

Cometer delitos y dictar sentencias: las colocaciones en los estudios de corpus. Pasado y futuro.

*Committing Crimes and Issuing Sentences: Collocations in Corpus Studies.
Past and future.*

Silvia Aguinaga Echeverría
Universidad de California, Davis
saguinaga@ucdavis.edu

Resumen

Hay una brecha en relación con la definición del término colocación y los estudios de corpus, además de una falta de estudios centrados en la enseñanza y adquisición de colocaciones en español como lengua extranjera (L2). El objetivo de este artículo es discutir las diferentes definiciones de colocación presentadas en los últimos años para ofrecer una definición práctica y transparente que ayude en futuras investigaciones de corpus y poder programar la enseñanza de estos ítems.

Palabras clave: adquisición de vocabulario, colocaciones, estudios de corpus, español como Lengua Extranjera.

Abstract

There is a gap between corpus studies and the definition of collocations, in addition to a lack of research focused on the teaching and acquisition of collocations in Spanish as Second Language (L2). The objective of this article is to discuss the different definitions of collocations presented in recent years to offer a practical and transparent definition that will help in future corpus investigations and to program teaching these items.

Keywords: Vocabulary development, collocations, corpus studies, Spanish as a Second Language.

1. DEFINICIÓN E IDENTIFICACIÓN

En 1964 se les pidió a los jueces que formaban el Tribunal Supremo de Justicia de los Estados Unidos que consideraran la legalidad de las leyes que hacían referencia a la «obscenidad» en el estado de Ohio. El juez Potter Stewart concluyó lo siguiente:

I have reached the conclusion... that under the First and the Fourteenth Amendments criminal laws in this area [obscenity] are constitutently limited to hard-core pornography. I shall not today attempt further to define the kinds of material I understand to be embraced within that shorthand description; and perhaps I could never succeed in intelligibly doing so. But I know it when I see it.¹

Sorprendentemente, este método de definir un concepto, que se basa en la intuición del que observa, lo han usado y aceptado durante años los académicos para reconocer colocaciones en un texto (Bahns, Burmeister y Vogel, 1986; Erman y Warren, 2000). No obstante, en el caso de las colocaciones, nos encontramos con diferentes definiciones (principalmente basadas en la fraseología o en análisis de frecuencia) y tipos (sustantivo + verbo, verbo + preposición + sustantivo, sustantivo + adjetivo, sustantivo + preposición + sustantivo, verbo + adverbio, verbo + adjetivo, adverbio + adjetivo) que explicaremos con más detalle a continuación, al igual que observamos colocaciones transparentes y opacas, las cuales difieren en su aspecto semántico. Entendemos, por tanto, que es necesario un mapeo sistemático de los estudios previos para comprender sus diferencias y puntos comunes y acceder a una definición que nos ayude a catalogar las colocaciones sin recurrir a la mera intuición del hablante.

O'Donnell, Römer y Ellis (2013) hacen un estudio con el objetivo de encontrar la metodología apropiada para reconocer los tipos de colocaciones. En su investigación comparan textos de hablantes nativos (HN) con los de hablantes no nativos (HNN) y utilizan cuatro medidas diferentes para el análisis. Una de las metodologías de análisis se basa en la «norma nativa»: algunas secuencias pueden presentar alta frecuencia de producción en aprendientes, pero no reflejan las normas nativas (2013: 90-91). Sorprendentemente, el análisis basado en la «norma nativa» no presenta variación en los resultados al comparar los corpus de lengua materna (L1) y lengua extranjera (L2), en contraposición a los otros métodos utilizados: (1) frecuencia de *n-gram* (frecuencia de ítems); (2) asociación de información mutua (IM): los autores afirman que algunas secuencias están *glued together* y, por lo tanto, las medidas de asociación entre los ítems tienen mayor relevancia que su frecuencia total; (3) o *phrase-frames (P-frame)*: con esta medida se pretende observar la variabilidad de las unidades del conjunto, es decir, si estas son fijas o no. Esto es, si únicamente usamos la perspicacia del hablante nativo para clasificar las colocaciones producidas por los alumnos de una L2 u otros nativos, nos podemos encontrar variación en los resultados. En esta misma línea de pensamiento, Wray (2002: 23) identifica varios problemas en el uso de la intuición del nativo como metodología: (1) solo se puede implementar con un corpus de tamaño reducido, (2) puede ocurrir que debido a la fatiga se produzcan juicios inconsistentes, (3) puede haber variación entre los criterios de los jueces, (4) ya que en ocasiones no hay una única respuesta, (5)

¹ Texto extraído de Nongbri, B., *Before Religion* (2013: 15).

finalmente, el uso de la intuición en este tipo de análisis puede llevar a equívocos porque «application of intuition in such a way may occur at the expense of knowledge we do not have at the surface level of awareness.» Asimismo, las expresiones idiomáticas, los proverbios, los modismos o las colocaciones son diferentes tipos de conjuntos léxicos dentro del amplio abanico del lenguaje formulaico, el cual a menudo se define por tener una alta concurrencia, una estructura fija y por no ser semánticamente composicional² (Wood, 2005). Sin embargo, al indagar en las particularidades de cada unidad léxica compleja se observa que las características antedichas no siempre se aplican a todas por igual. Wray (2002: 44) afirma que «the identification of formulaic sentences in a text is extremely difficult, largely because of the absence of any single definition able to capture all the relevant features at once».

Al desechar la intuición del nativo como método de análisis y reconocimiento de unidades fraseológicas, nos encontramos que son dos los acercamientos usados para su identificación, uno de tipo cualitativo enmarcado en la fraseología y otro más cuantitativo centrado en la frecuencia (Nesselhauf, 2005). Es importante entender ambos puntos de vista para justificar la elección de la definición que se propone al final del artículo, lo que entendemos como un punto intermedio entre la fraseología y la frecuencia.

El objetivo de este artículo es, por tanto, discutir las diferentes definiciones de colocación presentadas en los últimos años para ofrecer una definición práctica y transparente que ayude en futuras investigaciones de corpus y a programar la enseñanza de dichos ítems.

2. ESTUDIOS FRASEOLÓGICOS

Los principales autores de esta corriente conciben las colocaciones como asociaciones frecuentes de palabras; sin embargo, la semántica es el punto de unión entre el núcleo (*node*) y el colocativo o colocado (*collocate*) (Sinclair, 1991): el núcleo es el elemento independiente y el colocativo toma su significado al combinarse con el núcleo. Por ejemplo, en colocaciones como *palpitar el corazón*, *enemigo acérrimo*, *terrón de azúcar* o *locamente enamorado*, observamos que *palpitar* implica *corazón*, *acérrimo* implica *enemigo*, *terrón* implica *azúcar*, y *enamorado* implica *locamente*, pero no a la inversa, por lo que los primeros constituyen el núcleo y predicen la aparición de un determinado colocativo³. Asimismo, se conciben las colocaciones como conjuntos léxicos con un grado de fijación, lo que las distingue de las combinaciones libres; esto es, en el ámbito culinario se utiliza *diente de ajo* y *gajo de naranja* y no se aceptaría la combinación errónea de estos conjuntos léxicos, **diente de naranja*, o **gajo de ajo*. También se caracterizan por tener una transparencia semántica que no poseen los modismos o expresiones idiomáticas. No obstante, la transparencia semántica no siempre se observa en las colocaciones, por ejemplo, *dinero negro* o *tocar la guitarra*.

Las aportaciones realizadas por la Escuela soviética al campo de la fraseología son de suma importancia y también han ejercido una gran influencia en otras áreas de la lingüística

² No-composicional: cuando una unidad léxica compleja presenta un significado global diferente a la suma de sus partes individuales (Granger y Paquot, 2008: 4).

³ Ejemplos extraídos de Koike (2001: 44-60)

(Cowie, 1998). Una de las grandes contribuciones al estudio de la fraseología es la aportación del término *unidades fraseológicas* (UF) por Vinogradov (1947). Este autor muestra un profundo interés por las combinaciones de palabras; y afirma que las expresiones, los refranes, las citas, las locuciones, los proverbios y las combinaciones de palabras, aun siendo parcialmente composicionales, presentan restricciones de combinación en alguno de sus componentes. Se refiere con esta caracterización a lo que más adelante se denominó en la tradición anglosajona como *colocación* (Cowie, 1998). Estas combinaciones fijas pasan de una generación a otra y se reproducen en la práctica de la comunicación por tradición, diferenciándolas así de las combinaciones libres que se organizan en la lengua a voluntad del hablante. Asimismo, las combinaciones fijas se dividen en tres categorías: (1) uniones fraseológicas o modismos, (2) unidades fraseológicas y, finalmente, (3) combinaciones fraseológicas.

Los modismos son grupos de palabras indivisibles, fijas y semánticamente opacas o no composicionales. Por ejemplo, encontramos en español: *pañó de lágrimas, dar calabazas, ser harina de otro costal, dicho y hecho o de tal palo tal astilla*. Las unidades fraseológicas son expresiones que de forma directa o figurada expresan una enseñanza u opinión de tipo moral o práctico. Debido a su naturaleza se sitúan muy cerca de los modismos, la frontera entre ambas es borrosa y varía de acuerdo con la experiencia cultural y lingüística de cada hablante; esto se observa en la expresión *enterrar el hacha*, puesto que algunos hablantes podrían considerarla como puramente idiomática y otros, como una metáfora que no está totalmente fosilizada (Pérez Serrano, 2015: 41). La tercera categoría, las combinaciones fraseológicas, es la más interesante y la más compleja de delimitar o definir; en estas combinaciones, el significado de una palabra suele tener un significante dependiente, por ello, uno de sus componentes presenta un sentido literal y otro figurado. Estas son lo que se ha denominado posteriormente «colocaciones restringidas», en las que la acepción en la que utiliza uno de los elementos colocativos viene determinada por el otro elemento, como *correr un peligro, una suerte, aventuras* (Méndez Cedón, 2008: 120).

Autores como Igor Mel'čuk, Anthony Paul Cowie o Harald Burger son herederos de las ideas desarrolladas dentro del ámbito intelectual de la Escuela soviética y centran varios de sus estudios en la definición de colocación. Mel'čuk (2012) es el principal representante de la teoría sentido-texto en la que se otorga un valor principal al hablante frente al destinatario. Asimismo, busca unos principios generales que puedan aplicarse a todas las lenguas, por lo que defiende que las unidades léxicas se relacionan entre sí mediante un sentido semántico abstracto. Esto es, existen multitud de textos que pueden representarse con un solo sentido y a su vez, un único texto puede expresar varios sentidos. La teoría sentido-texto es, en palabras del autor, «rather a methodology than a theory», donde las relaciones entre el texto y el sentido de dicho texto se representan a través de funciones léxicas (FL) que asignan un peso semántico a cada elemento en una frase (2012: 87). En relación a las colocaciones, Mel'čuk (1998) explica que el factor más importante para analizarlas desde el punto de vista del significado es que los elementos no poseen el mismo valor semántico: el primer elemento apenas proporciona información, y todo el peso semántico se concentra en el segundo componente del conjunto léxico. Esta noción es muy similar a la idea de *node* y *collocate* expuesta por Sinclair (1991) en la cual uno de los componentes selecciona al otro, sin embargo, Mel'čuk (2012) explica que la relación semántica entre diferentes grupos léxicos puede ser la misma.

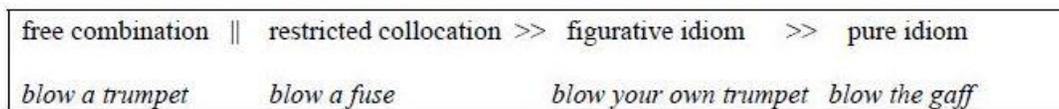
La teoría sentido-texto recoge unos 60 tipos de FL que permiten, entre otras cosas, la descripción de relaciones como la sinonimia (por ejemplo, comprar y adquirir son acciones idénticas), hiperonimia/hiponimia (un perro es un tipo de animal) u otras relaciones entre las unidades léxicas a un nivel oracional. Para ello, el autor distingue entre FL paradigmáticas y FL sintagmáticas. Las primeras hacen referencia a la relación semántica, esto es, al campo semántico de un grupo de palabras, y la segunda al principio de combinación. Por ejemplo, si nos basamos en esta teoría, colocaciones como *levantar un edificio*, *hacer un favor* o *formular una pregunta* entran en la misma función: Oper[N] o «realizar [algo]», pero *fumador empedernido* o *comprador compulsivo* se presentan mediante la función [N]Magn, siendo [N] la función léxica y Magn su carácter intenso. Ante este tipo de combinaciones, Benson, Benson y Ilson (1986) critican que si esta función explica conjuntos como *profundamente consternado*, del mismo modo podría generar combinaciones libres del tipo *muy consternado*. Mel'čuk (1998) argumenta que hay elementos universales que se asocian por defecto a una función, así ocurre con [N]Magn cuando el núcleo es un adjetivo.

Esta teoría enfatiza la variabilidad en la rigidez o *fixedness* de las FL y pone como ejemplo el contraste entre las expresiones en inglés *(to) pay attention* y *(to) give a look*. La primera se representa como Oper[attention] donde Oper siempre debe ser la palabra *pay*. Sin embargo, en la segunda Oper se expresa con el verbo *give* seguido de decenas de nombres (esto es, *give[N] a pull, a punch, a smile, a tug, a push, a kick, a strock, a kiss, a try*, etc.). No obstante, es posible caracterizar semánticamente la construcción resultante de [N], siendo esta «*give*[afecta a un objeto o se comunica con un ser] y voluntariamente se realiza una parte de la acción». Como resultado, encontramos FL con un grado muy bajo de fijación en las que los componentes pueden ser semánticamente transparentes y su concurrencia suficientemente previsible. El grado de rigidez es, por lo tanto, un parámetro independiente en los conjuntos léxicos y se define a través de propiedades tales como restricción de selección e irregularidad para combinarse (1998: 42).

Esta teoría, además de apoyar la idea de que las colocaciones están formadas por elementos que están a distinto nivel dentro del conjunto, esto es, uno de los componentes, el núcleo, selecciona al otro, el colocativo, también aporta ideas muy interesantes a nuestra definición. Por ejemplo, observamos que la característica generalizada de que las colocaciones son transparentes no siempre es real, por ejemplo, *cerrar la sesión* o *salir a bolsa*, por ello la importancia de la relación semántica entre elementos. La flexibilidad es otro elemento que hay que tener en cuenta, Wray (2002: 52) apunta que la mejor forma de entender la flexibilidad colocacional es por medio de la observación del «uso» y no del estricto «significado» de sus componentes.

Cowie (1998) también continúa la línea de investigación de los principales autores de Europa del Este y expone una distinción clara entre conjuntos que presentan la función sintáctica en el nivel de la frase o por debajo de este y las fórmulas que funcionan pragmáticamente como expresiones autónomas. Así también, crea un continuo donde muestra la complejidad del asunto. En un extremo, sitúa las combinaciones libres y en el otro, las locuciones o unidades léxicas complejas que presentan un alto grado de fijación. Amplía, de este modo, la línea de investigación de la Escuela soviética en donde se presta atención principalmente a los modismos o frases hechas.

Figura 1. Continuo fraseológico de A.P. Cowie (Extraído de Granger y Paquot, 2008: 34)



Como se observa en la figura 1, los conjuntos léxicos se subdividen en combinaciones libres, colocaciones restringidas, expresiones figuradas y modismos. Nos vamos a centrar en las tres últimas, ya que estas son las más interesantes para nuestro estudio. Por tanto, encontramos que estas tres categorías forman un continuo fraseológico, siendo las colocaciones restringidas los conjuntos léxicos más variables y los modismos, *pure idiom*, los más opacos y fijos. Los modismos tales como *estirar la pata* o *estar a dos velas* son semánticamente no-composicionales. Estos conjuntos léxicos también incluyen unidades oracionales como las fórmulas rutinarias como *buenos días* o *hasta pronto*, que funcionan pragmáticamente del mismo modo que los dichos o las frases hechas, en las que el hablante realiza un acto de habla y que tienen como objetivo organizar los mensajes e indicar una actitud como «¿sabes lo que quiero decir?», «¿vale?» o «¿entiendes?» (Cowie, 2001). Las expresiones figuradas tienen, como su nombre indica, un sentido figurado pero también presentan una interpretación literal, por ejemplo, en la carretera *hacer un cambio de sentido* y, como en los ejemplos anteriores, sus componentes no permiten sustitución. La categoría de colocaciones restringidas a menudo es designada simplemente como «colocaciones» e incluye combinaciones tales como *saltarse la ley*, *realizar una tarea* o *fuertes lluvias*. Estas se caracterizan por tener una colocatividad restringida basada en el sentido figurado de uno de sus elementos y por presentar un significado único en uno de sus elementos. Por ejemplo, el sentido figurado del verbo «cometer» en español limita su colocabilidad a una serie de sustantivos con connotaciones negativas como *error*, *adulterio*, *delito*, *infracción*, *falta*, *asesinato*, *fraude*, etc.⁴ Este autor también expone la dirección de la colocabilidad en la que el elemento con el sentido literal determina la elección del término con sentido figurado, así en el caso de los ejemplos anteriores, «error» determina el sentido del verbo «cometer». Del mismo modo, incluye combinaciones compuestas por verbo + sustantivo en las cuales el verbo está deslexicalizado o vacío de significado como *hacer un comentario* que también pueden expresarse con un solo verbo: «comentar». Alonso Ramos (2004) realiza un extenso estudio sobre este tipo de colocaciones en español, a las que denomina construcciones con verbo de apoyo (CVA).

En la definición que se presenta en este trabajo es de suma importancia la idea de continuo de lengua, ya que permite representar las unidades fraseológicas como una línea que va desde composicional a totalmente no-composicional con categorías intermedias. Este concepto también es aplicable a las mismas colocaciones y su transparencia semántica, porque nos permite hablar de diferentes grados de especialización semántica⁵ y no caer en la definición de falta de idiomatidad sin profundizar en la idea. Del mismo modo, el continuo de la lexicogramática permite clasificar patrones lexicogramaticales como más

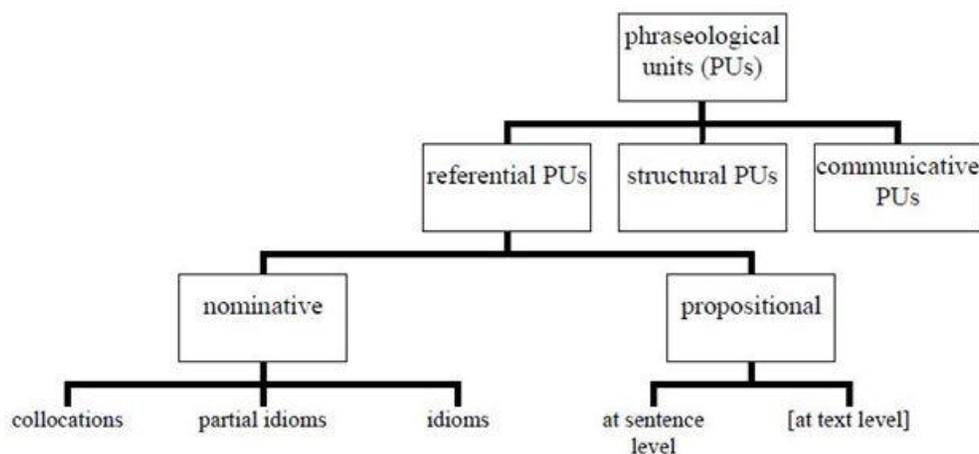
⁴ Ejemplos extraídos del diccionario en línea www.linguee.com.

⁵ Especialización semántica: cualquier cambio producido en el sentido primitivo de una unidad léxica (Koike, 2002).

léxicos o más gramaticales. Por tanto, el aprendizaje de lenguas implica el aprendizaje de construcciones, que son unidades asociadas a formas morfológicas, sintácticas y léxicas que poseen características semánticas, pragmáticas y funciones discursivas y, por ello, no dependen exclusivamente de la realización de categorías específicas de la lengua.

Burger (1998), al igual que Cowie o Mel'čuk, sigue la tradición de la Escuela soviética y considera que el léxico forma un sistema con sus leyes propias y relaciones internas, y en sus estudios marca la relación entre los factores internos y externos de la lengua. Por lo que respecta a las unidades fraseológicas que nos interesan, Burger (1998), al igual que los autores anteriores, las subdivide en tres tipos: expresiones idiomáticas o modismos, expresiones parciales y colocaciones. Sin embargo, estas tres clases de unidades son solo una subcategoría más dentro de las unidades fraseológicas que, como se muestra en la figura 2, se ramifican en diferentes categorías funcionales, creando una distinción entre unidades referenciales, unidades comunicativas y unidades estructurales.

Figura 2. Tipología de Burger (Extraído de Granger y Paquot, 2008: 36)



Un punto interesante de las investigaciones de Burger y útil en nuestra definición es su interpretación de la fijación de las unidades fraseológicas. Para este autor, la fijación es un concepto relativo y no es necesario que una combinación de palabras manifieste una estructura formal y léxica completamente fija para ser clasificada como fraseologismo. Aboga por una concepción flexible de la fraseología en lo que concierne a la fijación. Y aporta los conceptos de variación y modificación (Burger, 1998). La primera se refiere al hecho de que muchos fraseologismos presentan cierta variación léxica (*vivir en el quinto pino*, *vivir en el quinto infierno*). La siguiente hace referencia a la posibilidad de modificar creativamente un fraseologismo, especialmente de forma ocasional. Este proceso es muy habitual en ciertos ámbitos: en especial, en la publicidad y en algunos textos periodísticos como en los comentarios y artículos de opinión, *quien cría cuervos, tendrá cuervitos*. Los fines que persiguen con estos cambios son los de persuadir al lector, apoyar el hilo argumentativo o producir efectos cómicos o estilísticos (Corpas Pastor, 1996: 233-258). Esta flexibilidad también se observa en las variedades dialectales, por ejemplo, *hacer la cama* o *tender la cama*; *montar en bicicleta* o *andar en bicicleta*.

La interpretación ofrecida por estos autores se centra en la relación semántica entre los componentes de las colocaciones: entre el núcleo y el colocativo. Asimismo, estas son definidas sobre la base de lo que no son: no son combinaciones libres, porque presentan cierto grado de rigidez composicional; tampoco son expresiones idiomáticas, porque su significado se deduce del sentido literal de sus componentes. Aunque como hemos visto, este último concepto no es del todo cierto, como es el caso de *tocar la guitarra o perder los nervios*. Por ello, la idea del continuo toma fuerza a la hora de juzgar lo que es o no es una colocación. Durrant (2008) expone que uno de los principales problemas de estos enfoques es que dependen en gran medida de la interpretación del investigador a la hora de identificar los diferentes tipos de unidades fraseológicas, lo que hace que el análisis sea en cierto modo subjetivo y extremadamente laborioso: factible solo para los corpus de tamaño reducido. Asimismo, es de gran importancia el concepto de flexibilidad entre elementos y esto lo encontramos especialmente en las variedades dialectales.

3. ESTUDIOS DE CORPUS Y FRECUENCIA

Los estudios de corpus han aportado una información muy valiosa a la definición de lenguaje fraseológico y por ende, a la definición de colocación. Autores como J. R. Firth, Michael Halliday o John Sinclair iniciaron una corriente centrada en el análisis de conjuntos léxicos con una dura crítica al enfoque empleado por las escuelas tradicionales, donde el léxico se veía como una unidad aislada. Para este grupo de autores, el análisis probabilístico o basado en la frecuencia ha sido la principal herramienta metodológica, ya que el elemento fundamental dentro de la definición de colocación es la concurrencia del léxico; de este modo, las colocaciones se definen como combinaciones de palabras que aparecen en la lengua con una frecuencia significativa (Sinclair, 1991). Esto no quiere decir que la semántica quede totalmente fuera del ámbito de interés del análisis, ya que «the formalisation of contextual patterning of a given word or expression is assumed to be relevant to the identification of the meaning of that word or expression» (Granger y Paquot, 2008: 4). Hoy en día esta definición predomina dentro del análisis computacional, cuyo objetivo persigue la simulación de competencia comunicativa mediante el uso de programas informáticos y para ello, busca dar luz a las relaciones sintagmáticas (Tordera Yllescas, 2012). Se observa, de este modo, una preocupación por el contexto y las cadenas de palabras que desde el punto de vista de estos autores actúan como una sola unidad léxica y con un único significado.

Como se mencionó previamente, en nuestra definición hemos intentado aunar varias ideas extraídas de las teorías fraseológicas, como la importancia semántica entre el núcleo y el colocativo o la flexibilidad interna existente dentro de las colocaciones, pero también son importantes las corrientes centradas en la probabilidad y la frecuencia. La frecuencia es crucial para entender las colocaciones, por varios motivos. Primero, el uso establece que una colocación se imponga y sea tácitamente aceptada. Esto es, la frecuencia en la lengua es esencial para que un conjunto léxico sea adoptado por los hablantes, asimismo la frecuencia es crucial en el aprendizaje de una L2. Del mismo modo, la frecuencia de concurrencia entre las palabras que forman las colocaciones es de gran ayuda para detectarlas y hace que los hablantes tengan intuiciones con respecto a las colocaciones, a las que también podemos denominar combinaciones recurrentes de palabras. Es esta frecuencia

la que hace que al escuchar ciertas palabras puedas adelantarte a la siguiente; como es el caso de *atención*, que inmediatamente uno piensa en *llamar*, *llamar la atención* o *acaparar*, *acaparar la atención*; o si ignoramos en España lo hacemos olímpicamente, *ignorar olímpicamente*. Así observamos que la concurrencia léxica tiene un papel importante en la organización del lenguaje. A continuación, vamos a presentar las teorías más importantes de esta rama y explicaremos por qué nos interesan en nuestra definición.

J. R. Firth (1957) utiliza el término colocación en referencia a «frequent co-occurrences of lexical combinations» (1957: 196) y argumenta que el uso y significado de una palabra (*node*) se caracteriza por su colocación (*collocate*). Da como ejemplos *dark night*, *strong argument* o *heavy smoker* (1957: 197). El autor también distingue entre colocación y coligación; refiriéndose con el primer término a combinaciones léxicas como *tocar la guitarra* y con el segundo, a la fuerte afinidad de una palabra con una clase gramatical, por ejemplo, verbos como *agree*, *decide*, *expect* y otros de la misma clase rigen un complemento que forma cláusulas que comienzan con *that* o cláusulas de infinitivo con *to*. Sin embargo, otros verbos como *accept*, *explain* y *recognize* solo permiten como complemento cláusulas con *that*. Se entiende, por tanto, que el autor hace referencia a los complementos de régimen como *vivir con*, *acordarse de*, *casarse con*, etc. Sin embargo, en su obra nunca ofreció una definición clara de los parámetros que constituyen una colocación, ya que su enfoque es esencialmente semántico y hasta cierto punto estilístico. Firth (1957) estudió la frecuencia de las colocaciones para caracterizar el estilo de un autor y marcó la distinción entre colocaciones usuales: aquellas que presentan una frecuencia elevada y son producidas por el resto de hablantes de la lengua, y colocaciones inusuales: aquellas que representan la estilística de lo idiosincrático y personal. Para este autor la posibilidad de que una palabra se combine —colocabilidad— con otra es uno de los diferentes análisis que se pueden realizar para determinar la acepción de dicha palabra en contextos específicos.

Alonso Ramos (1994) ha criticado la obra de Firth debido a la estrecha relación que el autor establece entre el significado de una palabra y el conjunto fraseológico de su significado. Alonso Ramos (1994: 11) comenta que «el significado de ‘rubio’ [no] es su colocabilidad con *pele* pero sí que el hecho de que estos dos lexemas concurren está determinado por el significado de ‘rubio’: en su definición debe haber una mención a *pele*». No obstante, Firth aportó un sentido nuevo al significado global de las palabras y sus discípulos Halliday (1961, 1966) y Sinclair (1991, 1996) continuaron su línea de investigación en la que trataron las colocaciones esencialmente en términos de probabilidad, razón por la cual idearon fórmulas estadísticas para explicar tanto la frecuencia como su modo de aparición en los textos.

Halliday (1961: 276) propuso como definición de colocación la asociación sintagmática de ítems léxicos textualmente cuantificables y destacó la probabilidad de que esta asociación ocurra en n intervalos, o sea en una distancia de n unidades léxicas. Posteriormente, Halliday (1966: 159) propone que el criterio fundamental para definir una colocación debe ser la probabilidad de que dos palabras aparezcan juntas y para ello ofrece el ejemplo de los adjetivos *strong* y *powerful*, los cuales pueden ser yuxtapuestos al vocablo *argument*: *strong argument* o *powerful argument*. Sin embargo, estos adjetivos no siempre son intercambiables: *strong*, pero no *powerful*, puede combinarse con *tea*, y *powerful*, pero no

strong, puede combinarse con *car* (1966: 150). Para este autor las unidades léxicas que muestran cierto grado de probabilidad en su comportamiento colocacional son asignadas al mismo conjunto. De este modo, *strong* y *powerful* son agrupadas en el mismo conjunto ya que muestran una similar potencialidad de aparición en el entorno de *argument*, pero pertenecen a un conjunto diferente en relación con otros lexemas. Ante una colocación dada, un HN no la acepta o la rechaza, sino que «he will react to something as more acceptable or less acceptable on a scale of acceptability» (1966: 159). Es importante resaltar el problema que presenta Halliday en relación con la intercambiabilidad de sinónimos, pues algunas palabras, llamémoslas *x*, parecen combinar bien con otras, en este caso *y*; sin embargo, los sinónimos de *x*, que deberían sonar igual de bien al combinarlos con *y*, crean combinaciones que no suenan naturales a oídos de un HN, por ejemplo *entablar una conversación* o *una amistad*, pero **entablar un encuentro* o *una cita*.

Sinclair (1966) destaca la necesidad de estudiar la concurrencia de unidades léxicas con un método estadístico e introduce varios términos importantes, tales como núcleo (*node*), colocativo (*collocate*) y distancia (*span*). Este autor sostiene que la detección de una colocación puede efectuarse estadísticamente teniendo en cuenta no solo la distancia que media entre el núcleo y el colocativo, sino también la frecuencia de cada uno de ellos y el número de veces que aparecen juntos. Para Sinclair (1991: 170) la colocación es «the co-occurrence of two or more words within a short space of each other in a text. The natural measure of proximity is a maximum of four words». Recurre únicamente al criterio formal (la frecuencia de concurrencia y la distancia colocacional de cuatro palabras), sin entrar en detalles acerca de los aspectos semánticos o sintácticos. Asimismo, resalta que la asociación colocacional habitualmente es asimétrica: la probabilidad de que una palabra *A* esté asociada con una palabra *B* es diferente a la probabilidad de que *B* esté asociada con *A*. Una de las principales diferencias entre Sinclair y Firth con respecto a la definición de colocación, aparece en una entrevista realizada por Wolfgang Teubert publicada en el prefacio de *English Collocation Studies. The OSTI Report* (2004). En ella Sinclair afirma que para Firth «[o]ne of the meanings of *night* is its collocability with *dark*, and of *dark*, of course, collocation with *night*» (Firth, 1957: 196), para Sinclair «[t]he phrase *dark night* has its own meaning» (Sinclair, 2004: xxi).

Este análisis formal tiene la ventaja de ser puramente objetivo, pero al mismo tiempo plantea otro tipo de problemas; por ejemplo, la concurrencia frecuente de dos unidades léxicas no supone siempre la presencia de una colocación, por ejemplo, *compartir gastos* o *hacer un pastel*; además, tal y como señala Alonso Ramos (1994: 14), no todas las colocaciones aparecen a una distancia menor de cuatro palabras. Asimismo, Wray (2002: 27) aduce que este análisis estadístico depende en gran medida de la naturaleza de los corpus (oral, académico, informal, etc.), ya que en algunos de ellos podremos encontrar ciertas colocaciones, pero probablemente las que son típicas de un ámbito específico no aparecerán, por ejemplo, en textos jurídicos aparecerán con una alta frecuencia colocaciones como *dictar sentencia*, *informe pericial*, *comisión rogatoria*, *antecedentes penales*, sin embargo, en el habla cotidiana no son frecuentes. Esto crea una complicación añadida a la hora de estudiar la distribución real de ciertas fórmulas léxicas. Al mismo tiempo, un análisis basado en una metodología puramente probabilística no es capaz de dilucidar el comienzo o el fin de una combinación de palabras, por ejemplo, la cadena *thank you* no siempre aparece sola, también se manifiesta como *thank you very much* o *thank you*

very much indeed. Por tanto, Wray (2002: 28) se pregunta si se deben contar estas combinaciones como una sola o como si fueran diferentes conjuntos léxicos. Actualmente, y sin la ayuda de un análisis manual, no hay programa estadístico que pueda aclarar este problema metodológico. No obstante, el uso de corpus de gran extensión y la extracción automática de conjuntos léxicos mediante programas informáticos ha originado estudios como el *Collins Birmingham University International Language Database* (COBUILD) con resultados como el diccionario monolingüe *COBUILD English Language Dictionary* (1987). Este diccionario marcó un hito en la historia de la lexicografía ya que desde entonces el corpus ha servido como elemento principal para la descripción y como método de análisis del uso real de la lengua. En la parte introductoria del *COBUILD* se puede encontrar la siguiente entrada: «the title [of a word] not only explains what English words mean, but also provides information on how the words actually work».

Podemos colegir que la frecuencia es importante para reconocer ciertos elementos de las colocaciones como la relación entre las palabras que componen un conjunto léxico y la relación que existe con el texto, pero al mismo tiempo solo muestra una parte de dicha relación.

3.1 Tipos de análisis de corpus

Para entender mejor las diferentes metodologías e intentar acercarnos a una definición algo más clara de colocación, en esta sección se presenta un estudio que, siguiendo la línea probabilística, trabaja las colocaciones a través del análisis de corpus.

Evert (2009) presenta un profundo estudio de los múltiples análisis de colocaciones realizados a través de los estudios de corpus, en los cuales los conjuntos léxicos se entienden como realidades empíricas; esto es, son combinaciones observables y cuantificables en un grupo de textos. Este autor resalta dos medidas principales de análisis: (1) los análisis de asociación —medida cuantitativa para observar la relación entre palabras—; (2) la definición formal de concurrencia y el cálculo apropiado de datos de frecuencia de dicha concurrencia. Aunque la interpretación de las puntuaciones de asociación parece una tarea sencilla, —a mayor puntuación, mayor atracción entre palabras—, estas asociaciones pueden utilizarse de diferente manera para identificar las colocaciones. La primera diferencia la encontramos en la colocatividad de los términos, ya que la relación entre palabras puede tratarse como un fenómeno categórico, lo que pretende identificar «colocaciones verdaderas» que según afirma Hoey (1991: 7) «the relationship a lexical item has with items that appear with greater than random probability in its (textual) context», o como fenómeno continuo, en el que se clasifican pares de palabras en una escala de fuerza colocacional. El segundo análisis se refiere a la agrupación de palabras que componen una colocación, las cuales son vistas como unidades independientes. Este análisis núcleo-colocativo (*node-collocate*) se centra en la compañía del núcleo y observa su probabilidad de combinación con su colocativo. Se estudia, de este modo, «the company it keeps» (Firth, 1957: 11). Estos dos análisis son independientes entre sí, aunque la mayoría de las ocasiones el análisis *node-collocate* se combina con una metodología de asociación. En esta sección vamos a centrarnos en la segunda medida porque es esta la que usamos en nuestra definición.

Los métodos cuantitativos de extracción de colocaciones más utilizados son (1) t-score, (2) información mutua (IM) y (3) z-score. No obstante, como González Fernández y Schmitt (2015: 96) apuntan: «it is not clear which of these [MI-score and t-score] (or other) measures is the best to use in research, and to date, the selection of one or another seems to be somewhat arbitrary.» Por este motivo se ha incluido la fórmula (4) LogDice en esta descripción:

1) La fórmula matemática de un análisis t-score, como muestra Evert (2005: 82-83) no tiene una base muy transparente, por lo tanto, no es posible establecer puntos de corte válidos que ayuden a interpretar el puntaje de la relación entre las palabras que forman una colocación. Asimismo, los resultados habitualmente muestran que tanto los conjuntos léxicos como las palabras aisladas aparecen casi con la misma frecuencia, como es el caso de *producir un cambio* (*producir*: rango 239, frecuencia 682,939; *cambio*: rango 227, frecuencia 727,249 y *juntas* aparecen 51.097 veces)⁶. El problema principal de la fórmula t-score es que no utiliza una escala estándar y, por lo tanto, no se puede usar para comparar colocaciones en diferentes corpus (Hunston, 2002).

2) Información mutua (IM) usa una fórmula logarítmica para expresar la relación entre la frecuencia de la colocación y la frecuencia de la coocurrencia aleatoria de las dos palabras que forman el conjunto léxico. De este modo, se compara el número de veces que una colocación aparece dentro del corpus con el número de veces que aparecen las palabras que la componen por separado. Se ha observado que los resultados de IM favorecen las combinaciones de palabras de baja frecuencia que presentan una alta asociación entre ellas y, por el contrario, t-score se ha señalado como la fórmula para medir colocaciones de alta frecuencia (Granger y Bestgen, 2014). El resultado del análisis de IM es muy desigual, pues las palabras que se obtienen de él se perfilan como conjunto léxico en muchas ocasiones, pero rara vez de manera independiente. Por ejemplo, *placa tectónica*: en el *Corpus del Español* (Davies, 2002-) *placa* se sitúa en un rango de alta frecuencia (rango 2. 900, frecuencia 48. 385), sin embargo, *tectónica* está en un rango de baja frecuencia (rango 27. 965, frecuencia 544) y *juntas* aparecen en 49 ocasiones con una IM del 8,53. Otros ejemplos que encontramos son *guiñar un ojo* con un IM de 10,44 y *ojos almendrados* con un IM de 6,75. En relación a la escala que utiliza, el puntaje IM es un índice normalizado. Este índice es comparable a través de diferentes corpus, el problema es que opera en una escala que no tiene un mínimo o máximo teórico, es decir, no presenta un rango de valores específico.

3) Z-score compara la frecuencia entre un núcleo y sus colocativos; evalúa la diferencia entre estos valores por medio de una desviación estándar. En cuanto a los procedimientos, la puntuación z-score es bastante similar a la t-score; sin embargo, los resultados son similares a los obtenidos mediante un análisis de IM.

Estas fórmulas no solo analizan la frecuencia de los conjuntos léxicos, sino que también cuantifican la frecuencia marginal o en relación al texto. Evert (2009: 58) critica del análisis z-score la medida de aproximación utilizada en su derivación matemática, ya que es válida solo para léxico de alta frecuencia. Por otro lado, el análisis IM se centra en las

⁶ Datos extraídos de El corpus del español de Mark Davis (2002-) <https://www.corpusdelespanol.org/>

colocaciones más idiosincrásicas, como términos técnicos o frases más composicionales, por el contrario t-score cuantifica las combinaciones léxicas más frecuentes

4) LogDice es la fórmula estadística que utiliza el programa SketchEngine para identificar la frecuencia de las colocaciones. No obstante, todavía no se ha popularizado en el estudio de corpus. Este programa calcula la frecuencia de las palabras que componen la colocación a través de la media armónica de dos proporciones que expresan la tendencia de coocurrencia de dos palabras dentro del corpus. LogDice es una medida estandarizada que opera dentro una escala con un valor máximo de 14, lo que permite comparar los resultados a través de diferentes corpus. Por lo tanto, LogDice presenta características muy positivas para un análisis de corpus.

Este sistema de detección automática a través de programas informáticos y fórmulas estadísticas ha aportado notoriedad al estudio del léxico marcando un antes y un después en este campo. Como se mencionó previamente, el vocabulario ha pasado de estudiarse como una serie de elementos aislados a entenderse como una fórmula del lenguaje enmarcada en un contexto. No obstante, también ha recibido críticas. Primero por la necesidad de trabajar con corpus de gran tamaño, que además deben tener un formato específico para ser analizados con un *software*, y segundo porque los resultados no muestran la totalidad del asunto, ya que no reflejan las relaciones sintáctico-semánticas. Corpas Pastor (2001) hace hincapié en el hecho de que hay colocaciones muy frecuentes que no presentan un grado de estabilidad suficiente para ser consideradas colocaciones. Sin embargo, como afirma Wray (2002) también nos encontramos el caso contrario: colocaciones muy estables cuyos colocativos son palabras poco frecuentes, por lo que solo aparecerán representadas en ciertos corpus; por ejemplo, combinaciones específicas referentes a un área de estudio concreta, como *capa de ozono*, *bolsa de valores*, *renta fija/variable*. Por lo que se refiere al tipo de corpus, Bosque (2004) añade que entender la naturaleza de los datos también es de gran importancia ya que uno puede tropezar con el «estilo del autor».

4. LAS COLOCACIONES EN LAS FUENTES ESPAÑOLAS

Seco (1978: 17) introduce el término *colocación* en el marco de la lexicografía española para denotar lo que más tarde se denominará «contorno» o elementos contextuales. El autor utiliza el término en la definición que debe imperar en los diccionarios para los adjetivos, (los colocativos del sustantivo) y bajo la distinción entre «primer enunciado» y «segundo enunciado». De este modo, se presenta el término *colocación* no solo relacionado directamente con el de contorno, sino que no se delimita de forma clara. No es hasta la década de los noventa cuando varios investigadores retoman el estudio de las colocaciones en español.

Írsula (1992) trabaja con colocaciones del tipo sustantivo-verbo; dos años más tarde Írsula (1994) presenta lo que él considera las cuatro características principales de colocaciones: (1) la denotación: se refiere a la relación existente entre las expresiones léxicas y las estructuras, cuya función es apoyar la percepción e imaginación de la realidad circundante. (2) El contexto semántico-gramatical: es el conjunto de restricciones sintáctico-semánticas que cada lengua posee. Por ejemplo, en español se usa el verbo *comer* tanto para personas como para animales, sin embargo, las plantas *se alimentan*. (3) Otra característica cardinal

son las restricciones impuestas por el uso tradicional, lo que genera que estas combinaciones formen parte de la norma lingüística social. (4) Finalmente, el autor analiza el contexto situacional y comunicativo, el cual se relaciona con el empleo del material léxico que corresponde con la situación comunicativa y encontramos, por tanto, los aspectos diatópicos, diastráticos y diafásicos (1994: 278-279). En suma, el término *colocación* es para este autor: «combinaciones de dos o más unidades léxicas que resulten seleccionadas en los diferentes actos del habla, teniendo en cuenta los cuatro aspectos mencionados» (1994: 282).

Alonso Ramos (1993) analiza el léxico sobre la base de la concurrencia y la teoría sentido-texto de Me'lčuk (1981). A pesar de que su trabajo no se centra en las colocaciones, contiene un gran inventario de ellas. Dos años más tarde, Alonso Ramos (1994) realiza un estudio específico de colocaciones en el que las define como «combinación de dos lexemas A y B, de tal forma que su significado es la suma de los significados de los lexemas constituyentes y su significado incluye el significado del lexema A y un significado C que es:

- a) Bien 'C' 'B' y
 - i. 'C' es vacío: el lexema B es un auxiliar usado para sostener una configuración sintáctica (*dar un paseo*)
 - ii. 'C' no es vacío pero el lexema B expresa 'C' solo en combinaciones con 'A' o con otros pocos lexemas similares (*odio mortal, interés vivo*)

- b) Bien 'C' = 'B' y
 - i. El lexema B es seleccionado restringidamente en combinaciones con A. No puede ser reemplazado por otro posible sinónimo (*café fuerte/potente**)
 - ii. 'C' incluye el sentido 'A' (*pelo rubio, vino seco, nariz aguileña*)» (pp. 25-26).

La representación formal de una colocación es la que aparece en la figura 3.

Figura 3. Representación formal de una colocación (Extraído de Alonso Ramos, 1994: 26)

$$\frac{'A'}{/A/} + \frac{'B'}{/B/} = \frac{'A + C'}{/A + B/} = AB \quad \left| \begin{array}{l} 'C' = \emptyset \\ 'B' = 'C' \text{ sólo con A} \\ B \text{ está seleccionado restringidamente} \\ 'C' \supset 'A' \end{array} \right.$$

Corpas Pastor (1996) define las colocaciones como unidades fraseológicas formadas por dos lexemas que presentan relación sintáctica, pero que no constituyen por sí mismas actos de habla ni enunciados. Asimismo, debido a su fijación en la norma, presentan restricciones de combinación establecidas por el uso, generalmente de base semántica. Se observa, por lo tanto, que el término *colocación* se vincula con la noción de solidaridad léxica presentada por Coşeriu (1977) en la que la solidaridad es una relación de sentido único, ya que la implicación no es recíproca; se trata de una relación sintagmática en la que existe una clase *determinante* y una clase *determinada*.

De lo dicho previamente, se deduce que el estudio de las colocaciones en español ha seguido hasta no hace mucho una metodología puramente fraseológica, influenciada por autores de Europa del Este; por tanto, el contexto y el significado de cada componente es esencial no solo para definir las, sino también para identificarlas.

Castillo Carballo (1998) realiza una exhaustiva compilación de los estudios realizados hasta la fecha, la mayoría de ellos de índole teórica. Por otra parte, crea un cuadro (tabla 1) de clasificación sobre la base del tipo de léxico que componen las colocaciones:

Tabla 1. Tipos de colocaciones (Castillo Carballo, 1998: 54)

Tipos	Estructuras categóricas
Tipo 1	sustantivo (sujeto) + verbo <i>registrarse un incendio, desatarse una polémica</i>
Tipo 2	verbo + sustantivo (objeto) (también verbo + prep. + sustantivo) <i>asestar un golpe, interponer un recurso, poner en cuestión</i>
Tipo 3	sustantivo + adjetivo (también sustantivo + sustantivo) <i>ruido infernal, conducta intachable, paquete bomba</i>
Tipo 4	sustantivo + prep. + sustantivo <i>ronda de negociaciones, ciclo de confesiones</i>
Tipo 5	verbo + adverbio <i>rechazar categóricamente, llorar amargamente</i>
Tipo 6	adverbio + adjetivo <i>visiblemente afectado, altamente fiable</i>

Realmente no es hasta Koike (2001) cuando las colocaciones en español se estudian mediante un profundo análisis de corpus con el que amplía el concepto y afirma que «las solidaridades léxicas unilaterales no son colocaciones porque sus constituyentes no suelen coaparecer en el discurso» (2001: 18). Al mismo tiempo, introduce el concepto de ‘entorno’ (2001: 63): *contraer una deuda* está constituida por una base *deuda*, un colcativo *contraer* y un entorno, en este caso el sujeto animado. El autor propone su definición formal y semántica de las colocaciones léxicas en función de (1) la frecuencia en la concurrencia de las unidades; (2) las restricciones combinatorias; (3) la composicionalidad formal, esto es, la flexibilidad interna como la modificación adjetival o el cambio de categoría gramatical; (4) el vínculo de los lexemas; (5) la relación típica entre los componentes y (6) la precisión semántica. Asimismo, Koike propone algunos cambios en cuanto al tipo de colocaciones. Primero, Koike (2001: 55) discute que las colocaciones adverbio + adjetivo permiten la división entre «colocaciones complejas», aquellas compuestas por una unidad léxica simple y una locución, y «colocaciones simples», formadas por dos unidades léxicas simples. Después, añade la colocación del tipo verbo+ adjetivo: *salir indemne, resultar ileso*. La tabla 2 muestra una explicación más amplia y ejemplos de estas nuevas categorías.

Al contrario que Castillo Carballo (1998), Koike (2001) expone que colocaciones como *deseo ardiente* o *desear ardientemente* representan la misma colocación, puesto que es el vínculo entre lexemas y no entre unidades léxicas lo que las caracteriza. Además, los componentes de las colocaciones sustantivo + verbo guardan, para este autor, una relación semántica tipificada. Por ejemplo, el sustantivo *guitarra* solo puede establecer una relación

típica como instrumento musical, por eso *tocar* o *afinar la guitarra* son ejemplos de colocaciones, mientras que *limpiar la guitarra* es una combinación libre. No obstante, como otros autores han recalado previamente (Cowie, 1998; Wray, 2002) una de las características que hace que las colocaciones sean tan difíciles de delimitar es su variabilidad, lo que hace que esa relación de tipicidad sea difícil de valorar. Las construcciones con verbo de apoyo son un buen ejemplo de esto, afirmar que hay una relación típica entre el sustantivo y el verbo en *dar una clase* es problemático porque el colocativo verbal se ha vaciado parcialmente de su significado. Asimismo, Koike (2001: 223) defiende que las colocaciones presentan «una precisión semántica o un concepto inconfundible» frente a las expresiones idiomáticas. Pérez Serrano (2015) discute que este criterio es difícil de valorar puesto que las colocaciones no siempre designan un concepto preciso; la colocación *cometer un error* y la expresión idiomática *meter la pata* pueden significar lo mismo en numerosos contextos. Por último, Koike (2001) explica que las colocaciones se diferencian de las combinaciones libres a través de las restricciones combinatorias impuestas por el uso y afirma que las colocaciones léxicas habituales son las de tipo sustantivo + adjetivo y sustantivo + verbo y las menos habituales son las colocaciones sustantivo + sustantivo (2001: 42).

Blanco Escoda (2000) y Alonso Ramos (2004) estudian ampliamente un tipo de colocaciones verbo + sustantivo, que se denominan construcciones con verbo de apoyo (CVA) y están compuestas por verbos vacíos de significado léxico y sustantivos que denotan eventos. Estas son las que Cowie (1998) caracteriza como unidades con un verbo deslexicalizado, como *dar una explicación*, *hacer un análisis* o *tener dudas*. Blanco Escoda presenta las características más formales de estas construcciones, mientras que, desde la perspectiva de la teoría sentido-texto, la autora española afirma que el núcleo es el sustantivo. Es este, por tanto, el que selecciona al colocativo o al verbo, pero la selección de este verbo no es puramente arbitraria, de tal forma *dar* o *hacer* se combinan con nombres que denotan acciones y *tener* con nombres que denotan estados: *dar una explicación* frente a *tener frío*. Por el contrario, De Miguel (2008: 570) afirma que en la formación de las CVA se produce un proceso de concordancia de los rasgos léxicos del sustantivo y del verbo, indicando así que el verbo posee rasgos léxicos y que también tiene significado. Alonso Ramos (2012: 16) aduce estos argumentos semánticos con ejemplos que los contradicen para concluir que «no negamos que haya tendencias pero sí verdaderas reglas que permitan combinar el verbo y el nombre, basándose exclusivamente en correlaciones semánticas» y «que las distinciones se hacen tan complejas y tan hiperespecificadas que no pueden ser útiles [...] para el aprendizaje de una lengua».

Hasta aquí hemos ofrecido una amplia noción de lo expuesto a lo largo de los últimos años acerca de las colocaciones en la lengua española. Estos estudios han seguido principalmente una metodología fraseológica, lo que aporta una valiosa información a la definición de índole cualitativa. De igual modo, se puede concluir que existe una falta de estudios basados en una metodología de corpus; Bosque (2004), Calderan Campos (1994) o Koike (2001) son algunos de los autores que realizan investigaciones de estas características. Los dos primeros autores a la creación de diccionarios monolingües y el último en relación a la adquisición del español como lengua extranjera (ELE).

5. ESTUDIOS EN EL AULA DE ELE

En relación con la adquisición y aprendizaje de ELE, Higuera García (2006) presenta un acercamiento a la enseñanza de este tipo de unidades léxicas dentro del aula. La autora indaga en el concepto de colocación y revisa tanto sus características como las partes que lo integran. Sin embargo, Higuera García (2006) no deja claro cuáles son las semejanzas y diferencias con otros tipos de unidades léxicas, tales como las combinaciones libres o los modismos, aunque sí opina que las colocaciones deben enseñarse de manera explícita al alumno de ELE y facilitar de este modo la creación de redes de significados que ayuden en el aprendizaje del léxico. La parte más amplia del estudio, por tanto, la constituye una tipología de actividades.

Pérez Serrano (2015) da un paso más en el estudio aplicado de las colocaciones y, además de ofrecer una batería de actividades, pone a prueba varias metodologías de enseñanza para dilucidar hasta qué punto las prácticas propuestas por un enfoque léxico, como la captación de las colocaciones en el *input* y las actividades de tratamiento explícito, tienen un efecto significativo en el aprendizaje de estas unidades. Los resultados revelan que la captación de las formas meta en el *input* es una metodología efectiva que favorece el proceso de aprendizaje. No obstante, este tipo de rutinas no genera diferencias en la adquisición de las colocaciones de tipo CVA o construcciones con verbo deslexicalizado. Por otra parte, la realización de actividades explícitas tiene una incidencia positiva en el desarrollo de habilidades para el reconocimiento activo de las colocaciones. Al igual que en el caso anterior, la realización de actividades no genera diferencias en el aprendizaje de las CVA.

Jensen (2017) realiza un estudio centrado en el reconocimiento de las colocaciones y su enseñanza explícita en el aula de ELE con alumnos de nivel avanzado. Su estudio muestra que, aunque las colocaciones sean transparentes, no siempre son reconocidas por los estudiantes de ELE. Las colocaciones formadas por una palabra polisémica, por ejemplo, *hace buen tiempo* o *tomar partido*, pueden ser malinterpretadas por los alumnos por su limitada profundidad léxica. Esta falta de conocimiento del significado implica que los aprendientes no conocerán los significados extendidos de algunos de los constituyentes que integran la colocación. Por lo tanto, la enseñanza explícita de colocaciones se hace necesaria dentro del aula.

6. LAS COLOCACIONES Y LA PROSODIA

Los estudios prosódicos, además de esclarecer parte del procesamiento del lenguaje (cfr. (Bybee, 2002; Bybee y Eddington, 2006; Levelt, 1999), también han ayudado a la definición de lenguaje formulaico y por ende a la de colocación.

Estudios empíricos como los de Bybee y Eddington (2006) muestran cómo las palabras o frases de alta frecuencia presentan un ritmo más rápido que las secuencias de baja o media frecuencia. Asimismo, los conjuntos de alta frecuencia muestran una mayor reducción de sonidos, por ejemplo elisión de la *t/d* en inglés americano, o de la *d* en contextos intervocálicos en el español. Sin embargo, los estudios centrados en la prosodia describen el lenguaje formulaico no solo como conjuntos de alta frecuencia, sino también como unidades rutinarias o fijas con un significado convencional que presenta un uso

especializado. Por todo ello se ha considerado que «such expressions differ from novel, newly created language, whereby a speaker creates a new utterance based on individual words and morpho-syntactic rules» (Hallin y van Lancker Sidits, 2015: 2).

Lin (2010) enumera y explica los elementos fijos que se encuentran en dichos conjuntos léxicos:

- Rigidez en el lugar del acento prosódico. Para ejemplificar esta característica expone los siguientes ejemplos:

- «*To have a CHIP on one's shoulder* compared with *To have a BEE on one's shoulder*».
- «*POUR down*» (idiomatic) in compared with *pour Down* (literal)»

- Entonación: las colocaciones se presentan como unidades de un solo tono con un contorno ininterrumpido.

- Tempo: se refiere a la resistencia o a la falta de fluidez interna, sin vacilación o titubeo y pocas pausas.

Estos elementos conforman lo que se ha denominado «phonological coherence» y generan unidades producidas con fluidez sin ruptura de entonación y sin vacilación.

7. DEFINICIÓN DE COLOCACIÓN

Siguiendo la clasificación de Castillo Carballo (1998), pero ampliada por las nociones semánticas que aporta Corpas Pastor (1996), Pérez Serrano (2015: 37-39) elabora un cuadro con las colocaciones sobre la base del tipo de palabras que la componen. La TABLA 2 muestra este cuadro, pero con los nuevos registros presentados por Koike (2001). El núcleo aparece en negrita:

Tabla 2. Tipo de colocaciones (Ampliación de la tabla presentada por Pérez Serrano, 2015: 37-39)

Tipos	Estructuras categóricas
Tipo 1	<p>sustantivo + verbo: <i>estallar una guerra, soplar el viento, palpitar el corazón</i> En este tipo de colocaciones el verbo denota una acción característica de la persona o cosa designada por el sustantivo. En cuanto a la relación semántica entre estos componentes, el verbo expresa «realizarse» o «llevarse a cabo» (<i>confirmarse una sospecha</i>). También se dan otro tipo de relaciones como «empezar a sentir» (<i>entrarle miedo</i>) o «estropearse» (<i>cortarse la leche</i>). Asimismo, se incluyen construcciones pronominales impersonales como <i>declararse un incendio</i> o <i>desatarse una polémica</i>.</p>

<p>Tipo 2</p>	<p>verbo + sustantivo (también verbo + prep. + sustantivo): <i>rescindir un contrato, interponer un recurso, implorar misericordia</i> Corpas Pastor establece dentro de este grupo tres tipos de colocaciones que se ordenan en cuanto a la variabilidad del patrón combinatorio. En un extremo estarían las colocaciones que permiten cierta variabilidad, como <i>desempeñar un cargo/un papel/una función</i> hasta las que no lo hacen, como <i>conciliar el sueño</i>. En un estadio intermedio estarían todas las construcciones con verbo de apoyo del tipo <i>tomar una decisión, dar comienzo o tener repercusión</i>. Las relaciones semánticas que se establecen son numerosas. Corpas Pastor, basándose en Benson <i>et al.</i> (1986: XXV-XXVI), distingue en primer lugar dos, las de creación, en las que el verbo puede expresar «causar» (<i>crear una dificultad</i>) o «realizar» (<i>desempeñar un cargo</i>). Frente a estas, estarían las que expresan cancelación o liquidación, como <i>rescindir un contrato</i>. Pero hay muchas más relaciones semánticas que se dan en este numeroso grupo de colocaciones: <i>mantener una conversación, perder los nervios</i>, etc.</p>
<p>Tipo 3</p>	<p>sustantivo + adjetivo (también adjetivo + sustantivo): <i>discusión acalorada, santa paciencia, ojos almendrados</i> Estas colocaciones también presentan distintos grados de restricción, desde <i>oído/vista/olfato fino</i> (menos restringida) hasta <i>dinero negro</i> (más restringida). Corpas Pastor afirma que la relación semántica más frecuente es la de intensificación: el adjetivo suele intensificar al núcleo, aunque también pueden expresar «bueno» o «tal y como debe ser», como es el caso de <i>cuchillo afilado</i>.</p>
<p>Tipo 4</p>	<p>sustantivo + preposición + sustantivo: <i>ronda de negociaciones, ciclo de confesiones, banco de peces</i> Otros ejemplos de esta categoría son <i>rebanada de pan, pastilla de jabón, tableta de chocolate</i>, que ilustran una relación semántica de «porción, unidad de». Asimismo, están las que expresan «conjunto de», como <i>rebaño de ovejas o ciclo de conferencias</i>. Se trata de colocaciones con un alto grado de restricción, donde el segundo nombre suele estar contenido en la definición lexicográfica del primero.</p>
<p>Tipo 5</p>	<p>verbo + adverbio: <i>rechazar categóricamente, llorar amargamente</i> Los adverbios que forman parte de colocaciones con verbos suelen expresar también intensificación. Son <i>rogar encarecidamente, felicitar efusivamente, desear fervientemente, fracasar estrepitosamente, prohibir terminantemente</i>.</p>
<p>Tipo 6</p>	<p>adverbio + adjetivo: <i>terminantemente prohibido, visiblemente afectado, altamente fiable</i> Nuevamente la relación semántica «casi exclusiva» (Corpas 1996: 76) que se da entre adjetivos y adverbios es la de intensificación. Ejemplos de ello son <i>diametralmente opuesto, rematadamente loco, locamente enamorado</i>,</p>

	<p><i>firmemente convencido.</i></p> <p>A esta clasificación, Koike (2001) añade la novedad (Higueras, 2001) de distinguir las colocaciones formadas por unidades léxicas simples y las formadas por unidades léxicas complejas. Estas últimas están formadas por una unidad fraseológica y una unidad léxica simple, como ocurre en <i>dar el santo y seña</i>, donde las dos unidades serían <i>dar + el santo y seña</i>:</p> <ul style="list-style-type: none"> • Verbo + locución nominal: <i>dar el santo y seña.</i> • Locución verbal + nombre: <i>llevar a cabo un proyecto.</i> • Nombre + locución adjetival: <i>tener una salud de hierro.</i> • Verbo + locución adverbial: <i>reírse a mandíbula batiente.</i> • Locución adverbial + adjetivo: <i>fuerte como un roble.</i>
Tipo 7	<p>verbo + adjetivo</p> <p><i>resulta/salir ileso, andar liado, estar enojado</i></p> <p>Koike (2001:55) arguye que no hay muchas colocaciones de este tipo. Sin embargo, existen adjetivos que se combinan con una alta frecuencia con determinados verbos, por ejemplo <i>salir</i> o <i>resultar</i>, y añade que estos conjuntos léxicos «presentan las características de la colocación [...] (la coocurrencia frecuente de dos unidades léxicas, las restricciones combinatorias impuestas por el uso tradicional, etc.)»</p>

De lo dicho hasta ahora podemos colegir que las colocaciones son combinaciones léxicas determinadas por el uso de una comunidad de hablantes que se caracterizan por una concurrencia de elementos con cierto grado de fijación; esto motiva que este tipo de construcciones se cataloguen como unidades semiidiomáticas. Además, se puede observar que la frecuencia es una característica muy importante en la definición de colocación. Sin embargo, no es la única, ya que, por ejemplo, en cada conjunto léxico también hay un fuerte grado de composicionalidad semántica y en consecuencia una relación con el contexto. Se observa que las colocaciones son un elemento complicado de definir, ya que en los dos acercamientos principales (fraseológico o probabilístico) se presentan de manera muy diferente. Por ejemplo, en muchos estudios fraseológicos las colocaciones no se han tenido en cuenta por no ser totalmente no-composicionales, ya que el significado de algunos de ellos se extrae del sentido literal de cada uno de sus elementos. Sin embargo, los análisis probabilísticos han incluido todo tipo de unidades léxicas complejas en sus estudios ya que el criterio semántico no ha sido el fundamento base para delimitar los elementos léxicos que hay que estudiar. Como afirma Wray (2002: 43) «it may be simple that identification cannot be based on a single criterion, but rather needs to draw on a suit of features».

En este estudio se entiende, por tanto, que las colocaciones no deben confundirse con los modismos o locuciones como *escurrir el bulto* (esto es, ‘zafarse de un problema’) o *estirar la pata* (es decir, ‘fallecer’), en las cuales el significado no es una función del concepto de sus componentes computada a través de las relaciones formales entre ellos. Una definición de locución de gran influencia en la filología hispánica es la ofrecida por el lexicógrafo Casares (1950: 170): «combinación estable de dos o más términos que funciona como elemento oracional y cuyo sentido unitario consabido no se justifica, sin más, como una

suma del significado normal de sus componentes». Los modismos, por consiguiente, son conjuntos léxicos invariables con una fuerte idiomatidad. Por el contrario, el significado de las colocaciones se deriva del sentido estricto de cada palabra que la compone, puede deducirse de los significados de los elementos combinados, por ejemplo, *trasplantar un órgano*. De este modo, las colocaciones son asociaciones estadísticas entre dos palabras no idiomáticas en las que cada palabra, además de conservar su propio significado, puede estar asociada de manera independiente a otras palabras. En suma, al contrario que los modismos habitualmente definidos como no-composicionales, *un cuento chino* expresa ‘mentira’, las colocaciones se caracterizan por mantener su significado. De ahí que para los que aprenden una segunda lengua, las colocaciones presenten ciertos problemas de producción pero no de comprensión, pues en la medida en que se conoce el significado de cada uno de sus elementos, se podrá comprender el conjunto (Calderan Campos, 1994: 73-74). No obstante, es importante atender a la clasificación presentada por Wasow (2002: 84). Este autor divide las colocaciones en función de su composición y frecuencia:

- Composiciones libres, como *compartir gastos* o *hacer un pastel*. Estas unidades léxicas complejas permiten una amplia gama de sustitución entre sus elementos además de una transparencia semántica.
- Colocaciones transparentes, tales como *revisar al detalle* o *decir (algo) a la cara*.
- Colocaciones semiopacas, por ejemplo *tocar la guitarra*, *dinero negro* o *diente de ajo* en las que el alumno necesita un conocimiento más profundo de la lengua para comprender el sentido exacto de la combinación.

Por todo ello, en este trabajo se asume que la idea de no-composicional no muestra un corte radical a la hora de definir las colocaciones, siguiendo de este modo la afirmación de Mel’čuk en la que reconoce que este concepto se puede representar como una línea que va desde composicional a totalmente no-composicional con categorías intermedias. Atendiendo a Wray (2002: 33), que afirma en su definición de no-composicional que «[T]he string is no longer obliged to be grammatically regular or semantically logical», observamos que los ejemplos anteriores explican cómo las colocaciones pueden ser «semánticamente lógicas». A continuación se presentan diferentes ejemplos que versan sobre la gramática de dichos conjuntos léxicos; y es que otra diferencia que hay que tener en cuenta con los modismos es que las colocaciones presentan una flexibilidad interna que no permiten las locuciones. Así *el órgano fue trasplantado* es una oración correcta a diferencia de **el bulto fue escurrido*. Además las colocaciones admiten modificación adjetival:

hacer un aterrizaje/hacer un aterrizaje forzoso
*tomar tierra/ *tomar tierra forzosa*

Asimismo, permiten relativización:

Este libro marca la *línea* que deben *seguir* sus partidarios [colocación: seguir la línea]
**El ojo* que acabo de *echar* a ese libro [locución: echar el ojo a algo]

Koike (2001: 33) también incluye pronominalización:

Asumió el cargo de alcalde, pero su repentina enfermedad le impidió desempeñarlo.

Sin embargo, se ha observado que esta característica no es única de las colocaciones, ya que el HN en ocasiones produce oraciones como: *tomar el pelo* > *tomárselo* o *empinar el codo* > *empinarlo*.

Otros autores (Pérez Serrano, 2015; Val Álvaro, 1999) han resaltado las diferencias entre las colocaciones y los compuestos sintagmáticos como *ojo de buey*, *máquina de coser* u *hombre rana*, unidades léxico-semánticas construidas con varios sintagmas. A diferencia de la mayoría de los compuestos sintagmáticos, en las colocaciones se aplican las reglas morfosintácticas de concordancia tanto en número como en género, sirva de ejemplo *enemigo acérrimo/enemigos acérrimos*, o *dinero negro*. Asimismo, resaltan las diferencias con las expresiones institucionalizadas a las que Cowie (2001) denomina fórmulas rutinarias. Dichas fórmulas, aunque no dejan de ser otro tipo de conjuntos léxicos que poseen un significado específico, tales como «¿qué tal?», «pues eso», «eso es todo», poseen una función pragmática (Nattinger y DeCarrico, 1992). Por todo esto, en la definición presentada en este estudio se asume que las colocaciones no tienen un fuerte grado de fijación interna.

Son pocos los estudios acerca de la prosodia y el lenguaje formulaico; sin embargo, se afirma que una de las principales características de este es su falta de flexibilidad en el lugar del acento prosódico, en la entonación y el tempo. Lo que crea una serie de conjuntos léxicos caracterizados por una producción específica con relación al acento prosódico o a la entonación, sin vacilación o titubeo y sin pausas internas. Esto está íntimamente ligado a la fluidez de la lengua.

Se entiende, por lo tanto, que las características fundamentales de las colocaciones, las cuales nos permiten generar su separación de las combinaciones libres, de las locuciones, los compuestos sintagmáticos o las frases institucionalizadas o idiomáticas son las siguientes:

- 1) Concurrencia frecuente de dos unidades léxicas. Esta característica puede ser observada mediante programas informáticos que calculan la frecuencia de concurrencia.
- 2) Asimismo, las siguientes características también forman parte de la definición de colocación, pero a día de hoy deben examinarse a través de un análisis manual:
 - a. Restricciones combinatorias impuestas por el uso (combinaciones producidas por los HN y no creadas por hablantes de L2)
 - b. No lexicalizadas o semitransparencia semántica, relacionada esta con un significado literal subordinado al contexto. Por ejemplo, el significado del adjetivo *astronómico* (muy alto) depende del sustantivo *precio* en *precio astronómico*, pero el mismo adjetivo cambia su sentido cuando cambia de

sustantivo, *conocimientos astronómicos* (de la astronomía). En la colocación verbo + sustantivo *levantar la sesión*, el significado de levantar se adapta a *sesión*. Del mismo modo, las colocaciones no integran actos de habla ni enunciados en sí, eso significa, «que necesitan combinarse con otros signos lingüísticos y que equivalen a sintagmas» (Corpas Pastor, 1996: 51).

- c. Flexibilidad composicional (permiten modificación adjetival y relativización).
- d. Producción sin vacilación o titubeo y sin pausas internas.

Creemos que es de suma importancia entender las diferencias existentes entre los diferentes conjuntos léxicos para después poder crear no solo un continuo en el lenguaje formulaico, sino producir materiales concretos y acertados para el aula de ELE, donde se señalen las dificultades que conlleva el aprendizaje de cada uno de ellos.

Referencias bibliográficas

Alonso Ramos, M. 1993. *Las funciones léxicas en el modelo lexicográfico de I. Mel'cuk*. Madrid: Universidad Nacional de Educación a Distancia, España.

Alonso Ramos, M. 1994. Hacia una definición del concepto de colocación: de J. R. Firth a I. A. Mel'čuk. *Revista de Lexicografía* 1: 9-28.

Alonso Ramos, M. 2004. *Las construcciones con verbo de apoyo*. Madrid: Visor Libros.

Alonso Ramos, M. 2012. Naturaleza semántica de las colocaciones españolas. Ponencia presentada en Meanings, Texts and Other Exciting Things. A Festschrift to Commemorate the 80th Anniversary of Professor Igor Alexandrovich Mel'chuk, Moscow. http://www.ruslang.ru/doc/melchuk_festschrift2012/Ramos.pdf

Bahns, J., H. Burmeister y T. Vogel. 1986. The pragmatics of formulas in L2 learner speech: Use and development. *Journal of Pragmatics* 10, 6: 693-723.

Benson, M., E. Benson y R. F. Ilson. 1986. *Lexicographic Description of English*. Amsterdam: John Benjamins.

Blanco Escoda, X. 2000. Verbos soporte y clase de predicados en español. *Lingüística española actual* 22, 1: 99-118.

Bosque, I. (ed.). 2004 *Diccionario combinatorio del español contemporáneo*. Madrid: Ediciones SM.

Burger, H. 1998. *Phraseology: An introduction on the Example of the Germans*. Berlin: Erich Schmidt.

Bybee, J. 2002. Word frequency and context of use in the lexical diffusion of phonetically conditioned sound change. *Language Variation and Change* 14: 261-290.

- Bybee, J. y D. Eddington. 2006. A usage-based approach to Spanish verbs of 'becoming'. *Language* 8, 2: 323-355.
- Calderan Campos, M. 1994. Sobre la elaboración de diccionarios monolingües de producción: las definiciones, los ejemplos y las colocaciones. In P. Slagter (ed.), *Foro Hispánico* 6. "Aproximaciones a cuestiones de adquisición y aprendizaje del español como lengua extranjera o como segunda lengua", 105-119. Ámsterdam: Rodopi.
- Casares, J. 1950. *Introducción a la lexicografía moderna*. Madrid: CSIC Consejo Superior de Investigaciones Científicas.
- Castillo Carballo, M. A. 1998. El término 'colocación' en la lingüística actual. *Lingüística Española Actual* 20, 1: 41-54.
- Corpas Pastor, G. 1996. *Manual de fraseología española*. Madrid: Gredos.
- Corpas Pastor, G. 2001. Apuntes para el estudio de la colocación. *LEA: Lingüística española actual* 23, 1: 41-56.
- Coşeriu, E. 1977. *Principios de semántica estructural*. Madrid: Gredos.
- Cowie, A. P. 1998. *Phraseology: Theory, Analysis, and Applications*. Oxford: University Press.
- Cowie, A. P. 2001. Speech formulae in English: problems of analysis and dictionary treatment. In G. v. d. Meer & A. G. B. t. Meulen (eds.), *Making Senses: From Lexeme to Discourse. In Honor of Werner Abraham*, Vol. 44, 1-12. Groninger Arbeiten zur germanistischen Linguistik 44. Center for language and Cognition Groningen.
- Davies, M. (dir.). 2002. *Corpus del Español: 100 million words, 1200s-1900s*. <http://www.corpusdelespanol.org>
- De Miguel, E. 2008. Construcciones con verbos de apoyo en español. De cómo entran los nombres en la órbita de los verbos. Ponencia presentada en *Actas del XXXIX Simposio Internacional de la Sociedad Española de Lingüística*, Pamplona.
- Durrant, P. 2008. *High Frequency Collocations and Second Language Learning*. PhD thesis, University of Nottingham.
- Erman, B. y B. Warren. 2000. The idiom principle and the open choice. *Text* 20, 1: 29-62.
- Evert, S. 2005. *The statistics of word co-occurrences: Word pairs and collocations*. Stuttgart: University of Stuttgart.
- Evert, S. 2009. Corpora and collocations. In A. Lüdeling & M. Kytö (eds.), *Corpus Linguistics. An International Handbook*, 1212-1248. Berlin: Walter de Gruyter Co.
- Firth, J. R. 1957. *Papers in linguistics 1934-1951*. London: Oxford University Press.

- González Fernández, B., y Schmitt, N. 2015. How much collocation knowledge do L2 learners have?: The effects of frequency and amount of exposure. *International Journal of Applied Linguistics* 166: 94–126.
- Granger, S. y Y. Bestgen. 2014. The use of collocations by intermediate vs. advanced non-native writers: A bigram-based study. *International Review of Applied Linguistics in Language Teaching* 52, 3: 229-252.
- Granger, S., y Paquot, M. 2008. Disentangling the phraseological web. In S. Granger & F. Meunier (eds.), *Phraseology: An Interdisciplinary Perspective*, 27-49. Amsterdam/Philadelphia: Benjamins.
- Halliday, M. 1961. Categories of the Theory of Grammar. *Word* 17: 241-292.
- Halliday, M. 1966. Lexis as a Linguistic Level. In C. E. Bazell, J. C. Catford, M. A. Halliday & R. H. Robins (eds.), *In Memory of J.R. Firth*, 148-163. Longman: London.
- Hallin, A. E., y D. van Lancker Sidits. 2015. A Closer Look at Formulaic Language: Prosodic Characteristics of Swedish Proverbs. *Applied Linguistics* 38, 1: 68–89.
- Higueras García, M. 2006. *Las colocaciones y su enseñanza en la clase de ELE*. Madrid: Arco Libros.
- Hoey, M. 1991. *Patterns of Text*. Oxford: University Press.
- Hunston, S. 2002. *Corpora in applied linguistics*. Cambridge, UK: University Press.
- Írsula, J. 1992. Colocaciones sustantivo-verbo. In G. Wotjak (ed.), *Estudios de lexicografía y metalexigrafía del español actual*, 119-164. Tübinga: Max Niemeyer Verlag.
- Írsula, J. 1994. Entre el verbo y el sustantivo ¿quién rige a quién? El verbo en las colocaciones sustantivo-verbales. *Verbo e Estructura Frásicas: actas do IV Colóquio Internacional de Linguística Hispânica de Leipzig*, 277-286.
<http://ler.letras.up.pt/site/default.aspx?qry=id026id128&sum=sim>
- Jensen, E. 2017. *What's in a Phrase: The Acquisition and Comprehension of Collocations in an Advanced Spanish Classroom*. University of California, Davis, Davis, CA.
- Koike, K. 2001. Colocaciones léxicas en el español actual: estudio formal y léxico-semántico. Alcalá de Henares: Universidad de Alcalá & Takushoku University.
- Levelt, W. 1999. A Blueprint of the speaker. In C. Brown & P. Hagoort (eds.), *The neurocognition of language*, 83-123. Oxford: University Press.
- Lin, P. 2010. The Phonology of Formulaic Sequences: A Review. In D. Wood (ed.), *Perspectives on formulaic language: Acquisition and communication*, 174-193. London/New York: Continuum.
- Mel'čuk, I. 1998. Collocations and Lexical Functions. In A. P. Cowie (ed.), *Phraseology: Theory, Analysis and Applications*, 23-53. Oxford: University Press.
- Mel'čuk, I. 2012. *Semantics: From meaning to text*. Amsterdam: John Benjamins.

- Méndez Cedón, B. 2008. Las colocaciones en la prosa académica médica: *Análisis contrastivo inglés-español*. In C. M. Blanco (ed.), *Colocaciones y la fraseología en los diccionarios*, 117-130. Frankfurt: Peter Lang.
- Nattinger, J. R. y J. S. DeCarrico. 1992. *Lexical Phrases and Language Teaching*. Oxford: University Press.
- Nesselhauf, N. 2005. *Collocations in a Learner Corpus*. Philadelphia: John Benjamins Publishing.
- O'Donnell, M. B., U. Römer y N. Ellis. 2013. The development of formulaic sequences in first and second language writing. *International Journal of corpus linguistics* 18, 1: 83-108.
- Pérez Serrano, M. 2015. *Un enfoque léxico a prueba: efectos de la instrucción en el aprendizaje de las colocaciones léxicas*. Universidad de Salamanca, España.
- Seco, M. 1978. *Estudio de lexicografía española*. Madrid: Paraninfo.
- Sinclair, J. 1966. Beginning the Study of Lexis. In C. E. Bazell, J. C. Catford, M. A. Halliday & R. H. Robins (eds.), *In Memory of J.R. Firth*, 110-130. London: Longman.
- Sinclair, J. 1991. *Corpus, Concordance, Collocation*. Oxford: University Press.
- Sinclair, J. 2004. *Trust the Text: Language, Corpus and Discourse*. London and New York: Routledge.
- Tordera Yllescas, J. C. 2012. *El abecé de la lingüística computacional*. Madrid: ArcoLibros.
- Val Álvaro, J. 1999. La composición. In I. Bosque & V. Demonte (eds.), *Gramática descriptiva de la lengua española*, 4758-4841. Madrid: Espasa.
- Wasow, T. 2002. *Postverbal behavior*. Stanford: CSLI.
- Wood, D. 2005. *Fundamentals of Formulaic Language: An Introduction*. New York: Bloomsbury.
- Wray, A. 2002. *Formulaic language and the lexicon*. Cambridge, UK: University Press.