goo.gl/5oEXpR>
- Alvarado Ortega, M.^a B. (2010). *Las fórmulas rutinarias del español: teoría y aplicaciones*. Peter Lang.
- Alvar Ezquerro, M. (2005). La frecuencia léxica y su utilidad en la enseñanza del español como lengua extranjera, en M.^a A. Castillo Carballo (Coord.), *Las gramáticas y los diccionarios en la enseñanza del español como segunda lengua: deseo y realidad. Actas del XV Congreso Internacional de la Asele* (pp. 19-39). Universidad de Sevilla, disponible en: <https://bit.ly/2XDRCdI>
- Benigno, V., Kraiff, O., Grossmann, F. & Velez, A. (2016). La notion de collocation fondamentale: Une étude de corpus. *Cahiers de Lexicologie*, 108(1), 125-146.

- Bolaños Cuéllar, S. (2015). La lingüística de corpus: perspectivas para la investigación lingüística contemporánea. *Forma y Función*, 28(1), 31-54.
- Briz Gómez, A. & Albelda Marco, M. (2009). Estado actual de los corpus de lengua española hablada y escrita: I+D. En *El español en el mundo. Anuario 2009*. Madrid: Instituto Cervantes, disponible en: <https://bit.ly/2txdcBo>
- Buyse, K. (2017). Los corpus como herramientas de aprendizaje del léxico, en VV. AA., *Enseñar léxico en el aula de español. El poder de las palabras* (pp. 121-141). Barcelona: Difusión.
- Casares, J. (1950/1992). *Introducción a la lexicografía moderna*. Madrid, CSIC.
- Consejo de Europa. (2002). *Marco común europeo de referencia para las lenguas aprendizaje, enseñanza, evaluación*. Anaya y CVC, disponible en: <https://bit.ly/2IkNko4>
- Corpas Pastor, G. (1996). *Manual de fraseología española*. Gredos.
- Crida, A. & Alessandro, A. (Eds.). (2019). *Innovación en fraseodidáctica. Tendencias, enfoques y perspectivas*. Peter Lang.
- Cruz Piñol, M. (2014). Veinte años de tecnologías y ELE. Reflexiones en torno a la enseñanza del español como lengua extranjera en la era de Internet. *MarcoELE*, 19, disponible en: <https://bit.ly/2Kc8901>
- Cruz Piñol, M. (2017). *Lingüística de corpus y enseñanza de español como 2/L*. Arco Libros.
- Erman, B. & Warren, B. (2000). The idiom principle and the open-choice principle. *Text*, 20, 87-120.
- Ettinger, S. (2008). Alcances e límites de fraseodidáctica. Dez preguntas clave sobre el estado actual de investigación. *Cadernos de Fraseología Galega* 10, 95-127.
- Ferrando Aramo, V. (2012). *Aspectos teóricos y metodológicos para la compilación de un diccionario combinatorio destinado a estudiantes de E/LE*. Tesis doctoral. Universitat Rovira i Virgili, Tarragona, disponible en: <https://bit.ly/34c6wtm>
- García Salido, M. (2017). La frecuencia de corpus como criterio para nivelar colocaciones léxicas. *Études Romanes de Brno*, 38/2, 29-49.
- García Salido, M. & Alonso Ramos, M. (2018). Asignación de niveles de aprendizaje a las colocaciones del Diccionario de Colocaciones del español. *Revista Signos. Estudios de Lingüística*, 51(97) 153-174.
- Gómez Molina, J. R. (2004). Los contenidos léxico-semánticos. E J. Sánchez Lobato e I. Santos Gargallo (Eds.), *Vademécum para la formación de profesores, Enseñar español como segunda lengua (L2)/lengua extranjera (LE)* (pp. 789-810). Madrid: SGEL.
- González Fernández, A. (2017). La web como corpus: un esbozo. *Lengua y Habla*, 21, 126-150.
- González Rey, M.^a I. (2012). De la didáctica de la fraseología a la fraseodidáctica. *Paremia*, 21, 67-84.
- Instituto Cervantes (2006). *Plan Curricular del Instituto Cervantes. Niveles de referencia para el español*. Madrid: Instituto Cervantes, Biblioteca Nueva, disponible en: <https://bit.ly/21Whigd>
- Lennon, P. (1998). Approaches to the teaching of idiomatic language. *IRAL*, XXXVI/1, 11-30.
- Lewis, M. (1993). *The lexical approach. The state of ELT and a way forward*. London: Teacher Training.
- Lewis, M. (1997). *Implementing the lexical approach. Putting theory into practice*. London: Teacher Training.
- Lindstromberg, S., & Boers, F. (2008). *Teaching Chunks of Language. From Noticing to Remembering*. Helbling Languages.
- López Vázquez, L. (2011). La competencia fraseológica en los textos de los manuales de ELE de nivel superior". En E J. de Santiago Guervós et al. (Eds.), *Del texto a la lengua: la aplicación de los textos a la enseñanza-aprendizaje del español L2-LE*, (pp. 531-542). Salamanca: ASELE, disponible en: <https://bit.ly/2quqd1a>

- Martín Noguero, M. (2012). ¿Qué se dice en español cuando...? Las fórmulas rutinarias y las situaciones sociales de comunicación en los niveles iniciales. En O. Abenójar (Ed.), *Actas del III simposio internacional de didáctica del español para extranjeros* (pp. 57-64). Argel: Instituto Cervantes, disponible en: <https://goo.gl/PZcsdB>
- Martínez, R. (2013). A framework for the inclusion of multi-word expressions in ELT. *ELT Journal*, 67, 184-198.
- Martos Eliche, F. y Contreras Izquierdo, N.M. (2018). El empleo de corpus para el aprendizaje de secuencias formulaicas en ELE/EL2. La frecuencia de uso en el nivel B2 del PCIC. *CHIMERA: Romance Corpora and Linguistic Studies*, 5/1, 1-26, disponible en: <https://bit.ly/2WP71Z9>
- McGee, I.D. (2008). Word frequency estimates revisited – A response to Alderson (2007). *Applied Linguistics*, 29(3), 509-514.
- Moreno Teva, I. (2012). *Las secuencias formulaicas en la adquisición de español L2*. Tesis doctoral. Suecia: Universidad de Estocolmo, disponible en: <https://bit.ly/2OdGyQH>
- Nation, P. (2001). *Learning vocabulary in another language*. Cambridge University Press.
- Olímpio de Oliveira, M.ª E. (2006). Fraseología y enseñanza de español como lengua extranjera. *RedELE*, 5, disponible en: <https://bit.ly/33bVnHP>
- Penadés Martínez, I. (1999). *La enseñanza de las unidades fraseológicas*. Madrid: Arco Libros.
- Penadés Martínez, I. (2017). La enseñanza de las unidades fraseológicas, en A. M.ª Cestero e I. Penadés (eds.), *Manual del profesor de ELE*, (pp. 311-355). Servicio de Publicaciones de la Universidad de Alcalá.
- Pérez Serrano, M. (2015). Tratamiento de la combinatoria léxica en documentos de referencia y curriculares: el caso del MCER y del PCIC. *Revista de Investigación Lingüística*, 18, 213-232.
- Pérez Serrano, M. (2017). *La enseñanza-aprendizaje del vocabulario en ELE desde los enfoques léxicos*. Arco Libros.
- REAL ACADEMIA ESPAÑOLA. *Corpus de referencia del español actual* (CREA), disponible en: <http://www.rae.es/recursos/banco-de-datos/crea>
- REAL ACADEMIA ESPAÑOLA. *Corpus del Español del Siglo XXI* (CORPES), disponible en: <http://www.rae.es/recursos/banco-de-datos/corpes-xxi>
- Rojó, G. (2016). Los corpus textuales del español, en J. Gutiérrez-Rexach (Ed.), *Enciclopedia lingüística hispánica* (pp. 285-296). London: Routledge.
- Ruiz Gurillo, L. (1997). *Aspectos de fraseología teórica española*. Valencia: Universidad, Anejo XXIV de *Cuadernos de Filología*.
- Ruiz Gurillo, L. (2000). La fraseología, en A. Briz y Grupo VAL.ES.CO. (Coords.), ¿Cómo se comenta un texto coloquial? (pp. 169-189). Barcelona: Ariel.
- Sánchez Rufat, A. (2011). Léxico gramaticalizado y lengua formulaica: algunas precisiones al enfoque léxico. *Sintagma*, 23, 85-98.
- Sánchez Rufat, A. (2017). Estrategias para la enseñanza de secuencias formulaicas en el aula de español como lengua extranjera. E D. G. Nikleva (Ed.), *La formación del profesorado de español como lengua extranjera. Necesidades y tendencias* (pp. 257-282). Peter Lang.
- Serradilla Castaño, A. (2014). La fraseología en el aula de ELE: nuevos enfoques y propuestas didáctica. *RedELE* (volumen monográfico), VV.AA. (eds.), ¿Qué necesitamos en el aula de ELE?: reflexiones en torno a la teoría y la práctica (pp. 73-98), disponible en: <https://bit.ly/2D8wcvh>
- Sinclair, J (1991). *Corpus, Concordance and Collocation*. Oxford: Oxford University Press.

- Sinclair, J. (2004). Corpus and text - Basic principles, en M. Wynne (Ed.), *Developing Linguistic Corpora: a Guide to Good Practice* (pp. 1-16). Oxford: Oxbow Books, disponible en: <https://bit.ly/2XGFtUG>
- Sinclair, J. & Renouf, A. (1985). A lexical learning syllabus for language, en C. Ronald y M. McCarthy (Eds.), *Vocabulary and Language Teaching* (pp. 140-160). Longman.
- Szyndler, A. (2015). La fraseología en el aula de E/LE: ¿un reto difícil de alcanzar? Una aproximación a la fraseodidáctica. *Didáctica. Lengua y Literatura*, 27, 197-216.
- Timofeeva, L. (2013). La fraseología en la clase de lengua extranjera: ¿misión imposible? *Onomázein*, 28, 320-336.
- Tomé Cornejo, C. (2015). *Léxico disponible. Procesamiento y aplicación a la enseñanza de ELE*. Tesis doctoral, disponible en <https://bit.ly/2Uhod8n>
- Tomé Cornejo, C. (2016). Disponibilidad léxica y enseñanza de vocabulario, en N. Domínguez, M.ª C. Fernández y J. L. García (coords.), *Innovación y desarrollo en español como lengua extranjera* (pp. 89-105). Universidad de Salamanca.
- Velázquez Puerto, K. (2018). *La enseñanza-aprendizaje de fraseología en ELE*. Arco Libros.
- Villayandre Llamazares, M. & Llanos Casado, L. (2017). Los corpus electrónicos en la clase de español: reflexiones y aplicaciones, en B. Peña y A. Mª. Aguilar (Coords.), *Didáctica de la lengua y la literatura. Buenas prácticas docentes*. Tomo I (pp. 52-79). Madrid: ACCI.
- Wray, A. (2002). *Formulaic Language and the Lexicon*. Cambridge University Press.
- Zipf, G. (1935). *The Psycho-Biology of Language*. Cambridge: MIT Press.
- Zuluaga, A. (1980). *Introducción al estudio de las expresiones fijas* (Tesis doctoral inédita). Max Hueber, Verlag.