

# El empleo de corpus para el aprendizaje de secuencias formulaicas en ELE/EL2. La frecuencia de uso en el nivel B2 del PCIC

Fermín Martos Eliche<sup>o</sup>, Narciso M. Contreras Izquierdo\*

<sup>o</sup>Universidad de Granada, \*Universidad de Jaén

*Formulaic sequences* (FS), combinations of words that the natives memorize and use holistically to express a particular idea in certain communicative contexts, and that can be more than half of the speech, stand out in the lexical sub competence, central element of the development of communicative competence. The lexical approach determines that the mastery of these FS (*chunks*) has a positive effect on the improvement of the fluency in the communicative interaction of the learners. For Spanish as a foreign or second language (SFL), the *Plan Curricular del Instituto Cervantes* (PCIC) offers a non-exhaustive selection of these FS as an example of lexical units in each of the reference levels (A1-C2). For their selection, the criteria are frequency and communicative profitability, but it is done through an intuitive assessment based on the teaching experience. In this research, we set out to confirm the adequacy of this intuitive leveling, analyzing the frequency of FS of level B2 of the PCIC in various corpus available at present. Likewise, in an exploratory way, we will expose the usefulness of these corpus for data-driving learning (DDL), with real examples of these FS of the target language.

**Keywords:** Teaching of the lexicon in SFL, formulaic sequences, linguistic corpus, data-driving learning.

## 1. Introducción

Durante los últimos años, la competencia léxica se ha ido constituyendo en un elemento central del desarrollo de la competencia comunicativa. De entre los distintos tipos de unidades que integran el componente léxico, destacan las denominadas *secuencias formulaicas* (SF): combinaciones de palabras que se almacenan y recuperan de la memoria como un todo en el momento de su uso (Wray 2002: 9), y que los nativos emplean de forma natural para expresar una

idea particular en contextos comunicativos determinados (Sánchez Rufat 2017: 258).

Siguiendo los planteamientos del *enfoque léxico*, desarrollado por Michael Lewis (1993, 1997), el dominio de estas SF (*chunks*), que pueden llegar a constituir más de la mitad del discurso nativo, repercute positivamente en la competencia comunicativa de los aprendientes, mejorando su fluidez en la interacción comunicativa (Martín 2012: 61 y ss.).

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). En cuanto a 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.

Teniendo en cuenta esto último, en esta investigación nos planteamos si dicha apreciación intuitiva es suficiente para una adecuada adscripción de las SF a los niveles en el PCIC. Para confirmar la adecuación de dicha nivelación, analizaremos la frecuencia de SF del nivel B2 tanto en *Google* como en diversos corpus de español disponibles en la actualidad.

Igualmente, y solo de modo exploratorio, expondremos la utilidad de estos corpus para el aprendizaje de las SF, en la línea del denominado “aprendizaje lingüístico impulsado por datos” (*data-driven learning* o DDL), que se centra en el aprendizaje autónomo e inductivo a través de la reflexión sobre ejemplos reales de la lengua extraídos de corpus de la lengua meta (Cortina 2009: 109-110), contribuyendo así a la práctica de aspectos pragmáticos, discursivos y culturales mediante la exposición del alumno a muestras lingüísticas reales.

## 2. Competencia léxica, secuencias formulaicas y enfoque léxico

El *Marco común europeo de referencia para las lenguas* (MCER 2002: 126) define la competencia léxica como el conocimiento del vocabulario de una lengua y la capacidad para utilizarlo, algo que representa un importante reto para los profesores ELE/EL2, debido a la amplísima gama de conocimientos y destrezas sobre su forma, significado y uso necesarios para ello.

No es por tanto de extrañar que el nivel léxico haya centrado la atención en los últimos años de la lingüística y de la adquisición de lenguas, y que actualmente se considere el léxico como componente central en el discurso, de ahí la

necesidad de desarrollar nuevos avances metodológicos para su enseñanza y aprendizaje.

La competencia léxica está formada por diversos tipos de unidades: palabras simples y compuestas, fórmulas rutinarias o expresiones institucionalizadas, colocaciones, compuestos sintagmáticos, locuciones idiomáticas o modismos (Gómez Molina 2004). Dicha diversidad en la composición de esta competencia léxica ya aparece recogida tanto en el MCER como en el PCIC. En el primer caso (MCER), se indica que los elementos léxicos comprenden, además de unidades léxicas simples, un gran número de unidades fraseológicas o pluriverbales “que se utilizan y se aprenden como un todo”, y que incluyen fórmulas fijas (exponentes directos de funciones comunicativas como los saludos, refranes, proverbios, arcaísmos residuales), modismos (metáforas lexicalizadas, intensificadores, ponderativos o epítetos, estructuras fijas) y otras frases hechas (verbos con régimen preposicional, locuciones prepositivas, régimen semántico). Por su parte, el PCIC maneja igualmente un concepto amplio de unidad léxica “en la línea de investigación que parte de la base de que un hablante cuenta, además de con unidades léxicas simples o palabras, con un número amplio de bloques semiconstruidos que puede combinar al hablar”, de ahí que incluya en los inventarios nocionales una amplia serie de unidades léxicas pluriverbales como colocaciones y expresiones idiomáticas (sobre todo locuciones), bloques que se procesan holísticamente, y no solo palabra a palabra.

Como podemos apreciar, entre estas unidades destacan las pluriverbales, denominadas *secuencias formulaicas* (SF), para las que existen diversas y variadas definiciones. Una de la más extendida en la actualidad es la de Wray (2002: 9), para quien estas expresiones tienen en común el hecho de ser 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.

Igualmente, es importante destacar que son secuencias de palabras que los nativos sienten como la manera preferida y natural de expresar una idea particular o propósito, aunque haya otras formas expresivas también posibles (Sánchez Rufat 2011, 2017). De este modo, el dominio de estas unidades, que pueden llegar a constituir la mitad del discurso nativo (Erman y Warren 2000), mejora la competencia comunicativa de los aprendientes, pues aporta fluidez y precisión al discurso, tanto en la comprensión como en la producción (Martín 2012: 61 y ss.).

No obstante, y a pesar de su evidente interés en el desarrollo de la competencia léxica de los alumnos, faltan aplicaciones significativas de los avances

teóricos y de los resultados de investigaciones empíricas sobre estas unidades que promuevan un aprendizaje efectivo. En este sentido, una orientación metodológica adecuada para la enseñanza y aprendizaje de estas fórmulas es el enfoque léxico, desarrollado por Michael Lewis (1993, 1997)<sup>1</sup>. Para dicho autor, el léxico es el eje central del proceso de desarrollo de la competencia comunicativa (“la lengua consiste en léxico gramaticalizado, no en gramática lexicalizada”, Lewis 1993: 51), y para ello propone partir del aprendizaje de bloques prefabricados de palabras (*chunks*), segmentos léxicos empleados con mucha frecuencia por los hablantes nativos en sus interacciones.

Lewis sugiere que haya una mayor atención en la presentación del léxico en las clases de lengua extranjera y, por supuesto, un trabajo explícito proporcionando siempre un contexto (situación extralingüística) y un cotexto (entorno lingüístico). De ese modo, los aprendientes no solo deben conocer el significado de las palabras, sino también el uso y las restricciones de cada unidad léxica.

Como vemos, estas secuencias formulaicas representarían los segmentos léxicos (*chunks*) tratados por Lewis, y su tratamiento en el aula sería muy productivo, puesto que si los hablantes los perciben y almacenan en su lexicón de forma holística, como una unidad léxica, y asociados a contextos comunicativos determinados, su aprendizaje repercutiría positivamente en la competencia comunicativa de los aprendientes: al percibir, entender, recuperar y producir estas unidades como un bloque mejorará su fluidez y nivel de comprensión en la interacción comunicativa.

### 3. Uso de TIC y corpus en la enseñanza del léxico en ELE/EL2

Aunque parezca un tópico repetido hasta la saciedad, las Tecnologías de la Información y la Comunicación (TIC) están revolucionando, todavía más que antes, el proceso de enseñanza-aprendizaje de lenguas. En el ámbito de ELE, su implementación en el aula se va haciendo cada vez más necesaria, ya que sus conocidas ventajas superan con creces sus inconvenientes (Sitman 1998; Higueiras 2004; Pastor 2004; Autor, 2008). Muchas posibilidades de implementación han sido ya estudiadas, y así, desde finales de los noventa (Llisterri 1997) hasta la actualidad (Cruz Piñol 2017) hemos asistido a reflexiones sobre la pertinencia del enfoque comunicativo (Soria y Luzón 1999), el *e-learning* (Juan 2009), el Material Didáctico Digital (Juan 2016), la llamada *web social* (Hernández 2012), el uso de *blogs*, *wikis* y *podcast* (Herrera y Conejo 2009), trabajo desde diferentes perspectivas didácticas (Simons 2010), propuestas personales innova-

---

<sup>1</sup> Para conocer más sobre enfoque léxico y su aplicación en ELE, véase Pérez Serrano (2017).

doras de todo tipo (Jiménez, Mora y Cuadros 2016) e implementación de recursos propios de otros terrenos educativos como las *webquest* o cazas del tesoro (Hernández 2014).

Desde el punto de vista de la educación en general, se prevé que las tendencias hasta el 2020 apunten al aprendizaje móvil y la pedagogía inversa o *Flipped Classroom* (Pérez Tornero y Pi 2014), englobándose todo ello en lo que se denomina “pedagogía emergente”: “el conjunto de enfoques e ideas pedagógicas, todavía no bien sistematizadas, que surgen alrededor del uso de la TIC en educación y que intentan aprovechar todo su potencial comunicativo, informacional, colaborativo, interactivo, creativo e innovador en el marco de una nueva cultura del aprendizaje” (Adell y Castañeda 2012: 15).

Es por esto por lo que actualmente se aprecia un cambio en la denominación, extendiéndose la de Tecnologías del Aprendizaje y Conocimiento (TAC), que se refiere tanto a la variedad de tecnologías con las que puede trabajar el profesorado de L2 como a las competencias que alumnos y docentes han de tener para convertirse en agentes del cambio social y educativo (Román-Mendoza 2018: 12)<sup>2</sup>.

No cabe duda del reto que esto representa para el profesorado, al que se le exige cada vez más una preparación digital, estipulada incluso por el Instituto Cervantes como competencia clave del profesorado (Instituto Cervantes 2012). No obstante, es importante precisar que dicha preparación no solo es necesaria y deseable por parte de los docentes, sino también por parte del alumnado. Igualmente, es necesario señalar que “el grado de sofisticación y penetración de las tecnologías actuales implica que su implementación en la enseñanza-aprendizaje solo puede tener éxito si sus usuarios disponen de una competencia digital no solamente instrumental, sino que también abarque sus dimensiones cognitiva, socio-comunicativa, axiológica y emocional” (Román-Mendoza 2018: 166).

En cuanto a las ventajas de empleo de las TIC en clase de lengua, estas promueven un aprendizaje activo, autónomo y en colaboración, en relación con lo que en la actualidad se denomina “aprendizaje ubicuo”:

En lugar de que el instructor utilice el tiempo de clase para distribuir información, esa labor es realizada por cada estudiante después de clase y podría ser en forma de ver conferencias en vídeo, escuchar podcast, hojear los contenidos mejorados de un libro electrónico o colaborar con sus compañeros en las comunidades en línea. (...) El profesor puede dedicar más tiempo a interactuar con cada individuo. Después de la clase, los

---

<sup>2</sup> Esta obra también incluye una serie de recursos complementarios, entre ellos corpus lingüísticos, para facilitar la implantación de las TAC en el aula.

---

estudiantes gestionan el contenido que utilizan, el ritmo y el estilo del aprendizaje, y las formas en las que demuestran sus conocimientos (Johnson *et al.* 2015: 38-39).

Frente a los que propugnan que dicha autonomía del alumno puede hacer peligrar el papel del profesor en el proceso de enseñanza-aprendizaje, asumimos la opinión de Román-Mendoza (2018: 167): “se defiende que el aprendizaje autónomo siempre ha de estar contextualizado, que requiere la participación activa del profesorado y el alumnado, y que no se produce en aislamiento, sino que es interdependiente”.

Convencidos de los beneficios del empleo de las TIC en la enseñanza de lenguas, pensamos que el uso de la lingüística de corpus en ELE puede aportar importantes beneficios en este campo, pues pueden constituir herramientas didácticas eficaces para favorecer el desarrollo de la competencia comunicativa de nuestros alumnos. No obstante, 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, y así, encontramos obras como la de Cruz Piñol (2017), que explora adecuadamente este surgimiento desde la óptica ELE, presentando conceptos básicos de la lingüística de corpus, 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 (Briz y Albelda 2009; Cruz Piñol 2014, 2017; Nicolás 2012; Rojo 2016).

Los corpus constituyen recursos 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 (Albelda 2011: 87 y ss.). Desde la perspectiva del uso de muestras reales de lengua (*realia*), el MCER afirma que el alumno desarrollará su vocabulario mediante la simple exposición a palabras y expresiones hechas utilizadas en textos auténticos de carácter hablado y escrito (6.4.7.1). La selección léxica debe basarse en la elección de textos auténticos hablados y escritos y aprender o enseñar todas las palabras que contienen y se deben seguir unos principios léxicos-estadísticos que seleccionen las palabras más frecuentes en recuentos generales y amplios (6.4.7.3).

Por su parte, el *Plan Curricular del Instituto Cervantes* (2006) afirma:

Como criterios de selección, siempre partiendo de la apreciación intuitiva basada en la experiencia docente, han primado la frecuencia y la rentabilidad comunicativa, es decir, que se tratara de exponentes nocionales necesarios para realizar las funciones comunicativas y para desenvolverse con los textos orales y escritos correspondientes a cada nivel (Nociones específicas).

Más concretamente, Cruz Piñol (2017) señala la utilidad de los corpus al proporcionar concordancias y coapariciones (ocurrencias en contexto para el diseño de actividades), listas de frecuencias (para la confección de materiales o la enseñanza del léxico), corpus lematizados y analizados (con información sobre categoría gramatical, flexión, relaciones y estructura sintácticas o relaciones semánticas). En general, a propósito de la competencia comunicativa, sirven para confirmar si una unidad léxica, una estructura gramatical, un uso, etc. se realiza realmente en la lengua, y también como fuente de *input* lingüístico.

Por su parte, Villayandre y Llanos (2017: 53-55) recopilan ideas para la explotación de estas herramientas: contextualización de palabras y expresiones, el descubrimiento de patrones en el empleo de prefijos y sufijos, la extracción de frecuencias de palabras, el establecimiento de comparaciones en relación con la variación de registro, el análisis de la combinatoria de palabras. Tras analizar los principales corpus del español, destacan estas características como las más ventajosas para la enseñanza de ELE:

- 1 Ofrecen información estadística sobre frecuencia de uso, algo que ayuda al profesor a establecer prioridades en la enseñanza y seleccionar y gradar, así, los contenidos que son más habituales en la lengua.
- 2 Facilitan datos sobre distribución geográfica, temporal, tipología textual, ámbito del saber, géneros o temas de las unidades léxicas.
- 3 Presentan información sobre colocaciones o coapariciones, lo que propicia el desarrollo de la competencia léxica mediante la detección de patrones de palabras que suelen aparecer juntas.
- 4 Posibilitan comparaciones de formas con significados parecidos para facilitar la detección de diferencias en su empleo y combinatoria.
- 5 Muestran gráficamente relaciones sinonímicas.

Como conclusión, Buyse (2017: 122-123)<sup>3</sup> señala que, para el alumno, el corpus constituye:

---

<sup>3</sup> Este mismo autor (Buyse 2011) presenta diferentes tipos de corpus atendiendo a fines didácticos.

1. Una base sólida para elegir estructuras lingüísticas más frecuentes en las producciones reales de los hablantes nativos de una lengua.
2. Una herramienta que le otorga la autonomía de elegir por sí mismo qué aprender, cómo aprenderlo y en qué orden.
3. Un instrumento para encontrar respuestas a una tipología muy variada de dudas concretas y de profundizar en ellas por medio del acceso a amplios contextos reales.

Mientras que también presenta una ayuda importante para el profesor (nativo o no), ya que le permite:

1. Basarse no exclusivamente en su intuición en ejemplos elaborados *ad hoc*, sino en una fuente amplia y fiable de recursos lingüísticos.
2. Seleccionar un input suficiente y de calidad al que enfrentar a sus alumnos de forma que tenga lugar el *intake*, es decir, la adquisición de cualquier tipo de contenido.

A pesar de todas estas ventajas, en general en el ámbito de ELE la presencia de corpus es escasa, debido principalmente a que son un recurso relativamente reciente en nuestro idioma, y también a que los profesores no suelen poseer formación en lingüística de corpus (Villayandre y Llanos 2017: 58).

Como ejemplo de explotación del uso de corpus lingüísticos en ASL, en los últimos tiempos se está extendiendo el denominado aprendizaje lingüístico impulsado por datos (*data-driven learning*, DDL), término introducido por Johns (1990), y cuyo objetivo principal es potenciar el aprendizaje autónomo e inductivo mediante la reflexión con base en ejemplos reales de la lengua meta. Más concretamente, “el DDL propone la utilización de concordancias extraídas de un corpus de la lengua meta para que los alumnos observen y analicen el comportamiento lingüístico con el fin de que ellos mismos, aunque guiados por el profesor, extraigan generalizaciones sobre la lengua” (Cortina 2009: 109-110)<sup>4</sup>.

#### 4. Metodología

Este estudio parte de la hipótesis de que los corpus lingüísticos pueden emplearse como herramienta para la determinación de la frecuencia de uso de las SF en

---

<sup>4</sup> Otros autores que tratan sobre DDL: Hadley (2002), Boulton y Thomas (2012), Boulton y Pérez-Paredes (2014).

el discurso nativo, y que dicha frecuencia podría ayudar al profesor a establecer el nivel en el que deben enseñarse dichas SF en el aula de ELE.

Sobre esa base, 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 del nivel B2 presentes en el PCIC para comprobar si dicha nivelación se ajusta a la frecuencia de uso entre los hablantes nativos. En el MCER no es hasta el nivel C1 cuando se establece que el alumno, en cuanto a la riqueza del vocabulario, debe poseer un buen dominio de expresiones idiomáticas y coloquiales, así que el nivel B2 puede considerarse la antesala a esa preparación para ese dominio, y es lo que justifica nuestra elección de dicho nivel.

Para ello, pasamos a exponer los criterios de selección tanto de los corpus empleados en el estudio como de las SF del nivel B2 del PCIC.

#### 4.1. Selección de los corpus

En primer lugar, siguiendo a Lindstromberg y Boers (2008) recurrimos al motor de búsqueda de *Google* para analizar la frecuencia relativa de estas expresiones. Para estos autores, una SF cuya busca se sitúe en 3 millones de resultados puede considerarse como frecuente, si arroja un resultado de 230 000 se encuentra en el nivel medio de frecuencia relativa y, por último, si cuenta con 400, es una frecuencia rara<sup>5</sup>.

En cuanto al uso de este buscador como corpus, Buyse (2017: 123-124) comenta sus ventajas y desventajas, destacando entre los inconvenientes el hecho de no estar lematizado, es decir, que si realizamos la búsqueda de “trabajar” solo devuelve formas con el infinitivo, no con el verbo conjugado<sup>6</sup>.

En el ámbito hispánico destaca el *Corpus de referencia del español actual* (CREA) de la Real Academia Española de la Lengua<sup>7</sup>, compuesto por un amplio repertorio de textos de diversa procedencia, tanto escritos (libros, revistas, periódicos) como orales, procedentes en su mayoría de la radio y la televisión, desde 1975 hasta 2004, y que comprende más de 200 millones de palabras. Los principales criterios con los que se ha confeccionado dicho corpus son los cronológicos (1975-2004), geográficos (textos españoles y americanos distribuidos al

<sup>5</sup> 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.

<sup>6</sup> En este mismo trabajo (Buyse 2017: 136) se incluye un cuadro resumen comparativo con las características de varios corpus, entre ellos *Google*, CORPES XXI y *Corpus del español*, que son los que manejaremos en nuestro estudio.

<sup>7</sup> <http://corpus.rae.es/creanet.html>

50%, y en el caso de América, se intenta “reflejar la diversidad existente mediante el reconocimiento de diversas áreas lingüísticas, a cada una de las cuales se ha asignado un porcentaje diferente en función de su población y peso cultural”), medio (textos publicados en libros, revistas, periódicos, que conforman el 90% del total, y transcripción oral, 10%), y temáticos (ciencia, política, vida cotidiana, economía, ficción, etc.)<sup>8</sup>.

La RAE también trabaja actualmente en el *Corpus del Español del Siglo XXI* (CORPES XXI)<sup>9</sup> que, al igual que el CREA, está formado por textos escritos y orales procedentes de España, América, Filipinas y Guinea Ecuatorial comprendidos entre 2001 y 2012, y cuyo objetivo final es reunir, en 2018, un conjunto textual constituido por 400 millones de formas y palabras de la lengua común de casi 500 millones de hispanohablantes<sup>10</sup>. Por tanto, este recurso también permite conocer la distribución del término buscado en función de la zona geográfica, el medio (escrito, oral...), el soporte (Internet, libros...), el tema (actividades, arte...) y obtener datos sobre la frecuencia absoluta y el número de documentos en que aparece la expresión elegida<sup>11</sup>. Igualmente, ofrece la posibilidad de ver los datos estadísticos más importantes, así como las coapariciones (agrupaciones más frecuentes de palabras) y la concordancia de palabras cuando aparecen en proximidad.

Finalmente, otro interesante ejemplo es el *Corpus del español* de Mark Davies<sup>12</sup>, que consta de dos partes: la primera es el corpus original y más pequeño, y 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 dos mil millones de palabras de páginas web de 21 diferentes países de habla hispana, y permite hacer búsquedas en textos en español muy recientes (los textos se recopilaron en 2013 y 2014), por lo que será el que utilizemos para nuestro estudio.

Esta herramienta permite diferentes búsquedas con indicación de la frecuencia de uso, aunque sería amplio describir aquí todas las posibilidades que pre-

---

<sup>8</sup> Puede ampliarse esta información en el manual de consulta del corpus, accesible en [http://corpus.rae.es/ayuda\\_c.htm](http://corpus.rae.es/ayuda_c.htm)

<sup>9</sup> <http://www.rae.es/recursos/banco-de-datos/corpes-xxi>

<sup>10</sup> En junio de 2016 se publicó una nueva versión de consulta, la 0.83. Cuenta con 237 678 documentos, que suman unos 225 millones de formas.

<sup>11</sup> Para más información sobre la conformación y codificación del corpus véase [http://www.rae.es/sites/default/files/CORPES\\_Sistema\\_de\\_codificacion\\_12\\_2015.pdf](http://www.rae.es/sites/default/files/CORPES_Sistema_de_codificacion_12_2015.pdf)

<sup>12</sup> <http://www.corpusdelespanol.org>

senta: permite la búsqueda de palabras exactas, lemas, colocaciones y frases enteras, además de por etiquetas semánticas o categorías gramáticas, todo ello filtrado por períodos y/o géneros textuales. Además, a finales de 2016 se publicó una nueva interfaz que facilita el uso y atiende más a las variantes del español.

#### 4.2. Selección de las SF en el PCIC

Por lo que se refiere al catálogo de SF para nuestro estudio, las hemos extraído de los catálogos de Nociones Específicas del *Plan Curricular del Instituto Cervantes* (PCIC) (2007 [2006]) correspondientes al nivel B2. No obstante, hay que tener en cuenta que, tal y como se indica en el PCIC, sus inventarios deben considerarse abiertos y no exhaustivos “en cuanto que siempre podrían añadirse nuevos elementos, tanto en sentido horizontal (completando las series), como en sentido vertical (con nuevos elementos en cada nivel)” (2007 [2006]: 393-394)<sup>13</sup>. Por tanto, es labor de los programadores, profesores y creadores de materiales adaptar estos inventarios a las necesidades de los estudiantes y a la situación particular de enseñanza, así como añadir nuevos elementos que lo complementen.

En cuanto a la presentación de los contenidos léxicos del PCIC, este documento se asienta sobre la base teórica del enfoque nocio-funcional, en cuanto que presenta las unidades léxicas de acuerdo con la noción o nociones a la que aluden. Asimismo, en el tratamiento del componente nocional se deja ver la huella de autores representativos de los inicios del enfoque léxico como Lewis, Nattinger y De Carrico o Sinclair, que abogan por el principio de idiomatismo y por la segmentación del léxico en bloques semiconstruidos o prefabricados.

Seleccionamos aproximadamente el 30% de las SF del nivel B2 (en total 120)<sup>14</sup>, cantidad que estimamos representativa para realizar un estudio fiable, sin tener en cuenta la clase de unidad (colocación, modismo, expresión institucionalizada...), ya que primamos el criterio comunicativo al teórico, pues al alumno no le sirve esta clasificación teórica para el desarrollo de su competencia comunicativa. En primer lugar, hemos tomado como criterio de selección aquellas SF “semánticamente opacas”, es decir, cuyo significado global no sea la suma de los significados parciales de cada palabra, y que por lo tanto pueden provocar

---

<sup>13</sup> Para conocer más sobre combinatoria léxica en el PCIC, Pérez Serrano (2015: 229).

<sup>14</sup> Remitimos al apartado 4 (Metodología) para el criterio de selección del nivel B2. Por otro lado, hemos seleccionado unas 80 estructuras oracionales, pero el número de expresiones analizadas se amplía hasta las 120 porque debemos tener en cuenta que varias de dichas estructuras son del tipo verbo + sustantivo (“dar una patada/una torta/un puñetazo”) y similares. Presentamos el listado completo en el Anexo I.

problemas para los estudiantes de ELE, aunque también hemos seleccionado otras SF para completar el 30%. Además, como señala el PCIC “en todo caso, se ha procurado que los ejemplos fueran representativos, frecuentes y naturales.”

## 5. Análisis

### 5.1. Resultados de las SF en *Google*

En el primero de nuestros análisis, tomaremos como base los datos extraídos del buscador *Google*, y clasificaremos dichos datos, como ya hemos adelantado, según el criterio establecido por Lindstromberg y Boers (2008):

- A. 3 millones de resultados: frecuente.
- B. 230 000 resultados: nivel medio de frecuencia.
- C. 400 resultados: frecuencia rara.

De este modo, de mayor a menor, podemos agrupar las expresiones en las tablas que mostramos a continuación.

**Tabla 1.** Grupo A (3 millones de resultados)

Ir/salir/estar de fiesta (19 100 000) <sup>15</sup>	Dar de alta (la luz/el agua/el teléfono/el gas) (4 880 000)
Dar a (la calle/a un patio/al jardín) (8 540 000)	(Tener) salud de hierro (4 590 000)
(Estar) pasado de moda (8 180 000)	Dar de baja (la luz/el agua/el teléfono/el gas) (4 480 000)
Dar la vez (7 420 000)	Acceder a/pinchar en un enlace (4 050 000)
Tratar de tú (7 250 000)	Ser el centro de la fiesta (3 930 000)
Dar a luz (6 870 000)	Agotar los recursos naturales (3 730 000)
(Pagar) al contado (5 390 000)	Poner en escena (una obra teatral) (3 160 000)

<sup>15</sup> En estos casos de construcciones con distintas variedades, la cantidad de resultados que presentamos representan la suma de todas las variedades de estas construcciones.

En la Tabla 1, apreciamos que solo 14 expresiones se encuentran en este grupo, y que los resultados oscilan entre los 19 100 000 de “Ir/salir/estar de fiesta” y los 3 160 00 de “poner en escena (una obra teatral)”. Esto significa que estas SF presentan un alto índice de frecuencia, por lo que podrían tratarse antes del nivel B2. Además, es posible que los aprendientes las adquieran fuera de clase al escucharlas o leerlas frecuentemente, por lo que tal vez no sea necesario centrar la atención sobre ellas de forma tan específica en el aula. Esto confirma que la selección realizada por el PCIC, a pesar de basarse en un criterio intuitivo, es bastante fiable. No obstante, este método que proponemos puede servir al profesorado para realizar una secuenciación más precisa en la selección de las SF que trata en clase.

**Tabla 2.** Grupo B (230000 resultados)

Dar(se) de sí (2 880 000)	Arreglar la casa (536 000)	Tener buenos modales (57 700)
Pedir la vez (2 420 000)	Estar en buenas condiciones (alimentos) (479 000)	Hacer la digestión (55 200)
Hacerse una idea (2 390 000)	No tener dos dedos de frente (438 000)	Tener un ligue (44 800)
Ascender a +cargo profesional (2 370 000)	Dos por (el precio de) uno (394 000)	Estar retirado/a (38 400)
Ser objeto de estudio (2 210 000)	Tener un romance (370 000)	Pasar un control de carretera (37 000)
Dos por (el precio de) uno (2 130 000)	Entregar en mano (352 000)	Citar textualmente (33 500)
Pasar el rato (2 130 000)	Firmar la paz (343 000)	Estar como una cabra (31 100)
Ejercer de + profesión (2 030 000)	Cocer al vapor/a fuego lento/al baño maría (328 000)	Aparcar en zona azul (29 900)
(Pagar) a plazos (2 020 000)	Pasar(se)lo en grande (327 000)	Cortarse la comunicación (25 800)
Estar en obras (2 010 000)	Ceder el paso (326 000)	Sacar malas notas (25 600)
Limpia en seco (1 820 000)	Hacer doble clic (286 000)	Tener un sueño horrible (25 400)

(Hacer) doble clic (1 760 000)	Jugar al escondite/a las canicas/a la comba (267 200)	Quemarse la espalda (23 100)
Enterarse de una noticia casualmente (1 760 000)	Estar de oferta (257 000)	Estar castigado/a (22 500)
Preservar el medio ambiente (1 730 000)	Estar en malas condiciones (alimentos) (256 000)	Comer de menú/a la carta (22 000)
Sacar conclusiones (1 720 000)	Estar en huelga (255 000)	Llamar por el apellido (21 300)
Estar de baja (1 610 000)	Montar un viaje (242 000)	Tener un dolor horrible (19 300)
Perder el empleo (1 350 000)	Ser idénticos/as (230 000)	Enterarse de una noticia por (casualidad/la prensa) (17 800)
Ejercer como + profesión (1 340 000)	(Tener una salud) a prueba de bomba (225 000)	Enviar contra reembolso (17 000)
Hacer campaña (1 150 000)	Pagar a plazos (214 000)	Domiciliar un recibo (14 900)
Tener un lío (1 050 000)	Sacar buenas notas (210 000)	Pedir el libro de reclamaciones (14 400)
Dar un golpe de estado (991 000)	Dar un puñetazo (209 000)	Llamar a cobro revertido (14 200)
Ser despedido (880 000)	Salir gratis (199 000)	Ser clavados/as (14 000)
Dar una patada (846 000)	Llamar por el nombre (190 000)	Tener complejo de inferioridad/de superioridad (12 100)
Recoger la casa (812 000)	Pagar al contado (171 000)	(Agotarse) los recursos naturales (11 100)
Tratar de usted (768 000)	Dar una torta (120 000)	Pasar(se)lo fenomenal (9580)
Ir/salir/estar de juerga (737 000)	Ponerse guapo/a (118 700)	Minimizar una ventana (7080)
Cortar la línea (662 000)	Dejar una señal (115 000)	Tener malos modales (5610)

Declarar la guerra (653 000)	Hacer penitencia (112 000)	Maximizar una ventana (5420)
Tener un/a amante (641 000)	Resolver una duda (104 000)	Quemarse los hombros (3200)
Comer a la carta (636 000)	Tener vistas (103 000)	Estar acomplejado/a (2280)
Quitar el polvo (636 000)	Estar desempleado/a (96 100)	Cortarse la digestión (1850)
Reponer fuerzas (612 000)	Estar de baja por (paternidad/maternidad/por enfermedad) (79 700)	Domiciliar un pago (1710)
Quedarse en blanco (612 000)	Plantear/resolver una duda (59 700)	Agotarse un libro (1320)
Ser como dos gotas de agua (540 000)	Pinchar en un enlace (58 000)	

En este segundo grupo (Tabla 2), los resultados varían desde “dar(se) de sí” 2 880 000 a “agotarse un libro” (1320). Como hemos comentado para los resultados de la tabla anterior, el criterio seguido para su selección en el PCIC parece acertado. Ya hemos mencionado que las SF con este índice de frecuencia son muy habituales, por lo que es recomendable centrar en ellas nuestra labor docente. Es por esto que la amplia diferencia en el número de resultados en este grupo parece recomendar una subdivisión, siempre en función de la programación establecida por el profesorado, así como por las necesidades de los alumnos.

Finalmente, no hemos encontrado ninguna SF en el tercer grupo (400 resultados), lo que indica que ninguna de ellas presenta un nivel bajo de frecuencia relativa y, por tanto, sea aconsejable tratarlo en un nivel más alto.

## 5.2. Resultados en los corpus

Una vez establecido el índice general de frecuencia de las SF en *Google*, nuestra intención ahora es comparar dichos resultados con los obtenidos de los tres corpus que hemos seleccionado para nuestro estudio (CREA, CORPES XXI y *Corpus del español*). No obstante, debido a que no sería productivo manejar los resultados de todas las SF, hemos seleccionado las diez primeras más frecuentes (Tabla 3).

**Tabla 3.** Resultados más frecuentes en los corpus (comparados con *Google*)

GOOGLE	CREA	CORPES XXI	<i>Corpus del español</i>
Ir/salir/estar de fiesta (19 100 000)	Ir/salir/estar de fiesta (1096)	Ir/salir/estar de fiesta (1602)	Dar a luz (16 536)
Dar a (la calle/a un patio/al jardín) (8 540 000)	(Estar) de oferta (432)	(Estar) de oferta (659)	Sacar conclusiones (4545)
(Estar) pasado de moda (8 180 000)	Declarar la guerra (353)	Dar a luz (595)	Dar de baja (la luz/el agua/el teléfono/el gas) (3559)
Dar la vez (7 420 000)	Dar a luz (317)	Poner en escena (una obra teatral) (526)	Pasar el rato (2511)
Tratar de tú (7 250 000)	(Pagar/pago/s) a plazos (280)	Declarar la guerra (286)	Dar de alta (la luz/el agua/el teléfono/el gas) (2467)
Dar a luz (6 870 000)	(Pagar/pago/s) al contado (250)	(Pagar/pago/s) al contado (233)	Poner en escena (una obra teatral) (2339)
(Pagar) al contado (5 390 000)	Sacar conclusiones (180)	Sacar conclusiones (225)	Ir/salir/estar de fiesta (2022)
Dar de alta (la luz/el agua/el teléfono/el gas) (4 880 000)	Hacerse una idea (163)	Hacerse una idea (212)	Estar desempleado (1902)
(Tener) salud de hierro (4 590 000)	Ejercer de + profesión (160)	Ejercer de + profesión (204)	Ser despedido (1889)
Dar de baja (la luz/el agua/el teléfono/el gas) (4 480 000)	Firmar la paz (158)	(Pagar/pago/s) a plazos (178)	Ejercer como + profesión (1574)

Como podemos apreciar, en primer lugar destacan las diferencias cuantitativas en el número de resultados, que se cuentan por millones en *Google*, mientras que en los corpus el más abundante es el *Corpus del español*, seguido por este orden por el CORPES XXI y el CREA. La explicación es evidente, ya que mientras el buscador toma como base de búsqueda la inmensidad de Internet, los corpus, más o menos amplios, reducen dicha base en función de los criterios establecidos para su construcción. De este modo, los corpus siempre representan una selección con base en la modalidad, géneros, ámbitos temporales, temáticos, geográficos, etc. seleccionados.

A pesar de esto, sí que existe una notable coincidencia en las SF más habituales en estos cuatro recursos, puesto que un alto número de expresiones aparecen en los cuatro, o al menos en tres de ellos: “Ir/salir/estar de fiesta”, “dar a luz”, “pagar al contado”, “sacar conclusiones”, “ejercer como/de + profe-

sión”)<sup>16</sup>. Por otro lado, es evidente y comprensible la casi total coincidencia en los dos corpus académicos (en nueve de las diez SF) debido a que parten de bases documentales similares.

Lo dicho parece confirmar que, aunque *Google* sea la herramienta más adecuada para establecer la frecuencia de las SF por su amplitud y por el número de resultados arrojados, estos tres corpus también pueden tomarse como referencia para determinar dicha frecuencia. Además, como mostramos a continuación, estos corpus sí que son muy útiles para extraer información muy significativa sobre el uso real de las expresiones, información que los profesores de ELE pueden aprovechar para su enseñanza y aprendizaje.

### 5.3. Información de uso de las SF en los corpus

Aunque no sea el objetivo principal de nuestro estudio, no podemos dejar de mencionar una de las principales utilidades de los corpus para la enseñanza y aprendizaje de las SF en ELE/L2: nos referimos a la diversa tipología de información sobre el uso de estas expresiones de la que podemos disponer mediante el empleo de estas herramientas.

Año	%	Casos	País	%	Casos	Tema	%	Casos
1997	10.15	92	ESPAÑA	54.35	587	7.- Ficción.	31.75	348
1995	9.16	83	MÉXICO	8.42	91	5.- Ocio, vida cotidiana.	21.89	240
1996	7.94	72	ARGENTINA	7.87	85	2.- Ciencias sociales, creencias y pensamiento.	15.32	168
2001	6.95	63	VENEZUELA	5.18	56	4.- Artes.	12.59	138
2002	6.29	57	COLOMBIA	4.90	53	3.- Política, economía, comercio y finanzas.	8.66	95
1994	6.18	56	PERÚ	4.07	44	9.- Oral.	6.56	72
2000	5.62	51	CHILE	3.42	37	8.- Miscelánea.	1.36	15
1989	4.96	45	CUBA	3.14	34	1.- Ciencia y Tecnología.	1.09	12
1987	4.52	41	PUERTO RICO	1.57	17	6.- Salud.	0.72	8
Otros	38.18	346	Otros	7.03	76			

Imagen 1. CREA “de fiesta”. Estadística

Así, podemos conocer la distribución temporal de la SF (lo que nos indicaría, por ejemplo, su vigencia de uso, es decir, si se trata de una expresión actual o que ha dejado o está dejando de emplearse hoy día), así como la zona geográfica en la que se emplea, con la valiosa información que esto representa para deter-

<sup>16</sup> Esta coincidencia es incluso más amplia, ya que algunas de las expresiones que no coinciden en este rango, sí que lo hacen si lo ampliamos hasta las quince o veinte SF más frecuentes en cada recurso.

minar un uso más general en el ámbito hispánico, o bien un empleo más restringido a ciertas zonas hispanohablantes (Imagen 1 y 2).



Imagen 2. CORPES XXI “como una cabra”. Estadística

Además, ofrecen la posibilidad de precisar el medio (escrito, oral...), el soporte (Internet, libros...), el tema (actividades, arte...) y obtener datos sobre la frecuencia absoluta y el número de documentos en que aparece la expresión, con lo cual, por ejemplo, se puede apreciar si la SF se encuentra presente en la modalidad escrita (libros, periódicos...) y/o en la oral (entrevistas, radio,...), aunque debemos precisar que la modalidad oral está mucho menos representada que la escrita en estos corpus (Imagen 1, 2 y 3).

Pantalla: 1 de 1.

Párrafo nº 1.

Es que a unos resulta y a otros no. Los escándalos sociales tienen a veces resultados contradictorios. La princesa Diana le ganó el pleito a su marido, quien tuvo que darle más de veinte millones de dólares, después de decirle al mundo (en todos los idiomas) que ella le había sido infiel con cuanto amante le dio la gana (y parece que las ganas eran muchas). Después de eso se ha convertido en una atracción mundial que eclipsa hasta la misma Reina de Inglaterra. Y su concubina, la esposa del príncipe Andrés, Sarah Ferguson, ya no sabe qué hacer con el dinero que le llega por la publicación de sus escándalos sexuales.

El deportista O.J. Simpson, después de asesinar a su ex mujer y a un compañero de ésta, se convirtió en un ídolo de las multitudes. Al parecer, ya empiezan a notar la estupidez, y le han retirado su admiración.

Sí, estamos por creer lo que decía Jardiel Poncela: La Humanidad está **como una cabra**.

OTRO CORRUPTO QUE VUELVE.

El ex presidente venezolano Carlos Andrés Pérez, destituido de su segunda presidencia en 1993 y convicto por corrupción, se lanzó de nuevo a la arena política. (Volverá a hablar de decencia y corrección administrativa?)

Es interesante saber cómo lo recibirá el pueblo venezolano...

LA PREGUNTA DEL DÍA.

¿Cómo puede esperar que su vida tenga una dirección si no sabe adónde va?

-----

AÑO: 1997  
AUTOR: PRENSA  
TÍTULO: El Siglo, 27/03/1997 :  
PAÍS: PANAMÁ  
TEMA: 02\_Testimonios varios  
PUBLICACIÓN: (Panamá), 1997

Ir arriba Pantalla: 1 de 1.

Imagen 3. CREA “como una cabra”. Párrafos

Igualmente, estos corpus nos permiten disponer de muestras reales en las que se emplean las SF, ya que podemos recuperar los textos en los que aparecen insertas (Imagen 3). En este sentido, es importante destacar la posibilidad de consul-

tar las palabras que aparecen cerca de las expresiones (“colocaciones” o “agrupaciones”), lo que proporciona información muy relevante sobre su significado y su uso. Por ejemplo, podemos consultar con qué otros nombres o verbos se combinan las expresiones (Imagen 4).



Imagen 4. *Corpus del español* “de fiesta”. Colocaciones

Todo este volumen de información posibilita el “aprendizaje lingüístico impulsado por datos” (*data-driven learning* o DDL), potenciando el aprendizaje autónomo e inductivo de los estudiantes precisamente mediante la reflexión sobre ejemplos reales de la lengua. Estos recursos contribuirían, así, a la práctica de aspectos pragmáticos, discursivos y culturales mediante la exposición de los aprendientes a muestras lingüísticas reales.

## 6. Conclusiones

Es innegable la importancia del componente léxico en el proceso de enseñanza y aprendizaje de ELE/L2, tanto que hoy día se considera el eje central de dicho proceso, y a partir del cual se desarrolla la competencia comunicativa del estudiante. De entre las distintas unidades que lo forman, destacan las SF por su gran número en el discurso nativo, y esto, por tanto, las dota de una importante utilidad comunicativa, puesto que su conocimiento incrementa la fluidez en la producción y la capacidad de comprensión de los alumnos.

En la actualidad, el empleo de medios técnicos como las TIC y TAC permite disponer de un importante volumen de información sobre estas expresiones, como su frecuencia, aspectos de uso y discursivos como el ámbito geográfico y

temático, la modalidad (oral y/o escrita) en la son más empleados, su registro, así como la combinatoria léxica en la que participan.

En el caso del ELE/L2, encontramos una amplia selección de estas SF en el PCIC, agrupadas en torno a las nociones específicas que presenta esta obra. La selección realizada en este documento nos parece a grandes líneas correcta, pero nuestra propuesta sirve, como hemos podido comprobar en el estudio realizado, en primer lugar para precisar y afinar la frecuencia de uso de estas SF, y que puede servir de orientación al profesor, en función del currículo y de las necesidades del alumnado. En segundo lugar, y aunque solo lo hemos esbozado, estas herramientas ofrecen un amplio e importante volumen de información de uso de estas SF muy útil para su enseñanza y aprendizaje.

En este sentido, podemos imaginar *Google* como un océano de información, mientras que los corpus serían las redes que nos permiten recoger un volumen más concreto de datos según unos criterios establecidos por los compiladores (modalidad, género, temática, ámbito geográfico y temporal...). Por tanto, nuestra propuesta es emplear *Google* como índice de frecuencia de las SF, lo que guiaría su enseñanza en un nivel más o menos determinado, mientras que los corpus podrían afinar dicha selección y, sobre todo, poner a disposición del profesor un amplio abanico de datos sobre el uso real de las expresiones.

Tomando como base todo lo expuesto en el trabajo, nos proponemos realizar un estudio similar a este en todos los niveles del PCIC, así como elaborar un modelo de propuesta didáctica de estas SF, tomando como base el DDL, que pueda servir como guía para la enseñanza y aprendizaje de estas expresiones en ELE/L2.

## Referencias bibliográficas

- Adell Segura, J, y Castañeda Quintero, L. 2012. Tecnologías emergentes, ¿pedagogías emergentes? En J. Hernández *et al.* (eds), *Tendencias emergentes en educación con TIC*, Barcelona: Espiral, 13-32.
- 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*. Salamanca: ASELE, 83-95. <https://goo.gl/5oEXpR> (consultado el 7 de abril de 2018).
- Boers, F., & Lindstromberg, S. 2009. *Optimizing the Lexical Approach to Instructed Second Language Acquisition*. Nueva York: PalgraveMacmillan.
- Boulton, A. & Thomas, J. 2012. Corpus language input, corpus processes in learning, learner corpus product. En J. Thomas & A. Boulton (eds), *Input, Process and Product: Developments in Teaching and Language Corpora*. Brno: Masaryk University Press, 7-34.

- Boulton, A. & Pérez-Paredes, P. 2014. Researching uses of corpora for language teaching and learning. *ReCALL* 26: 121-127.
- Briz Gómez, A. y Albelda Marco, M. 2009. Estado actual de los corpus de lengua española hablada y escrita: I+D. *El español en el mundo. Anuario 2009*. Madrid: Instituto Cervantes. <https://bit.ly/2txdcBo> (consultado el 16 de abril de 2018).
- Buyse, K. 2011. ¿Qué corpus en línea utilizar para qué fines en la clase de ELE?. En J. De Santiago Cuervós *et al.* (eds), *Del texto a la lengua: la aplicación de los textos a la enseñanza-aprendizaje del español L2-LE*. Salamanca: ASELE, 277-289. <https://bit.ly/2ttkEx6> (consultado el 7 de abril de 2018).
- 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*. Barcelona: Difusión, 121-141.
- Consejo de Europa. 2002. *El Marco común europeo de referencia para las lenguas aprendizaje, enseñanza, evaluación*. Madrid: Anaya y CVC. <https://bit.ly/2IkNKo4> (consultado el 22 de febrero de 2018).
- Contreras Izquierdo, N. 2008. La enseñanza-aprendizaje de lenguas extranjeras y las TICs: el caso del Español como Lengua Extranjera (ELE). *Iniciación a la investigación*. Universidad de Jaén.
- Cortina Pérez, B. 2009. Una propuesta didáctica para la mejora de la competencia conversacional de una LE mediante secuencias formulaicas. *Didáctica. Lengua y Literatura* 21: 99-115.
- 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. Revista de Didáctica ELE*. <https://bit.ly/2Kc8901> (consultado el 3 de marzo de 2018).
- Cruz Piñol, M. 2017. *Lingüística de corpus y enseñanza de español como 2/L*. Madrid: Arco Libros.
- Erman, B. & Warren, B. 2000. The idiom principle and the open-choice principle. *Text* 20: 87-120.
- Gómez Molina, J. R. 2004. Los contenidos léxico-semánticos. En J. Sánchez Lobato e I. Santos Gargallo (eds), *Vademécum para la formación de profesores. Enseñar español como segunda lengua (L2)/lengua extranjera (LE)*. Madrid: SGEL, 789-810.
- Hadley, G. 2002. An introduction to data-driven learning. *RELC Journal* 33(2): 99-124.
- Hernández, M. P. 2012. La integración de las TIC en la clase de ELE. Panorama de una (r)evolución. *Revista Internacional de Lenguas Extranjeras* 1: 63-99. <https://bit.ly/2yBCTq3> (consultado el 5 de febrero de 2018).
- Hernández, M. P. 2014. Recursos para la integración de las TIC en el aula de ELE: las cazas del tesoro. En *XV Encuentro Práctico de Profesores de ELE*. International House: Difusión. <https://bit.ly/2K1EF9b> (consultado el 13 de abril de 2018).
- Herrera, F. y Conejo, E. 2009. Tareas 2.0: la dimensión digital en el aula de español lengua extranjera. *MarcoELE* 9: 1-20. <https://bit.ly/2yCGHHv> (consultado el 16 de abril de 2018).
- Higueras, M. 2004. Internet en la enseñanza del español. En J. Sánchez Lobato e I. Santos Gargallo (eds), *Vademécum para la formación de profesores. Enseñar español como segunda lengua (L2)/lengua extranjera (LE)*. Madrid: SGEL, 1067-1077.
- Instituto Cervantes. 2006. *Plan Curricular del Instituto Cervantes. Niveles de referencia para el español*. <https://bit.ly/2IWhigd> (consultado el 27 de febrero de 2018).

- Instituto Cervantes. 2012. *Competencias clave del profesorado de lenguas extranjeras*. <https://bit.ly/2tu4117> (consultado el 7 de febrero de 2018).
- Jiménez, D., Mora, M. y Cuadros, R. 2016. La importancia de las nuevas tecnologías en el proceso educativo. Propuesta didáctica TIC para ELE: Melendien7días. *Fuentes18* (2) <https://bit.ly/2lrsfZI> (consultado el 18 de febrero de 2018).
- Johns, T. 1990. From printout to handout: Grammar and vocabulary teaching in the context of data-driven learning. *CALL Austria* 10: 14-34.
- Johnson, L., Adams Becker, S., Estrada, V., & Freeman, A. 2015. *NMC Horizon Report: Edición Educación Superior 2015*. Austin, Texas: The New Media Consortium. <https://bit.ly/1MdNICM>. Versión en español disponible en <https://bit.ly/1J9oQQS> (consultado el 3 de febrero de 2018).
- Juan, O. 2009. Web 2.0, comunicación y material didáctico digital para el aprendizaje del español: El Aula Virtual de Español el Instituto Cervantes y su actualización. *RLA. Revista de Lingüística Teórica y Aplicada* 47(2): 13-34. <https://bit.ly/2MjhRyu> (consultado el 5 de febrero de 2018).
- Juan, O. 2016. La tecnología en el aula y fuera del aula: actitudes y valoraciones del profesor en la integración del Aula Virtual de Español, AVE. Hacia propuestas de mejora continua y renovación. *RedELE* 28. <https://bit.ly/2ywSWW6> (consultado el 5 de febrero de 2018).
- 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*. Londres: Helbling Languages.
- Llisterri, J. 1997. Nuevas tecnologías y enseñanza del español como lengua extranjera. En F. Moreno, M. Gil y K. Alonso (eds), *El español como lengua extranjera: del pasado al futuro*. Alcalá de Henares: Servicio de Publicaciones de la Universidad de Alcalá-ASELE, 45-75.
- 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*. Argel: Instituto Cervantes, 57-64. <https://goo.gl/PZcsdB> (consultado el 5 de mayo de 2018).
- Nicolás Martínez, C. 2012. *Corpus C-Or-DiAL (Corpus Oral Didáctico Anotado Lingüísticamente)*. Madrid: Liceus.
- Pastor, S. 2004. *Aprendizaje de segundas lenguas. Lingüística aplicada a la enseñanza de idiomas*. Alicante: Publicaciones Universidad de Alicante.
- 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*. Madrid: Arco/Libros.
- Pérez Tornero y Pi, M. 2014. *Perspectivas 2014: Tecnología y pedagogía en las aulas. El futuro inmediato en España*. Universidad Autónoma de Barcelona: aula Planeta. <https://bit.ly/1vrSLmo> (consultado el 11 de febrero de 2018).
- REAL ACADEMIA ESPAÑOLA. *Corpus de referencia del español actual (CREA)* <http://www.rae.es/recursos/banco-de-datos/crea>

- REAL ACADEMIA ESPAÑOLA. *Corpus del Español del Siglo XXI (CORPES)*  
<http://www.rae.es/recursos/banco-de-datos/corpes-xxi>
- Rojo, G. 2016. Los corpus textuales del español. En J. Gutiérrez-Rexach (ed.), *Enciclopedia lingüística hispánica*. London: Routledge, 285-296.
- Román-Mendoza, E. 2018: *Aprender a aprender en la era digital. Tecnopedagogía crítica para la enseñanza del español LE/L2*. New York: Routledge.
- 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. En D. G. Níkleva (ed), *La formación del profesorado de español como lengua extranjera. Necesidades y tendencias*. Bern: Peter Lang, 257-282.
- Simons, M. 2010. Perspectiva didáctica sobre el uso de las TIC en la clase de ELE. *MarcoELE* 11. <https://bit.ly/1SVFZsB> (consultado el 7 de mayo de 2018).
- Sitman, R. 1998. Divagaciones de una internauta. Algunas reflexiones sobre el uso y abuso de la Internet en la enseñanza del E/LE. *Espéculo. Revista de estudios literarios* 10. <https://bit.ly/2to9ISu> (consultado el 8 de marzo de 2018).
- Soria, I. y Luzón, J. M. 1999. El enfoque comunicativo en la enseñanza de lenguas. Un desafío para los sistemas de enseñanza y aprendizaje abiertos y a distancia. En D. Gallego y C. Alonso (eds), *Informática Educativa. Congreso Internacional de Informática Educativa*. Madrid: UNED, 405-432.
- Villayandre Llamazares, M. y Llanos Casado, L. 2017. Los corpus electrónicos en la clase de español: reflexiones y aplicaciones. En B. Peña y A. M<sup>a</sup>. Aguilar (coords), *Didáctica de la lengua y la literatura. Buenas prácticas docentes*. Tomo I. Madrid: ACCI, 52-79.
- Wray, A. 2002. *Formulaic Language and the Lexicon*. Cambridge: Cambridge University Press.

## Anexo

### Unidades pluriverbales (SF) del nivel B2 en el PCIC

Nociones específicas	Unidades pluriverbales
<b>Tabla 1: Individuo: dimensión física</b>	
Partes del cuerpo	
Características físicas	ser ~ clavados/idénticos/como dos gotas de agua
Acciones y posiciones que se realizan con el cuerpo	dar ~ una patada/una torta/un puñetazo
Ciclo de la vida y reproducción	dar a luz
<b>Tabla 2: Individuo: dimensión perceptiva y anímica</b>	
Carácter y personalidad	estar acomplejado, tener complejo ~ de inferioridad/de superioridad no tener dos dedos de frente estar como una cabra
Sensaciones y percepciones físicas	tener ~ un dolor/un sueño ~ horrible/espantoso
Modales y comportamiento	tener ~ buenos/malos ~ modales
<b>Tabla 3: Identidad personal</b>	
Datos personales	llamar ~ por el nombre/por el apellido tratar ~ de tú (tutear)/de usted
<b>Tabla 4: Relaciones personales</b>	
Relaciones sociales	tener ~ un lío/un romance/un amante/un ligue
Celebraciones y actos familiares, sociales y religiosos	ser el centro de la fiesta
<b>Tabla 5: Alimentación</b>	
Dieta y nutrición	estar ~ en buenas/en malas ~ condiciones hacer/cortarse ~ la digestión
Recetas	cocer ~ al vapor/a fuego lento/al baño María
Restaurante	comer ~ de menú/a la carta
<b>Tabla 6: Educación</b>	
Aprendizaje y enseñanza	sacar conclusiones, quedarse en blanco
Exámenes y calificaciones	sacar ~ buenas notas/malas notas
Lenguaje de aula	plantear/resolver ~ una duda estar castigado
<b>Tabla 7: Trabajo</b>	
Profesiones y cargos	ejercer ~ de/como ~ abogado/veterinario
Actividad laboral	ascender a ~ subdirector/jefe de sección/responsable de unidad
Desempleo y búsqueda de trabajo	perder el empleo, ser despedido, estar desempleado
Derechos y obligaciones laborales	estar de baja ~ por paternidad/por enfermedad estar en huelga estar retirado
<b>Tabla 8: Ocio</b>	

Tiempo libre y entretenimiento	pasar el rato ir/salir/estar ~ de juerga/de fiesta
Juegos	pasar(se)lo ~ fenomenal/en grande jugar ~ al escondite/a las canicas/a la comba
<b>Tabla 9: Información y medios de comunicación</b>	
Información y comunicación	enterarse de una noticia ~ por casualidad/casualmente/por la prensa
Correspondencia escrita	entregar en mano
Teléfono	cortar la línea, cortarse la comunicación llamar a cobro revertido
Internet	dar de ~ alta/baja ~ la línea (telefónica) entrar en/acceder a/pinchar en ~ un enlace
<b>Tabla 10: Vivienda</b>	
Acciones relacionadas con la vivienda	dar ~ de alta/de baja ~ la luz/el agua/el teléfono/el gas
Características de la vivienda	dar ~ a la calle/a un patio/al jardín
Actividades domésticas	arreglar/recoger ~ la casa quitar el polvo
<b>Tabla 11: Servicios</b>	
Servicio postal	enviar contra reembolso
Servicios financieros	domiciliar un pago / un recibo
<b>Tabla 12: Compras, tiendas y establecimientos</b>	
Lugares, personas y actividades	dar/pedir la vez hacerse una idea
Ropa, calzado y complementos	dar(se) de sí limpiar en seco estar pasado de moda
Pagos	dejar una señal salir gratis dos por (el precio de) uno
<b>Tabla 13: Salud e higiene</b>	
Salud y enfermedades	reponer fuerzas tener una salud de hierro tener una salud a prueba de bomba
Heridas y traumatismos	dar puntos
Estética	ponerse guapo
<b>Tabla 14: Viajes, alojamiento y transporte</b>	
Viajes	montar un viaje quemarse la espalda/los hombros
Alojamiento	tener vistas pedir el libro de reclamaciones
Sistema de transporte	pasar un control de carretera aparcar en zona azul ceder el paso
<b>Tabla 15: Economía e industria</b>	
Finanzas y bolsa	pagar a plazos pagar al contado

Comercio	estar de oferta
Entidades y empresas	estar de baja
<b>Tabla 16: Ciencia y tecnología</b>	
Cuestiones generales	ser objeto de estudio
Informática y nuevas tecnologías	hacer doble clic
	maximizar/minimizar una ventana
<b>Tabla 17: Gobierno, política y sociedad</b>	
Política y gobierno	hacer campaña
	dar un golpe de Estado
Ejército	declarar la guerra
	firmar la paz
<b>Tabla 18: Actividades artísticas</b>	
Literatura	agotarse un libro
	citar textualmente
Cine y teatro	poner en escena (una obra teatral)
<b>Tabla 19: Religión y filosofía</b>	
Religión	hacer penitencia
<b>Tabla 20: Geografía y naturaleza</b>	
Espacios urbanos o rústicos	estar en obras
Problemas medioambientales y desastres naturales	agotarse los recursos naturales
	preservar el medio ambiente