

## Combinatoria léxica y corpus como *input*

ANNA SÁNCHEZ RUFAT  
*Universidad de Córdoba*  
FRANCISCO JIMÉNEZ CALDERÓN  
*Universidad de Extremadura*

**Abstract:** *Computer based corpora which have been statistically analyzed have lead to the recognition of frequent lexical sequences (also known as multiword expressions, prefabricated chunks or collocations –in a loose use of the term– as a very important component of language production. Spontaneous interactions demand the storage of the phrasal element during the encoding and decoding of the message. Corpus studies, apart from improving the descriptions of lexis contribute, by means of the Applied Cognitive Linguistics proposals, to the learning process of these units with activities conveniently designed. This article brings an approach which combines the applications of corpora with the benefits of collocations –in a restricted use of the term– when the lexical competence is being developed. The aim is to provide the ability to identify collocations by using the input contained in the corpora, which it will lead to the production of many different combinations, once the connections between a predicate and its arguments have been established.*

**Key words:** *lexical combinations, collocation, vocabulary acquisition, textual corpus, Applied Cognitive Linguistics.*

### 0. Introducción

El presente estudio trata de conjugar el creciente interés en la parcela lingüística del léxico, tanto desde el punto de vista de la investigación como desde la perspectiva del aprendizaje, con la aplicación, también en ambas vertientes, de las herramientas que suponen los corpus textuales.

Por lo que respecta al léxico, revisten especial importancia para los propósitos de estas líneas sus propiedades combinatorias, que suelen dar lugar a

*Language Design* 14 (2013): 61-81)

conflictos terminológicos y conceptuales. Según se comprobará más abajo, resulta de gran utilidad contar con una definición bien delimitada del concepto de *colocación*, ya que esta permite, en primer lugar, aclarar la clasificación de las diferentes combinaciones léxicas y, en segundo lugar, diseñar estrategias verdaderamente efectivas para el aprendizaje del léxico.

En lo que se refiere a los corpus textuales, su análisis ha desembocado en una de las líneas de investigación que se desarrollan al calor de la revalorización del componente léxico. Pero, además de las vías abiertas por esta línea de estudio, los corpus textuales pueden utilizarse en el proceso de aprendizaje de una unidad léxica, en las fases de retención, fijación y reutilización. Es decir, el objeto de estudio puede convertirse en herramienta para el aprendizaje. Se trata, por tanto, de concretar un determinado tipo de combinación léxica que resulte operativa (la *colocación*) y, posteriormente, diseñar estrategias para su aprendizaje utilizando las posibilidades que los corpus ofrecen. Si bien semejante objetivo desborda los límites de un solo trabajo, trataremos a continuación de abrir un camino en esa dirección, sugiriendo una propuesta de aprendizaje de las colocaciones mediante la aplicación al proceso de los corpus textuales.

### **1. A propósito de las combinaciones léxicas**

Cuando se analizan las redes léxicas en torno a una palabra en el eje sintagmático, esto es, con qué palabras o tipos de palabras suele o puede coaparecer, comprobamos que tanto en la literatura como en los manuales de ELE las coapariciones son agrupadas –pese a su heterogeneidad– en un mismo concepto, al que se alude con diferentes denominaciones. Así, la coaparición o *colocación* –en sentido amplio– es también *expresión multipalabra*, *bloque semiprefabricado* o *secuencia formulaica*. Entre dichas unidades se hallan marcadores del discurso (*por otra parte*), locuciones (*darse cuenta*), símiles estandarizados (*fuerte como un roble*), proverbios (*mientras hay vida, hay esperanza*), exclamaciones (*¡anda ya!*), fórmulas sociales (*¡que cumplas muchos más!*), citas (*pienso, luego existo*), esquemas fijos abiertos

(*cuanto más* [presente subjuntivo], *más* [futuro]), combinaciones libres<sup>1</sup> (*reservar habitación*), colocación en sentido estricto (*derrumbar* + nombres que designan objetivos o esperanzas, como *planes, proyectos, ilusiones...*), etcétera. Todas estas combinaciones léxicas tienen en común la coaparición frecuente de sus constituyentes, por lo que su almacenamiento en el léxico mental tiende a ser holístico en los miembros de una misma comunidad de hablantes. Con la implantación de las nuevas teorías lingüísticas y de adquisición fundamentadas en la primacía del léxico y en el principio de idiomática, surgen enfoques didácticos<sup>2</sup> que promueven la enseñanza de estos bloques semiconstruidos (Nattinger y DeCarrico, 1992; Lewis, 1993, 1997 y 2000), los cuales, una vez reconocida su abundante presencia en el discurso nativo, han llegado a proponerse como elemento clave para lograr una producción nativa (Wray, 2002; Skehan, 1998), pese al mencionado problema terminológico-conceptual que existe en torno a ellos.

En la actualidad, por tanto, no se duda de los beneficios que aporta el conocimiento de estas secuencias léxicas, especialmente en la interacción espontánea, contexto en el que prima la fluidez –de ahí la ventaja del almacenamiento y recuperación holística de estas secuencias– tanto en la codificación como en la descodificación lingüística. No obstante, se plantea el problema de que el estudiante de lenguas no es capaz de producir estos segmentos léxicos activando sus conocimientos gramaticales, tal y como se entiende tradicionalmente la gramática, con sus reglas abstractas y, a menudo, demasiado generalizadoras. Así, la cuestión que debe plantearse no es si estas unidades deben o no ser enseñadas, sino cómo ha de llevarse a cabo el proceso para que el aprendizaje resulte exitoso. Partiendo de la experiencia docente y del tratamiento de la parcela léxica que presentan los manuales didácticos, puede afirmarse que el paradigma metodológico del enfoque comunicativo aplicado a la enseñanza del léxico no resulta eficaz, pues no se aprende lo que se considera enseñado: la exposición al *input*, la práctica de destrezas, las actividades –cuando las hay– con bajo procesamiento cognitivo

---

<sup>1</sup> Cabe poner en duda, al hilo de lo que sostienen muchos sintactistas, la ausencia en una combinación léxica composicional de restricción semántica entre un predicado y sus argumentos; varía, en cualquier caso, la naturaleza de la restricción. Por ello, acaso el término *libre* no resulte apropiado.

<sup>2</sup> El más aplicado en el aula de idiomas es el enfoque léxico de Lewis (1993, 1997, 2000).

y sin atender al comportamiento combinatorio de la palabra y el énfasis en la producción abierta no conllevan el aprendizaje lexical. De los manuales se infiere que el enfoque comunicativo considera el aprendizaje del léxico como incidental y autónomo, como el de la L1<sup>3</sup>. Sin embargo, en un contexto formal de enseñanza la adquisición no se produce, pues requiere que el alumno, para poder asimilar una pieza o combinación léxica, se exponga numerosas veces a ella, lo cual resulta difícil en ese contexto. Aunque, mediante este método, el alumno aprende a comunicarse mejor –pues adquiere estrategias comunicativas al tiempo que desarrolla su competencia gramatical–, parece estancarse en un nivel B en lo que concierne a los aspectos de la precisión y la fluidez, debido a que su conocimiento léxico no evoluciona: la cantidad de palabras que tiene en el lexicón mental apenas se ve incrementada al pasar de una unidad didáctica a otra ni de un nivel a otro, y muchos errores se cometen a la hora de combinarlas en el eje sintagmático. Por tanto, estos problemas se deben, sobre todo, a que los materiales se centran mayoritariamente en el significado de las palabras aisladas, desatendiendo su combinatoria; y a que las actividades dedicadas al aprendizaje del léxico son muy escasas y, si las hay, no ofrecen oportunidades para que las palabras sean absorbidas y se puedan recuperar del lexicón mental ni a medio ni a largo plazo, lo que radica en el bajo nivel de procesamiento cognitivo que exigen tales actividades.

De lo expuesto anteriormente se concluye que no se puede esperar que el léxico de la L2 se adquiera de manera inconsciente, como el de la L1, cuando la cantidad de *input* a la que el alumno está expuesto en la L2 es infinitamente menor; e incluso ante la exposición repetida al *input*, conviene tener muy presente que “picking up target language forms from input when they do not carry information crucial to the task, appears unlikely for adults. Paying attention to language form is hypothesized to be facilitative in all cases” (Schmidt, 1990: 49). A su vez, como muchas de estas expresiones no se obtienen al aplicar las reglas gramaticales, se ha de pensar en una enseñanza explícita del léxico pero que no se limite a exigir un esfuerzo cognitivo

---

<sup>3</sup> El manual *Bitácora* (Difusión, 2011) marca el comienzo de una nueva etapa en la que se incide en la importancia del léxico en general y más concretamente en el de las combinaciones léxicas. El tratamiento del léxico es sistemático y no se trabaja de manera aislada; las palabras se muestran en su contexto combinatorio.

superficial, sino profundo; solo así se logrará el objetivo de recuperar la combinación a medio y largo plazo en el proceso de producción.

## **2. Hacia un concepto operativo de *colocación***

De acuerdo con lo señalado en el apartado anterior, la enseñanza de las expresiones multipalabra en el ámbito de las segundas lenguas puede resultar muy productiva. El siguiente paso al respecto tiene que ver con la delimitación de las expresiones englobadas en el concepto de *bloque semiconstruido*, cuyo elemento común se refiere a la frecuencia de coaparición de sus integrantes. Lewis, a la hora de seleccionar las expresiones multipalabra que deben enseñarse, parte de la definición aportada por la corriente estadística, que es la que se toma como referencia en la enseñanza del inglés y la que se está extendiendo en los estudios sobre aprendizaje y enseñanza del léxico en ELE, según se refleja, además, en los materiales didácticos. Desde su enfoque léxico, Lewis (2000: 132) define así la colocación: “collocation is the way in which words co-occur in natural text in statistically significant ways”. Colocaciones en sentido restringido, fórmulas sociales, marcadores discursivos, compuestos, locuciones, símiles estandarizados, proverbios, clichés, exclamaciones, combinaciones *libres* frecuentes, etcétera, conviven en el territorio de la secuencia formulaica, con la coaparición frecuente de sus integrantes y la recurrencia automática y habitual por parte de los usuarios de la lengua como únicos elementos comunes. La razón de que los criterios de determinación de estas combinaciones sean el de la frecuencia, en el primer caso, y el de la naturalidad, en el segundo, se debe a que los principales intentos de caracterización de las expresiones prefabricadas proceden de las perspectivas del análisis de corpus y de la psicolingüística; el aspecto de la naturalidad, además, conduce a determinar –a través de la discusión con nativos y aprendientes de modelos de adquisición y procesamiento de estas expresiones– si las secuencias se almacenan como una sola unidad en el lexicón mental (Stengers, 2009).

Sin embargo, la frecuencia no es un criterio que aporte información sobre la naturaleza lingüística de una combinación (Bosque, 2001; Sánchez Rufat, 2010). Frecuentes pueden ser estados, acciones o eventos que sean repetidos o

habituales sin tener por qué serlo desde el punto de vista del sistema lingüístico, como sucede con las acciones diarias del tipo *lavarse la cara*, *tender la ropa* o *arreglar el coche*. Se trata, además, de un concepto relativo, pues una combinación puede ser más o menos frecuente en relación con los tipos de textos que constituyan el corpus analizado. Por estos motivos, el alumno no puede trabajar con este criterio de forma autónoma<sup>4</sup>, es decir, no podrá distinguir lo que es una secuencia formulaica de lo que no lo es<sup>5</sup>. Lo mismo sucede con la naturalidad, pues incluso los nativos no tienen por qué coincidir en si una secuencia de palabras es natural, y se almacena, por tanto, en nuestra memoria como una sola unidad. Por todo ello, si el término *secuencia formulaica* (o *colocación*, *expresión multpalabra* o *bloque semiconstruido*) no restringe –en términos léxicos y sintácticos– los dominios que necesitan ser establecidos en los estudios lingüísticos, el empleo de este término y el concepto que designa no será útil como unidad funcional (diferenciada) de descripción en la didáctica de lenguas<sup>6</sup>. Para poder enseñar a identificar las combinaciones léxicas propias de la norma y del sistema, independientemente de su frecuencia, y evitar así los errores combinatorios, es necesario sistematizar estas combinaciones atendiendo a criterios puramente lingüísticos, dado que es evidente que no es posible dar un contenido homogéneo –al menos, no comprensible conceptualmente en términos teóricos– al variado grupo de fenómenos al que se aplica el término *secuencia formulaica*.

Partiendo del concepto de selección léxica de Bosque (2001, 2005), basado en la motivación de la restricción de las combinaciones, pueden distinguirse dos tipos de combinaciones léxicas: restricciones no sistemáticas y restricciones sistemáticas. En las restricciones no sistemáticas incluimos las unidades

---

<sup>4</sup> Recuérdese que el objetivo del enfoque léxico (Lewis, 1993, 1997, 2000) es lograr la autonomía del aprendiente equipándolo con estrategias para identificar en un *input* auténtico las expresiones multpalabra de la L2 dentro y fuera del aula.

<sup>5</sup> Stengers (2009) hace referencia a experimentos que demuestran el fracaso del alumno a la hora de tener que identificar estas secuencias partiendo de tal criterio.

<sup>6</sup> Bosque (2001) y Sinclair *et al.* (2005) señalan la falta de utilidad que supone recoger en un diccionario combinatorio ciertas unidades por el mero hecho de ser frecuentes; del mismo modo, nosotros cuestionamos la validez del criterio de la frecuencia en el ámbito de la enseñanza de lenguas.

fraseológicas<sup>7</sup> y las preferencias léxicas, en las que la restricción no viene impuesta por el sistema lingüístico, sino por factores en ocasiones difíciles de sistematizar, ya sean culturales, sociales o históricos. En muchos casos, desconocemos de qué son producto estas restricciones no sistemáticas, cuál es su motivación; si se las considera ajenas a la descripción del sistema lingüístico, deben situarse en el terreno de la norma. Tómese como ejemplo el adjetivo *fúnebre*. El alumno debe aprender que *fúnebre* selecciona a *coche* y a *honras*, mientras que su sinónimo *mortuario* se combina con *cámara* o *esquela*. Se desconoce la explicación semántica para estas preferencias léxicas, que, por otro lado, muestran el problema que supone definir un término en el aula por medio de un sinónimo al tiempo que se ignora su comportamiento léxico-sintáctico.

Por lo que respecta a las restricciones sistemáticas, en su delimitación deben considerarse las propiedades combinatorias de las palabras<sup>8</sup>; es decir, estas combinaciones son producto del carácter restrictivo que forma parte del sistema lingüístico. Los predicados seleccionan sus argumentos restringiendo el conjunto de piezas léxicas que pueden denotar de acuerdo con sus rasgos semánticos, por lo que las unidades seleccionadas no son piezas aisladas – como en el caso de la preferencia léxica–, sino clases léxicas. La teoría de los predicados de Bosque explica la capacidad de generación de estas restricciones sistemáticas: atender a estas restricciones, no como combinaciones en las que la relación entre los elementos es biunívoca, sino como casos de selección léxica, implica adquirir –de entrada– un

---

<sup>7</sup> La expresión *unidad fraseológica* designaría, en un primer grupo, las unidades que son conmutables por un enunciado, como las paremias (refranes, proverbios, citas y enunciados de valor específico) y las fórmulas rutinarias; y, en un segundo grupo, las unidades fraseológicas que no conforman un enunciado completo, esto es, las locuciones. A diferencia de las demás combinaciones léxicas, las unidades fraseológicas no son composicionales desde un punto de vista semántico, y no se forman aplicando las reglas de la gramática con libertad.

<sup>8</sup> Existen dos diccionarios combinatorios en español, ambos coordinados por I. Bosque: *Redes* (2005), centrado en las restricciones sistemáticas, principalmente en las intralingüísticas, que son las más complejas; y *Práctico* (2006), que muestra las combinaciones más habituales en las que suele aparecer una palabra, ya sea preferencia léxica o combinación sistemática.

conocimiento de la lengua generativo (basado en reglas)<sup>9</sup>, pues una misma relación léxico-sintáctica permite generar diferentes combinaciones<sup>10</sup>.

Las restricciones que los predicados imponen a sus argumentos pueden ser de dos tipos, de ahí que haya dos tipos de restricciones sistemáticas: si están fundamentadas únicamente en nociones de naturaleza física (Bosque, 2005), podemos construir el paradigma extensional de acuerdo con el conocimiento que tenemos del mundo; por el contrario, si la restricción no está basada únicamente en nociones físicas, tendremos que recurrir a nuestro conocimiento del idioma para obtener las piezas léxicas que conforman la extensión, de ahí que las denominemos restricciones sistemáticas intralingüísticas o colocaciones –en sentido restringido. Los siguientes ejemplos ilustran la diferente naturaleza de la predicación correspondiente a un mismo verbo: *La clara de huevo no ha cuajado* y *El plan no ha cuajado*. La restricción de naturaleza extralingüística de la predicación impuesta por *cuajar*, en su interpretación física ‘tomar consistencia sólida’, se combina con una cantidad de argumentos que pueden agruparse en una clase léxica de gran extensión: los sustantivos que designan diversas sustancias líquidas o no<sup>11</sup>. Una vez conocido el significado del predicado, tanto si se es hablante nativo como si no, pueden añadirse piezas léxicas a esta clase aplicando el sentido común, como *leche*, *yogur* o *nieve*. De igual manera, podemos añadir más sustantivos al paradigma extensional de este verbo si lo tomamos con el sentido figurado de ‘llegar a buen término’: se combina con sustantivos de persona, especialmente si se refieren al que desarrolla o pretende desarrollar una labor profesional (*torero*, *centrocampista*, *escritor*). Tenemos, por lo tanto, otro caso de restricción basada en una noción física. Sin embargo, aunque las restricciones sistemáticas intralingüísticas continúan siendo casos de selección léxica, la restricción, a diferencia de lo que ocurre en el grupo

<sup>9</sup> Frente al conocimiento memorístico, basado en la memoria, que es el único que se puede desarrollar en los casos de restricciones no sistemáticas.

<sup>10</sup> Obsérvese que esta productividad se pierde si la direccionalidad de la selección de naturaleza semántica parte de los argumentos para llegar a los predicados, como sucede con los trabajos derivados de la Teoría Sentido-Texto.

<sup>11</sup> A partir de la definición que aparece en el *DRAE*: “transformar una sustancia líquida en una masa sólida y pastosa. Se utiliza especialmente referido a sustancias que contienen albúmina”. Las combinaciones que señalamos en torno al verbo *cuajar* están tomadas de *Redes* (615-616).



anterior, viene dada en función de las propiedades del predicado, ya sea este un adverbio, un adjetivo, un verbo, una preposición o un sustantivo. No basta con el conocimiento del mundo: se necesita un conocimiento profundo del predicado para obtener el paradigma extensional, que no tiene por qué coincidir en otra lengua, de ahí que sea una fuente de errores para el estudiante. De esta manera, en su sentido figurado, *cuajar* se combina, entre otras clases léxicas, con la formada por sustantivos que denotan plan, iniciativa o proyecto, así como los que se refieren a algunos de sus contenidos, en particular a actividades relativas al cambio o la renovación de alguna cosa, como *plan*, *proyecto*, *idea*, *reforma*, *programa*. Pese a que este verbo se combina con sustantivos que pertenecen a diferentes clases léxicas que, a priori, partiendo del significado del verbo, no pueden señalarse, el hecho de que el sentido figurado no sea más que una extensión natural del literal –de ahí que las combinaciones resultantes sean composicionales– provoca que la deducción de aquel nos permita descodificar, en muchos casos, la combinación, aunque no codificarla.

Teniendo en cuenta las anteriores características de cada tipo combinatorio, puede darse un paso más con respecto a la propuesta de Lewis y ahondar en el concepto de léxico gramaticalizado, que él mismo plantea, pero cuyo tratamiento resulta un tanto superficial. En coapariciones frecuentes del tipo *conceder un deseo* y *conceder un permiso*, en lugar de tratar de que el alumno fije y almacene en la memoria estas expresiones como relaciones biunívocas, se puede aprovechar la oportunidad para considerar estas combinaciones como integrantes de la restricción intralingüística o colocación a la que pertenecen. De esta manera, se obtienen varias combinaciones léxicas de una misma relación léxico-sintáctica, lo que resultaría muy rentable para el alumno; así, junto a *deseo*, también constituyen argumentos de *conceder* otros sustantivos que designan lo que se desea o se espera obtener, como *capricho* o *antojo*, que dan lugar a combinaciones de una misma colocación; y junto a *permiso*, *conceder* selecciona otros sustantivos que conforman la clase léxica integrada por sustantivos que denotan licencia, consentimiento o aprobación,

como *permiso, licencia, autorización, visado, beneplácito* (Redes: 524-525)<sup>12</sup>, que conformarían una segunda colocación en torno al predicado *conceder*.

### 3. Enfoques de la lingüística cognitiva

Hasta no hace mucho tiempo, plantearse una enseñanza explícita del léxico implicaba únicamente una *atención a la forma*<sup>13</sup> (Long, 1991), concepto por el que se entiende que dentro de una tarea comunicativa se deben incorporar actividades esporádicas de *toma de conciencia*<sup>14</sup> de los fenómenos lingüísticos y, según el cual, el docente, además, debe hacer un uso no intensivo de las técnicas de recuperación e identificación de errores. Lewis parte de estas premisas y propone la enseñanza explícita de las expresiones multipalabra a través de su ejercicio clave conocido como *segmentación* de la lengua<sup>15</sup>, prestando así atención a la forma dentro de una actividad comunicativa. Esta práctica consiste en equipar al alumno en el aula con las estrategias para identificar del *input* los bloques prefabricados de un texto oral o escrito, para que más adelante, y fuera del aula, sea capaz de reconocer y seleccionar de forma autónoma aquellos que sean más necesarios y útiles para su ampliación del léxico en cada momento<sup>16</sup>. No obstante, varios experimentos (Boers, 2008, 2009a, 2009b; Stengers, 2009) demuestran que segmentar bloques durante todo un curso académico no supone adquirir la capacidad de identificar segmentos nuevos, por lo que no es adecuado equiparar esta práctica al logro de una autonomía en el aprendizaje de las combinaciones léxicas. Por otro lado, los resultados muestran que los bloques segmentados *enseñados* durante el curso no han sido *aprendidos* o retenidos a medio plazo. Lo cierto es que la

---

<sup>12</sup> Nótese el problema de aportar sinónimos en el aula sin atender al contexto sintáctico de las palabras, pues es difícil que sean sustituibles dos términos en todos los contextos de aparición. Así, a menudo se propone el verbo *dar* como sinónimo de *conceder*. Estos sinónimos funcionan en la segunda colocación, pero, en lo que se refiere a la primera, solo son intercambiables (si prescindimos del componente pragmático) en el caso de *capricho*.

<sup>13</sup> Traducción nuestra de *focus on form*. A este concepto hace referencia Schmidt en la cita recogida más arriba.

<sup>14</sup> Traducción nuestra de *noticing*.

<sup>15</sup> Traducción nuestra de *chunking*.

<sup>16</sup> Nótese que la anterior cita de Schmidt –paradójicamente recogida por Lewis (2000: 163)– refuta que el aprendizaje sistemático de estas unidades pueda ser autónomo, objetivo último del enfoque léxico.

capacidad de procesamiento que demanda este ejercicio es muy limitada y requiere múltiples encuentros con la misma secuencia léxica; esto es aplicable a la mayoría de las actividades para la enseñanza de las expresiones multipalabra que se acogen al enfoque léxico.

Como alternativa al enfoque léxico, pueden tomarse los principios de la lingüística cognitiva, que concibe el comportamiento lingüístico como un elemento integrante de la cognición y no como una facultad aislada; por ello, las teorías generales del procesamiento cognitivo pueden aplicarse al aprendizaje de una segunda lengua. Son componentes básicos del procesamiento cognitivo –aplicables al aprendizaje de una palabra– los estadios de la retención, la fijación y la reutilización. En estas tres fases la memoria es fundamental, a corto plazo –en el primer estadio– y a largo plazo –en el segundo y en el tercero. Existen teorías cognitivas relacionadas con la memorización como la *teoría del rastro*<sup>17</sup> (Baddeley, 1997), que defiende que encontrarse repetidamente con una unidad refuerza su permanencia en la memoria. En segundo lugar, la teoría de los niveles de procesamiento establece que, cuanto más profundo sea el nivel en el que se procesa mentalmente la información, más posibilidades hay de que se memorice a largo plazo. Por procesamiento profundo se entiende ‘elaborar’ la información léxica (Barcroft, 2002); la operación mental que exige la elaboración se caracteriza por su complejidad, y se opone al bajo nivel de procesamiento cognitivo que requieren tanto las actividades sobre léxico que se incluyen en los manuales comunicativos como la actividad de *segmentación* propuesta por Lewis en su enfoque léxico, y basada en el concepto de *atención a la forma*. Existen dos tipos de elaboración: semántica y estructural (Barcroft, 2002). La primera se refiere a la operación mental relacionada con el significado de una palabra o expresión multipalabra, y la segunda se corresponde con la operación mental realizada respecto de las propiedades formales de la palabra o del bloque semiconstruido. A esta distinción añadimos un tercer tipo de elaboración que hemos denominado léxico-sintáctica, que consiste en la operación mental que hay que realizar para relacionar el significado de una palabra con su combinatoria léxica; esto es, que se logre aprender el uso combinatorio de una palabra a través de su significado. Por lo tanto, para que

---

<sup>17</sup> Traducción nuestra de *trace theory*.

la lengua pase a estar analizada y almacenada, hay que lograr que el alumno *elabore* la información léxica. Si no participa en la elaboración, el alumno se queda estancado tras una mera percepción de la unidad, y no será capaz de recuperarla para su producción; necesita, así, una mayor implicación cognitiva y más posibilidades de ensayo en el aula. Para ello, la intervención del profesor es fundamental; la enseñanza ha de ser no solo expresa –como defiende Lewis–, sino también intensiva. Deben activarse procedimientos mnemotécnicos que ayuden al alumno a retener y a fijar la unidad, que deberá recuperar más adelante en el proceso de aprendizaje<sup>18</sup>.

Aunque se ignora el modo específico en el que el alumno asimila la lengua, enseñar las combinaciones léxicas desde la perspectiva de la lingüística cognitiva, basada en la motivación de muchas de estas expresiones, se presenta como una opción más eficaz que las anteriores. La lingüística cognitiva, al concebir el fenómeno lingüístico como un reflejo del proceso cognitivo general, parte de la idea de que la lengua no es arbitraria, como sucede con el pensamiento figurativo. De esta manera, los significados de las formas lingüísticas están motivados por la experiencia física, social y cultural de la comunidad de hablantes. Si el alumno es consciente de que la L2 no es el resultado de combinaciones azarosas, sino de restricciones motivadas, ya sean sistemáticas o no sistemáticas, es más probable que realice prácticas mnemotécnicas en lugar de centrarse en otras menos eficaces asociadas a la memorización ciega. Y para desbloquear este potencial mnemotécnico, como se ha mencionado anteriormente, la figura del profesor es muy importante, pues posibilita y potencia el aprendizaje.

#### 4. Sobre la utilización de los corpus

Como ya hemos señalado, una de las líneas de investigación que surge en la tendencia de revalorización del léxico proviene del campo del análisis de corpus<sup>19</sup>. Estos estudios, basados en la frecuencia de coaparición de las

---

<sup>18</sup> Recuérdese que en el enfoque léxico el papel del profesor en la enseñanza del léxico se reducía a atender a la forma; y en el enfoque comunicativo ni siquiera se reconoce su intervención.

<sup>19</sup> Sobre los orígenes, avances y progresos en la descripción de la lengua como fruto de la aparición de la lingüística de corpus, *vid.* Sinclair *et al.* (2005).

palabras, subrayan que la capacidad de almacenamiento del lexicón mental en un hablante nativo es muy superior a lo que se pensaba (Carter y McCarthy, 1997), y que el cincuenta por ciento del discurso nativo está constituido por expresiones multipalabra prefabricadas (Erman y Warren, 2000). La hipótesis de trabajo de estos autores está cimentada en el concepto de *principio de idiomática* (Sinclair, 1991), que establece que gran parte del proceso de producción del lenguaje se fundamenta en la recuperación en la memoria de bloques semiconstruidos, caracterizados por la coaparición frecuente de los elementos que los integran e identificados como naturales por el oído nativo, de acuerdo con los parámetros con los que se trabaja en la psicolingüística. Esta visión conlleva la superación del *principio de selección libre*, que basaba el proceso de producción lingüística en la elaboración de estructuras sintácticas –determinadas por unas reglas gramaticales– en las que se insertan las piezas léxicas. Con este cambio se reconoce la interdependencia entre la gramática y el léxico, y surge el concepto de lengua como léxico gramaticalizado, concepto por el que entendemos que “las palabras no significan algo y ADEMÁS se combinan de cierta manera, sino que en gran medida se combinan de cierta manera PORQUE expresan precisamente esos significados”<sup>20</sup> (Bosque, 2005: XVIII); la sintaxis llega a ser concebida como un fenómeno emergente (Tomasello, 2003) que se adquiere, en gran medida, a partir del uso de segmentos memorizados.

Así, los corpus se sitúan en el origen del proceso que permite extraer conclusiones útiles en torno al comportamiento combinatorio de las unidades léxicas y a su significado, proceso que desemboca, por ejemplo, en la delimitación del concepto de colocación llevada a cabo anteriormente. Pero, además, la aplicación de los corpus puede contribuir al aprendizaje de dichas combinaciones mediante actividades convenientemente diseñadas. En este sentido, cobran especial importancia los corpus informatizados<sup>21</sup>, que permiten el acceso inmediato en el proceso del aprendizaje y que ofrecen, claro está, usos lingüísticos que, aunque constituyen una muestra parcial de la lengua, son reales y representativos. Entre los corpus que pueden consultarse en Internet destacan los de la Real Academia Española, fundamentalmente el

<sup>20</sup> Las mayúsculas son del original.

<sup>21</sup> Como síntesis del asunto, *vid.* Pitkowski y Vázquez (2009).

*CREA (Corpus de referencia del español actual)*. Este ingente corpus permite búsquedas filtradas por el medio de comunicación (oral, escrito y, dentro de este, libro, prensa y otros tipos de texto), periodos de tiempo (de cuatro años desde 1975), área (España, América y, dentro de esta, las franjas lingüísticas tradicionales) y, finalmente, por tema (hasta veinte ámbitos). El gran inconveniente es que el *CREA* no está lematizado, es decir, no permite relacionar en una búsqueda un término con todas las formas de su misma familia. Lo mismo ocurre con otros corpus informatizados, como el *Corpus del español* (<http://www.corpusdelespanol.org/>), *SOL (Spanish Online)*: <http://spraakbanken.gu.se/konk/rom2>) o *Corpus Concordance Spanish* ([http://www.lex tutor.ca/concordancers/concord\\_s.html](http://www.lex tutor.ca/concordancers/concord_s.html)); los tres son útiles para la búsqueda de contextos léxicos, pero los resultados se limitan a la forma introducida en el buscador. Corpus lematizados son, por ejemplo, *AnCora\_Es* (<http://elic.ub.edu/corpus/es>) y *Arthus* (<http://adesse.uvigo.es/>), pero, dado que sus principales objetivos no tienen que ver con las propiedades combinatorias de las palabras, sino con funciones sintácticas o semánticas, no disponen de un sistema de concordancias orientado a la combinatoria léxica; es decir, los resultados se muestran de una manera que no permite discernir, a primera vista, las palabras, antepuestas o pospuestas, con las que se combina la forma introducida en el buscador. En relación con ello, resultan especialmente útiles el *CREA* y el *Corpus del español*, que ofrecen la posibilidad de detallar la posición de la forma buscada en la oración y de incluir otras palabras en la búsqueda, explicitando la posición en que se desea que aparezca (también lo permite el *Arthus*, pero, como decimos, la búsqueda se orienta más hacia las funciones sintácticas). Para contribuir, por tanto, a la adquisición y fijación lexicales de una manera eficaz, resulta muy útil que las actividades que se diseñen tengan en cuenta la rentabilidad que ofrece el concepto de colocación en la enseñanza del léxico y aprovechen las búsquedas que permiten los corpus informatizados, sin perder de vista, finalmente, los enfoques de la lingüística cognitiva.

### **5. Actividad para el aprendizaje de colocaciones con corpus como *input***

En relación con lo dicho, la siguiente actividad pretende combinar la aplicación de los corpus textuales en el aula de ELE con el aprovechamiento de las colocaciones en el apartado específico de la enseñanza del léxico. Se

pretende aquí, no que simplemente el alumno identifique en el *input* los bloques semiconstruidos, sino que, utilizando el *input* que los corpus suponen, sea capaz de producir otros, toda vez que haya descifrado la relación que se establece entre un predicado y sus argumentos. Para esto último nos servimos del concepto de colocación expuesto más arriba, según el cual los predicados seleccionan sus argumentos restringiendo el conjunto de piezas léxicas que pueden denotar de acuerdo con sus rasgos semánticos; y también de los enfoques de la lingüística cognitiva, mediante los que el alumno concebirá la lengua, no como un conjunto de relaciones azarosas, sino como el resultado de relaciones lógicas desde los planos semántico y léxico; todo ello facilitará la elaboración de la información léxica, es decir, ayudará al alumno a producir, con la ayuda de los corpus como *input*, combinaciones léxicas basadas en el conocimiento de las propiedades combinatorias de los predicados y de su carácter restrictivo.

La actividad que se propone está integrada en una hipotética unidad didáctica fundamentada en los principios teóricos del enfoque comunicativo –esto es, se apoya en sus teorías lingüística, pedagógica y de aprendizaje. Dicha unidad está planteada para un grupo de nivel B2. Para alcanzar los correspondientes objetivos comunicativos, los componentes lingüísticos que serán trabajados en el proceso de enseñanza-aprendizaje a través de tareas posibilitadoras y de comunicación versarán, en lo que se refiere al componente léxico, sobre el tema o área de interés de la unidad didáctica. En principio, muchos temas caben en esta propuesta; en este caso, y como muestra, hemos optado por el ámbito detectivesco, a partir del cual pueden establecerse vínculos con otros campos de interés. La fase de presentación –de acuerdo con el paradigma metodológico del enfoque comunicativo– aporta como muestra de lengua un texto motivador acerca del tema sobre el que versa la unidad, texto que constituye la herramienta que propiciará las actividades subsecuentes. El pasaje escogido es el que sigue<sup>22</sup>:

*Este texto no pretende ser una guía para cometer delitos. Es simplemente un artículo informativo sobre la delincuencia y la detección de crímenes. Todos*

---

<sup>22</sup>Adaptado de <<http://www.taringa.net/posts/info/11647496/como-cometer-el-crimen-perfecto.html>> (27/10/2011).

*los días se cometen crímenes que no son resueltos y quedan en el más absoluto misterio, e incluso algunos pasan desapercibidos. Se podría decir que muchos de estos casos son el crimen perfecto, que se consigue con precauciones como las que se señalan a continuación.*

*El ADN es la forma más segura para identificar a la persona que cometió un delito. Para no correr el riesgo, el crimen debe realizarse en un lugar donde se pueda encontrar gran cantidad de ADN de personas extrañas, como un parque o un centro comercial. La víctima debe elegirse al azar, sobre todo si el crimen es el asesinato, para que no puedan establecerse relaciones entre ella y el autor del crimen.*

*El mejor momento es por la mañana, cuando las personas están aún adormiladas: no reparan en detalles y son más vulnerables. No se debe intentar ocultar el crimen, sino la relación con el crimen. Si se mata a alguien, no se debe arrojar el cuerpo al río o a un descampado. Deshágase del arma y de la ropa que llevaba.*

*Por último, debe evitarse ver la televisión y leer periódicos, pues la policía puede usar estos medios como herramienta psicológica para presionar al autor del crimen. No es conveniente celebrar el crimen o jactarse de él. Aunque el responsable sea detenido, no podrá ser condenado si toma las precauciones necesarias, pues no habrá pruebas para que lo incriminen.*

Una vez leído el texto, los alumnos desarrollarán las actividades correspondientes, entre las que se incluyen en la fase de conceptualización las encaminadas a la *elaboración* de colocaciones mediante el uso de los corpus como *input*<sup>23</sup>. He aquí algunas muestras de ejercicios que potencian el descubrimiento del funcionamiento léxico-sintáctico de las piezas léxicas que han de ser aprendidas:

*1) En el texto aparece la expresión cometer delitos. ¿Con qué otras palabras crees que puede combinarse el verbo cometer? Este ejercicio se ampara en la teoría de los niveles de procesamiento; se pretende que el alumno elabore el comportamiento léxico-sintáctico del predicado cometer.*

---

<sup>23</sup> No contamos aún con una relación de colocaciones distribuidas por niveles, lo cual es comprensible teniendo en cuenta que no hay acuerdo en torno a la delimitación del concepto. Las colocaciones que deben aprenderse en cada nivel dependen, de momento, del criterio del profesor.



Una actividad de este tipo potencia el descubrimiento de las reglas combinatorias gracias a la muestra de lengua aportada que expone al alumno a ejemplos contextualizados que fomentan la comprensión del valor de *cometer* y posibilitan el análisis y la formulación de hipótesis –el alumno observará que los sustantivos seleccionados por el verbo designan resultados de la acción de quebrantar leyes o normas y están marcados negativamente.

2) Busca en los corpus CREA (<http://corpus.rae.es/creanet.html>), SOL (<http://spraakbanken.gu.se/konk/rom2/>) y en Google (<http://www.google.es>) las combinaciones (concordancias) con *cometer*. Elabora un asociograma con el verbo *cometer* en el centro. ¿Qué tienen en común las palabras que rodean al verbo? Este ejercicio se fundamenta, en primer lugar, en la teoría de los niveles de procesamiento, pues continúa con el proceso de elaboración. El propio aprendiente puede contrastar su hipótesis en el corpus real y aceptarla o refutarla; tras la búsqueda se comprueba que los sustantivos seleccionados por el verbo están marcados negativamente.

Antes de desarrollar ambas actividades, el profesor habrá explicado a los alumnos qué es un corpus textual. La elección del corpus para trabajar en el aula depende, igualmente, del criterio del profesor. Si se utiliza el CREA, conviene filtrar la búsqueda, ya que el volumen de información que se obtenga puede ser abrumador; para la actividad propuesta, sugerimos buscar únicamente en el apartado de “prensa” y en España. Si se prefiere, pueden utilizarse corpus más pequeños como SOL, o, incluso, puede aprovecharse Internet a través del buscador Google. En cualquier caso, el alumno de este nivel puede clasificar los sustantivos que aparecen a la derecha o izquierda del verbo y que son seleccionados por él. Así, por un lado, obtendrá sustantivos que denotan error, como *fallo*, *equivocación*, *desliz*, *torpeza*, *lapsus*, *patinazo* o *desacierto* y, por otro, los que designan resultados de la acción de quebrantar leyes, como *delito*, *falta*, *crimen*, *infracción*, *robo*, *asesinato*, *pecado* o *violación*.

Esta actividad se apoya también en la teoría del rastro, puesto que el alumno se va a encontrar repetidamente con una misma pieza léxica y con una misma relación léxico-sintáctica –es decir, colocación–, que se concreta en las diferentes concordancias recogidas (y clasificadas) por el alumno. Esta búsqueda en el corpus reforzará la permanencia de la colocación en la

memoria, contribuyéndose así a potenciar el conocimiento generativo o basado en reglas. De igual manera, muchas de esas coapariciones léxicas se encuentran, además, muy repetidas, lo que contribuirá a su conocimiento memorístico o formulaico; así, pasarán a formar parte del lexicón mental del alumno como secuencias formulaicas a las que podrá recurrir de manera holística tanto en la producción como en la codificación lingüística. Por lo tanto, el alumno asimilará el significado amplio de *cometer* y podrá producir combinaciones a partir de él, lo cual resulta más rentable que otros procedimientos, como la clásica memorización de las listas de vocabulario, ya sea de palabras o de combinaciones. Y con este tipo de ejercicio se desarrolla la autonomía del aprendiente sobre su propio aprendizaje, pues es una práctica que puede desarrollar fuera del aula.

Por otra parte, los corpus propuestos para las actividades anteriores no están lematizados, es decir, no se obtienen resultados con todas las formas flexionadas de la palabra introducida en la búsqueda. Así, en principio, las búsquedas deben realizarse a partir de una forma concreta. No obstante, también es posible utilizar corpus lematizados (*AnCora\_Es, Arthus*), de modo que, al introducir el infinitivo, el buscador aportará combinaciones con todas las formas del verbo. Aportamos a continuación otras dos actividades:

3) *En el texto puede también leerse arrojar el cuerpo. ¿Con qué otras palabras crees que puede combinarse el verbo arrojar?*

4) *Busca en el corpus las concordancias con arrojar. ¿Hay combinaciones en las que el significado de arrojar sea diferente al de arrojar el cuerpo? ¿En qué consiste esa diferencia?*<sup>24</sup> *Elabora un asociograma con el verbo arrojar en el centro. Con estas últimas actividades, basadas de nuevo en la teoría de los niveles de procesamiento, el alumno avanzará desde el significado literal del verbo hasta su uso figurado, derivado del anterior; podrá elaborar la información léxica y producir nuevas combinaciones.*

La unidad incluiría en la fase de práctica actividades para entrenar la fijación y producción de formas, por lo que se incluirían ejercicios de manipulación de

---

<sup>24</sup> El profesor puede servirse de los diccionarios combinatorios *Redes* y *Práctico* para completar los asociogramas elaborados por los alumnos.

enunciados (huecos, sustitución de la palabra subrayada...) y de producciones pautadas (con roles, intenciones y formas lingüísticas predeterminados). Dado que, como se ha dicho, los manuales de ELE no trabajan el léxico sistemáticamente de manera expresa, y puesto que en el Instituto Cervantes no se especifica qué colocaciones se deben aprender en cada nivel, el profesor puede servirse de los textos aportados en los manuales –que contienen los elementos lingüísticos contextualizados que serán trabajados, en el caso de la gramática, en la unidad– y abordar el componente léxico a partir de propuestas como la descrita aquí.

## 6. Conclusiones

En definitiva, las aportaciones de la lingüística cognitiva permiten que, en el aprendizaje del léxico, el alumno retenga eficazmente las unidades, las recupere posteriormente para su producción y las enriquezca con aportaciones propias. En ese proceso, las colocaciones, entendidas como combinaciones producto del carácter restrictivo que forma parte del sistema lingüístico, resultan unidades muy rentables, pues ofrecen la posibilidad de ser asimiladas de acuerdo con sus propiedades semánticas, léxicas e incluso sintácticas. Partiendo de estas premisas, es posible diseñar actividades que cuenten con la ayuda de los corpus textuales informatizados, los cuales pueden funcionar como un *input* muy provechoso que los alumnos pueden manejar autónomamente. Se trata, en definitiva, de abrir un camino que contribuya al aprendizaje lexical y que supere los obstáculos de enfoques que, si bien han desempeñado un papel muy importante en este ámbito, se revelan insuficientes.

## Bibliografía

- BADDELEY, A. (1997): *Human Memory: Theory and Practice*. Hove: Psychology Press.
- BARCROFT, J. (2002): “Semantic and structural elaboration in lexical acquisition”. *Language Learning*, 52: 323-36.
- BOERS, F. (2008): “How cognitive linguistics can further vocabulary teaching”. En F. Boers y S. Lindstromberg (eds.), *Cognitive Linguistic Approaches to Teaching Vocabulary and Phraseology*. Berlín: Mouton de Gruyter.

- BOERS, F. Y LINDSTROMBERG, S. (2009): *Optimizing a Lexical Approach to Instructed Second Language Acquisition*. Basingstoke: Palgrave Macmillan.
- BOERS, F. (2009): "Optimizing a lexical approach in TEFL". Conferencia inédita impartida en la Universidad de Extremadura (25/5/2009).
- BOSQUE, I. (2001): "Sobre el concepto de colocación y sus límites". *Lingüística española actual*, XXIII-1: 9-41.
- BOSQUE, I. (2005): *Redes. Diccionario combinatorio del español contemporáneo*. Madrid: SM.
- BOSQUE, I. (2006): *Práctico. Diccionario combinatorio práctico del español*. Madrid: SM.
- CARTER, R. Y MCCARTHY, M. (1997): *Exploring Spoken English*. Cambridge: Cambridge University Press.
- ERMAN, B. Y WARREN, B. (2000): "The idiom principle and the open choice principle". *Text*, 20: 29-62.
- LEWIS, M. (1993): *The Lexical Approach*. Londres: Language Teaching Publications.
- LEWIS, M. (1997): *Implementing the Lexical Approach*. Londres: Language Teaching Publications.
- LEWIS, M. (2000): *Teaching Collocation. Further Developments in the Lexical Approach*. Londres: Language Teaching Publications.
- LONG, M. (1991): "Focus on form: A design feature in language teaching methodology". En K. de Bot, R.B. Ginsberg y C. Kramersch (eds.), *Foreign Language Research in Cross-Cultural Perspective*. Amsterdam/Philadelphia: John Benjamins, 39-52.
- NATTINGER, J.R. Y DECARRICO, J.S. (1992): *Lexical Phrases and Language Teaching*. Oxford: Oxford University Press.
- PITKOWSKI, E.F. Y VÁSQUEZ GAMMARRA, J. (2009): "El uso de los corpus lingüísticos como herramienta pedagógica para la enseñanza y aprendizaje de ELE". *Tinkuy: boletín de investigación y debate*, 11: 31-51.
- SÁNCHEZ RUFAT, A. (2010): "Apuntes sobre las combinaciones léxicas y el concepto de colocación". *Anuario de Estudios Filológicos*, XXXIII: 291-306.
- SCHMIDT, R. (1990): "The role of consciousness in second language learning". *Applied Linguistics*, 11: 129-158.

- SINCLAIR, J. (1991): *Corpus, Concordance, Collocation: Describing English Language*. Oxford: Oxford University Press.
- SINCLAIR, J. *et al.* (2005): *English Collocation Studies: The OSTI Report*. Londres: Continuum.
- SKEHAN, P. (1998): *A Cognitive Approach to Language Learning*. Oxford: Oxford University Press.
- STENGERS, H. (2009): *The Idiom Principle Put to the Test: An Exercise in Applied Comparative Linguistics* (tesis doctoral inédita). Bruselas: Universidad Vrije de Bruselas.
- TOMASELLO, M. (2003): *Constructing a Language: A Usage-Based Theory of Language Acquisition*. Cambridge: Harvard University Press.
- WRAY, A. (2002): *Formulaic Language and the Lexicon*. Cambridge: Cambridge University Press.