



DIVERSIDAD Y SOFISTICACIÓN DE COLOCACIONES EN LA PRODUCCIÓN ESCRITA DE APRENDICES Y HABLANTES NATIVOS DE ESPAÑOL

Rocío CUBEROS-VICENTE¹
Universidad Complutense de Madrid

Elisa ROSADO²
Universidad de Barcelona

Resumen

El desarrollo del lenguaje formulaico en segundas lenguas (L₂) ha sido analizado en estudios que señalan el uso de colocaciones como indicador fiable de competencia lingüística. Sin embargo, las colocaciones han recibido poca atención en la investigación sobre complejidad lingüística. Este estudio pretende contribuir a dicha investigación examinando la complejidad fraseológica, constructo que refiere a la diversidad y la sofisticación de las colocaciones utilizadas en la producción escrita de aprendices de nivel intermedio y avanzado y hablantes nativos de español. Se analizan las colocaciones verbo-sustantivo y sustantivo-adjetivo en dos tipos de texto: narrativo y expositivo. Los datos provienen de dos tareas de producción escrita espontánea a partir de un estímulo visual. Los resultados indican que la complejidad fraseológica aumenta con el nivel de competencia en ambos tipos de texto, pero su desarrollo varía según la estructura sintáctica. Se discuten las implicaciones de estos hallazgos para la enseñanza de L₂ y se sugieren orientaciones para futuros estudios.

Palabras clave: español lengua segunda; colocaciones; complejidad fraseológica; diversidad; sofisticación.

DIVERSITY AND SOPHISTICATION OF COLLOCATIONS IN WRITTEN PRODUCTION OF LEARNERS AND NATIVE SPEAKERS OF SPANISH

Abstract

The development of formulaic language in second language (L₂) has been analyzed in studies that point to the use of collocations as a reliable indicator of linguistic proficiency.

1. marcubert@ucm.es;  <https://orcid.org/0000-0002-9089-5092>
2. erosado@ub.edu;  <https://orcid.org/0000-0002-1459-284X>

However, collocations have received little attention in research on linguistic complexity. The present study aims to contribute to this line of research by analyzing phraseological complexity, a construct that refers to the diversity and sophistication of collocations used in written production of intermediate and advanced learners and native speakers of Spanish. Verb-noun and noun-adjective collocations are analyzed in narrative and expository texts. The data comes from two spontaneous writing production tasks based on a visual stimulus. The results show that phraseological complexity increases with proficiency, with equivalent patterns across text types, but the developmental pattern did vary according to the syntactic structure of collocations. The implications of these findings for L2 teaching are discussed and directions for future research are suggested.

Keywords: Spanish as a second language; collocations; phraseological complexity; collocational diversity; collocational sophistication.

RECIBIDO: 21/02/2025

ACEPTADO: 14/11/2025

1. INTRODUCCIÓN

El lenguaje formulaico desempeña un papel fundamental en el desarrollo del lenguaje nativo y no nativo, en tanto que el dominio de determinadas expresiones prefabricadas puede facilitar el aprendizaje de lenguas primeras (L1) y segundas (L2) (Ellis et al., 2015; Hyland, 2008; Schmitt, 2004). La reflexión metalingüística sobre la composición de dichas expresiones puede suponer para el aprendiz una valiosa estrategia de análisis que, en última instancia, contribuya al aprendizaje de la L2 (Ellis et al., 2015). Además, el conocimiento y uso de estas expresiones aporta fluidez y naturalidad al discurso, por cuanto constituyen opciones lingüísticas preferidas frente a otras alternativas gramaticalmente posibles, pero menos frecuentes (Kecske, 2007). Estudios previos sobre la relación entre la calidad de los textos y sus características léxicas (Crossley et al., 2012; Paquot, 2018) coinciden en señalar la necesidad de considerar el uso de unidades léxicas multipalabra para alcanzar una correcta valoración de la calidad de los textos producidos por aprendices de L2. Por ejemplo, en Cuberos (2019) se demostró cómo el uso de colocaciones por parte de hablantes nativos y no nativos de español está asociado no solo a niveles de competencia más altos, sino también a la evaluación de la calidad textual, por cuanto su uso contribuyó a explicar variaciones en las calificaciones asignadas a los textos por parte de jueces expertos.

Dentro del estudio del lenguaje formulaico en el ámbito del español L2, las colocaciones léxicas han recibido especial atención en los últimos años (véase, por

ejemplo, García-Salido et al., 2019; Pérez-Serrano y Mavrou, 2022), principalmente, porque su dominio representa un desafío para los aprendices de L2 (Fan, 2009; Nesselhauf, 2004). Los estudios relacionados con el uso de colocaciones léxicas en español L2 se han centrado sobre todo en caracterizar el repertorio de colocaciones en las producciones de aprendices en contraste con el equivalente nativo y, en menor medida, en compararlo en distintos niveles de competencia. Todos ellos coinciden en señalar el carácter deficitario del repertorio de colocaciones empleado en español L2, de forma que los aprendices no solo suelen emplear colocaciones erróneas que se desvían de la equivalente nativa sino que, además, tienden a sobreutilizar determinadas colocaciones, ya sea por su alta frecuencia en la lengua meta o por la similitud con la L1 (García-Salido, 2016; García-Salido y García, 2017; García-Salido et al., 2019). Con respecto a su uso según el nivel de competencia, se observa un aumento progresivo en la cantidad, diversidad y corrección de las colocaciones a medida que aumenta el grado de dominio lingüístico (Cuberos, 2019; Cuberos et al., 2024; García-Salido et al., 2019). Sin embargo, incluso en niveles avanzados, el repertorio de colocaciones empleado sigue sin ser equivalente al de los hablantes nativos (García-Salido et al., 2019). Estos resultados coinciden con los obtenidos sobre el uso de colocaciones por parte de aprendices de otras lenguas como el inglés (Durrant y Schmitt, 2009), el francés (Vandeweerd et al., 2021) o el alemán (Boone y Eyckmans, 2023).

Este estudio pretende contribuir a la caracterización del repertorio de colocaciones del español nativo y no nativo en dos corpus equivalentes. Para ello, examinamos dicho repertorio atendiendo al concepto de *complejidad fraseológica*, que alude a dos propiedades del uso que se hace de las colocaciones, a saber, la *diversidad colocacional* y la *sofisticación colocacional* (Paquot, 2019). Mientras que la *diversidad colocacional* refiere a la variedad de colocaciones utilizadas en un texto, el concepto de *sofisticación colocacional* identifica el uso de colocaciones menos comunes y propias de niveles avanzados. En español como L2, estas propiedades han sido analizadas en el conjunto de colocaciones producidas por los aprendices de modo general (véase, por ejemplo, García-Salido et al., 2019), sin distinguir su comportamiento en distintas estructuras sintácticas. No obstante, para alcanzar una adecuada caracterización del repertorio de colocaciones en L2, estimamos pertinente considerar también la configuración interna de las mismas. Esto es así, principalmente, porque el desarrollo y dominio de las colocaciones en aprendices de L2 está determinado, entre otros factores, por la complejidad sintáctica de las colocaciones (Bardovi-Harlig y Stringer, 2017; Gitsaki, 1996). Además, estudios previos realizados en inglés, francés y holandés como L2 (Paquot, 2018, 2019; Vandeweerd et

al., 2021; Rubin et al., 2021) coinciden en identificar patrones de desarrollo distinto en función de la estructura sintáctica de las colocaciones. Por ejemplo, Vandeweerd et al. (2021) encontraron que la sofisticación de las colocaciones de verbo-sustantivo utilizadas por aprendices de francés L2 aumentó significativamente con el nivel de competencia, mientras que la sofisticación de otras combinaciones analizadas, a saber, las colocaciones de sustantivo-adjetivo y las de adverbio-adjetivo, se mantuvieron similares en los distintos grupos de aprendices. En el presente estudio se analiza la diversidad y la sofisticación de las colocaciones en dos combinaciones distintas, definidas según las categorías gramaticales implicadas y su función sintáctica: 1) las colocaciones verbo-sustantivo(objeto), caracterizadas por formar un sintagma verbal en el que el sustantivo desempeña la función de complemento directo (p. ej.: *dar un paseo*) o de complemento de régimen (p. ej.: *carecer de dinero*), y 2) las colocaciones formadas por un nombre y un adjetivo que lo modifica (p. ej.: *ayuda mutua*). Estos dos tipos de colocaciones son las más frecuentes tanto en los corpus analizados (Cuberos et al., 2024), como en español en general (Koike, 2001).

Por otra parte, estudios previos sobre el desarrollo del repertorio lingüístico en L1 y L2 coinciden en señalar la necesidad de evaluar el dominio lingüístico de hablantes nativos y no nativos en distintos tipos de texto, por cuanto el desarrollo lingüístico se ve condicionado de forma determinante por la situación comunicativa (Aparici, 2010; Cuberos, 2019; Ravid y Tolchinsky, 2002; Rosado et al., 2014). A pesar de ello, la mayoría de los trabajos previos analiza el uso de colocaciones en producciones escritas sin considerar la influencia de la tipología textual. En este estudio, analizamos el uso de colocaciones en dos tipos de textos escritos: narrativos y expositivos. Estos géneros discursivos resultan idóneos para analizar el desarrollo del repertorio lingüístico porque contrastan por sus características textuales definitorias y, además, se dominan en momentos evolutivos distintos (Aparici, 2010).

1.1. *De nuevo sobre el concepto de colocación*

El fenómeno lingüístico al que remite el término *colocación* se ha definido de diversas maneras tanto en lingüística como en el ámbito de la enseñanza de lenguas, y se ha denominado también con términos como unidades léxicas multipalabra o combinaciones prefabricadas, entre otros³. Tanto es así que, como señala Nesselhauf

3. En este artículo, utilizamos los distintos términos empleados por distintos autores, por ejemplo, colocaciones, adoptado, entre otros, por Cowie (1998) o unidad fraseológica, adoptado, entre otros, por Vinogradov (1947), como equivalentes.

(2004), parece que el único aspecto sobre el que existe consenso generalizado es que el fenómeno hace referencia a algún tipo de relación sintagmática entre un grupo de palabras (p. 11).

Tradicionalmente, la noción de *colocación* se aborda en lingüística desde dos perspectivas principales: el enfoque distribucional o basado en la frecuencia, y el enfoque fraseológico (Nesselhauf, 2004, p. 11). La diferencia principal entre estos enfoques radica en las propiedades definitorias que atribuyen a las colocaciones: mientras que el primero describe estas unidades lingüísticas en términos de la frecuencia de coocurrencia de palabras (Firth, 1957, p. 182; Sinclair, 1987, p. 320), el segundo se centra en las posibilidades combinatorias de las palabras que las componen en contextos sintácticos específicos, de modo que describe las colocaciones por medio de criterios estrictamente lingüísticos (Cowie, 1998; Hausmann, 1989).

Desde el enfoque distribucional se sostiene que ciertas palabras tienden a aparecer juntas en determinadas combinaciones y la noción de *colocación* refiere específicamente al contexto en el que aparece una palabra (Laso, 2009, p. 25). A partir de la noción de colocación de Firth (1957, p. 179): «you shall know a word by the company it keeps», lingüistas como Halliday (1961, p. 276) y Sinclair (1991, p. 170) destacan la frecuencia de coocurrencia de ciertas palabras teniendo también en cuenta la distancia entre los elementos que conforman la colocación. De acuerdo con esta aproximación, las colocaciones se identifican aplicando criterios de frecuencia (Sinclair, 1987, p. 320) y también considerando la fuerza de asociación de los elementos que conforman la colocación (Kilgarriff, 2006, p. 38). Por ejemplo, la fuerza de asociación del sintagma *coche oficial* resulta débil porque ambas palabras pueden aparecer en numerosas combinaciones, mientras que, por ejemplo, la colocación *piedra angular* muestra una fuerza de asociación fuerte porque el adjetivo *angular* no suele coaparecer con otras palabras para transmitir un significado unitario.

Los análisis realizados desde esta perspectiva han sido, sin duda, de gran utilidad para comprender mejor las relaciones sintagmáticas que se establecen entre distintos tipos de expresiones multipalabra (Laso, 2009, p. 24). Además, han facilitado la identificación de combinaciones léxicas que desde otros enfoques habrían pasado desapercibidas (Granger y Paquot, 2008, p. 14). A pesar de ello, desde el enfoque fraseológico, autores como Howarth (1996, p. 25) o Mel'čuk (1998, p. 24) sostienen que es fundamental considerar criterios lingüísticos. Así, en el marco del enfoque fraseológico se sostiene que las colocaciones deben necesariamente definirse mediante criterios gramaticales, léxicos y/o semánticos, en lugar de basarse exclusivamente en la frecuencia de coaparación. Desde esta perspectiva se adopta, por tanto, una estrategia

de identificación de colocaciones guiada por criterios lingüísticos previamente establecidos. Así, las unidades léxicas multipalabra se sitúan en un continuo que incluye desde combinaciones libres de palabras semánticamente transparentes en un extremo, hasta expresiones idiomáticas cuyo significado no es composicional en el otro (Cowie, 1998, p. 6). El grado de fijación de las combinaciones se define también como gradual (Corpas, 1996, p. 20), de modo que el *continuum* en el que se ubican las colocaciones implica, sin embargo, áreas difusas cuyos límites resultan difíciles de definir. A pesar de ello, de manera generalizada, se sostiene que las colocaciones son combinaciones prototípicas de dos palabras léxicas que mantienen algún tipo de relación gramatical entre sí (Mittmann, 2013, p. 500) y entre las que se establece un vínculo semántico (Corpas, 2001, p. 42). Si bien los dos miembros de la colocación contribuyen a su significado, uno de ellos (la base) es semánticamente autónomo mientras que el significado del otro (el colocativo) viene semánticamente restringido y seleccionado por la base (Hausmann, 1998, p. 65). Por ejemplo, en la colocación *dar apoyo*, el significado del verbo *dar* (el colocativo) está contextualmente condicionado por el del sustantivo *apoyo* (la base).

Junto a estos dos enfoques tradicionales sobre el concepto de colocación, la propuesta de Bosque (2001, 2011) nos ofrece un marco para analizar las colocaciones desde una perspectiva gramatical en el que defiende que las colocaciones no forman parte del ámbito de la fraseología, sino que se sitúan en la interfaz léxico-sintáctica. Sostiene, además, que la frecuencia por sí misma no proporciona información sobre las propiedades lingüísticas de las colocaciones. En línea con el enfoque fraseológico, Bosque también refiere a la relación semántica que se establece entre los elementos de una colocación; sin embargo, recurre al concepto de 'selección léxica' para defender que es el colocativo, y no la base, el que selecciona palabras de una determinada clase léxica. Siguiendo con el ejemplo anterior, Bosque sostiene que es el verbo *dar* el que selecciona a palabras semánticamente relacionadas como *apoyo*, *ayuda* o *soporte* (Bosque, 2006, p. 382), y no a la inversa.

En este estudio, y en línea con autoras como Granger y Paquot (2008, p. 14) o Laso (2009, p. 9), sostenemos que estos enfoques complementarios son complementarios y no excluyentes, puesto que contribuyen a una comprensión más exhaustiva de las relaciones sintagmáticas que se establecen entre los constituyentes de las unidades léxicas multipalabra. Además, como destaca Wood (2020, p. 42), la combinación de criterios favorece la identificación de unidades fraseológicas. En la mayoría de los estudios basados en corpus, las colocaciones se identifican a partir de criterios distribucionales o fraseológicos y, solo en algunas ocasiones, las investigaciones aplican

ambos criterios (véase, por ejemplo, García-Salido et al., 2019). En este último caso, el proceso suele consistir en la extracción de las colocaciones utilizando medidas estadísticas de frecuencia y, posteriormente, en la desambiguación manual de los resultados para descartar aquellas combinaciones que no cumplan con los criterios fraseológicos previamente establecidos. Adoptar en primer lugar criterios de frecuencia, no obstante, tiene ciertas limitaciones, especialmente cuando se aplica a conjuntos pequeños de datos, puesto que las colocaciones con solo una o dos ocurrencias pueden fácilmente pasar desapercibidas (Wood, 2020). Así pues, en línea con trabajos anteriores (Cuberos, 2019; Cuberos et al., 2024), adoptamos una definición deliberadamente amplia de colocación basada en el enfoque fraseológico que se complementa, por una parte, con un criterio lingüístico de autoridad y, por otra, con un criterio de frecuencia (v. Método). Así, con el objetivo de superar las limitaciones inherentes tanto a los criterios exclusivamente fraseológicos como a los basados únicamente en la frecuencia, el enfoque que adoptamos pretende ofrecer una comprensión más completa del comportamiento de las combinaciones léxicas multipalabra.

2. OBJETIVOS

En este estudio se analiza la diversidad y la sofisticación de las colocaciones en producciones escritas de estudiantes universitarios nativos y no nativos de español. El objetivo es determinar si la diversidad y la sofisticación de estas colocaciones varían según el nivel de competencia de los participantes y el tipo de texto producido. En particular, se examina las colocaciones más frecuentes en español, colocaciones de verbo-sustantivo(objeto) y colocaciones de sustantivo-adjetivo, en dos tipos de textos escritos: narrativos y expositivos.

3. MÉTODO

3.1. *Participantes*

En el estudio participaron 50 estudiantes universitarios: 30 hablantes no nativos de español como L2 (18 mujeres) (rango de edad: 20 – 43;6) y 20 hablantes nativos (14 mujeres) (rango de edad: 18;11 – 28;4). Los participantes de L2 tenían como L1 árabe, chino o coreano. Para determinar su nivel de competencia, docentes experimentados de español como L2 evaluaron y clasificaron los textos en niveles, que fueron distribuidos en dos grupos: inicial-intermedio (A2 y B1) ($n = 11$) e intermedio-avanzado (B2 y C1) ($n = 19$).

3.2. *Tareas de producción*

Los datos proceden de dos corpus análogos: el corpus GRERLI-ES1 (Tolchinsky y Aparici, 2015) y el corpus GRERLI-ES2 (Tolchinsky et al., 2023) los cuales incluyen textos narrativos y expositivos escritos y orales producidos por hablantes nativos y no nativos de español, respectivamente. Las tareas de producción fueron equivalentes para L1 y L2 (v. Berman y Verhoeven, 2002)⁴. Los participantes vieron un vídeo sin diálogos que mostraba situaciones de conflicto en un entorno escolar. A continuación, se les pidió que relataran oralmente y por escrito una experiencia personal similar (texto narrativo) y que discutieran, también oralmente y por escrito, sobre el tema del vídeo (texto expositivo). Como resultado, cada participante produjo cuatro tipos de textos: narrativo escrito, narrativo oral, expositivo escrito y expositivo oral. Para este estudio, se seleccionan únicamente los textos escritos.

3.3. *Identificación de colocaciones*

Las colocaciones se identificaron manualmente mediante la Prueba para la Identificación de Unidades Fraseológicas (TIPUS, por sus siglas en inglés) (Cuberos et al., 2024). TIPUS es un procedimiento para la identificación de colocaciones que combina tres tipos de criterios: fraseológicos, distribucionales y de autoridad lingüística. Para que una unidad léxica se considere una colocación, se deben cumplir dos de los tres criterios establecidos.

El primer criterio consiste en identificar manualmente unidades léxicas candidatas a ser consideradas colocación siguiendo una definición deliberadamente amplia basada en el enfoque fraseológico (véase apartado 1.1.). En concreto, en línea con Corpas (1996; 2001), entendemos las colocaciones como combinaciones sintagmáticas recurrentes entre las que se establece un vínculo léxico-semántico. Para guiar este proceso de identificación, consideramos dos propiedades fundamentales de las colocaciones propuestas por Higueras (2006):

1. Unidades multipalabra compuestas por al menos dos lexemas.
2. Unidades multipalabra con un significado unificado, cierta fijación y coherencia interna.

4. Más detalles sobre el procedimiento de elicitation de los textos puede consultarse en la página web de CHILDES (GRERLI-ES1: <https://childe.talkbank.org/access/Spanish/GRERLI.html>; GRERLIES2: <https://slabank.talkbank.org/access/Spanish/GRERLI-ES2.html>).

Solo se anotaron aquellas combinaciones léxicas canónicas en español, por lo que las colocaciones creativas y/o erróneas de los aprendices (p. ej., *tener un paseo*^{*}, por *dar un paseo*), no se identificaron como candidatas a colocación, principalmente, porque no cumplen con los criterios fraseológicos establecidos, pero también porque estas colocaciones que se desvían de la norma no forman parte del objeto de estudio. Para garantizar la fiabilidad de los *juicios intuitivos* (Wray, 2008) que guían este proceso, se discutieron los casos dudosos entre las investigadoras y se calculó la fiabilidad de la anotación en el 20% de los textos mediante el coeficiente Alfa de Cronbach donde se obtuvo un acuerdo significativo entre juezas ($\alpha \geq .784$) (Field, 2009). Como resultado de este primer paso, se anotaron las combinaciones léxicas candidatas a colocación. A modo de ejemplo, incluimos a continuación algunos ejemplos para ilustrar las combinaciones léxicas anotadas en este primer paso:

- (1) *Si lo ha hecho mal, que pida perdón*
- (2) *Hay tantas imágenes violentas como en la guerra o en conflictos laborales*
- (3) *Gente no le gusta sentarse al lado porque tiene envidia*
- (4) *Nos cogían como boxeadores y nos daban masajes pequeñitos en la espalda*

El segundo paso, basado en el enfoque distribucional, consiste en medir la fuerza de asociación y frecuencia de coocurrencia de las unidades léxicas que conforman las combinaciones candidatas a colocación mediante la puntuación *logDice*. Se trata de una medida de asociación léxica que combina información sobre la frecuencia de coocurrencia y la fuerza de asociación de los constituyentes de la colocación (Rychlý, 2008). Estas puntuaciones se obtuvieron mediante la aplicación Word Sketch disponible en el programa de análisis de corpus Sketch Engine (Kilgariff et al., 2014). Las búsquedas, que se realizan tomando como referencia la base de la colocación, ofrecen listados organizados según las relaciones gramaticales que se establecen entre la base y los colocativos. Por ejemplo, en la búsqueda de *línea*, sus colocativos aparecen agrupados según se trate de verbos que coocurren con la base como objeto directo (p. ej., *trazar; trazar una línea*), o de adjetivos que la modifican (p. ej., *recto; línea recta*). Decidimos utilizar la medida de asociación *logDice* en lugar de otras alternativas como la puntuación *MI* o la puntuación *T* porque su valor no depende del tamaño del corpus y porque, a diferencia de estas otras medidas, se basa en un máximo teórico, lo que resulta fundamental para establecer comparaciones entre distintos corpus (Gablasova et al., 2017). Se ha demostrado, además, que *logDice* funciona mejor que otras medidas en la identificación de colocaciones (Gablasova et al., 2017). En concreto, a las combinaciones candidatas

se les asignó la puntuación *logDice* obtenida en el corpus de español esTenTen (Kilgariff y Renau, 2013), disponible también en Sketch Engine. El máximo teórico de esta puntuación es 14, sin embargo, los valores que alcanza suelen ser inferiores a 10 (Rychlý, 2008). Por esta razón convinimos que las combinaciones candidatas a colocación debían obtener una puntuación *logDice* igual o superior a 5 para cumplir con el criterio distribucional. En los casos en que colocaciones identificadas en el primer paso son instancias de la misma colocación (p. ej., *sacar buenas notas* y *sacar malas notas*), la puntuación se calculó a partir de la media de las puntuaciones de cada combinación. La Tabla 1 recoge las colocaciones identificadas en el primer paso junto con sus puntuaciones *logDice*, aquellas sombreadas en gris no alcanzaron el umbral mínimo de 5 puntos.

Colocaciones candidatas	Puntuación <i>logDice</i>
Pedir perdón	9.97
Conflictivo laboral	7.63
Dar un masaje	4.23
Tener envidia	2.99

Tabla 1. Puntuaciones *logDice* de las colocaciones candidatas

Por último, como criterio lingüístico de autoridad, se verifica la presencia de las colocaciones en el *Diccionario combinatorio práctico del español contemporáneo (Práctico)* (Bosque, 2006). Decidimos utilizar este diccionario porque está basado en la noción de colocación de Bosque (2011) que, recordamos, sostiene que las colocaciones responden a casos de selección léxica en las que el colocativo, interpretado en sus obras como predicado, selecciona la base, interpretada en sus obras como argumentos. De este modo, las búsquedas en *Práctico* se realizaron tomando como referencia el colocativo en lugar de la base.

Una vez asignadas las puntuaciones *logDice* a las colocaciones candidatas y tras verificar su presencia en *Práctico*, se identificó la lista final de colocaciones. Si bien encontramos casos en los que algunas colocaciones no estaban recogidas en *Práctico* (p. ej., *conflicto laboral*) o no alcanzaban el umbral de frecuencia establecido (p. ej., *tener envidia*), solo se descartaron aquellas colocaciones que no cumplían ninguno de los dos criterios.

Como resultado de aplicar este procedimiento, en los corpus de análisis se identificaron 1392 colocaciones: 710 en el corpus de L1 y 682 en el corpus de L2. En este estudio, como se ha dicho, se analizan las colocaciones de

verbo-sustantivo(objeto) (p. ej., *ocasionar problemas*; *ir al teatro*) y las de sustantivo-adjetivo (p. ej., *año pasado*; *ayuda mutua*), que son los dos tipos de colocaciones más frecuentes en español y en los corpus analizados. En la Tabla 2 se muestra el número total de colocaciones y el número de colocaciones de cada tipo por nivel de competencia.

Nivel de competencia	N.º verbo-sustantivo	N.º sustantivo-adjetivo	N.º total
Inicial-intermedio	72	31	167
Intermedio-avanzado	189	77	515
Nativo	321	90	710

Tabla 2. Número de colocaciones total y número de verbo-sustantivo(objeto) y sustantivo-adjetivo por nivel de competencia

3.3. Complejidad fraseológica

La complejidad fraseológica es, como se ha dicho, un constructo que describe el uso de dos propiedades del uso que se hace de las colocaciones: la diversidad colocacional y la sofisticación colocacional (Paquot, 2019). La diversidad, que refiere a la variedad de colocaciones utilizadas, se obtuvo mediante la proporción de colocaciones diferentes (lemas)⁵ sobre la raíz cuadrada del número total de colocaciones (Paquot, 2019). Por su parte, la sofisticación, que refiere al uso de colocaciones menos comunes y avanzadas, se calculó mediante las puntuaciones *logDice*. Esta medida de asociación léxica combina información sobre la frecuencia de coaparación y la fuerza de asociación de los elementos que conforman la colocación (Rychlý, 2008), de tal modo que aporta información sobre cuán frecuente es la combinación al mismo tiempo que indica cuál es la fuerza de la asociación entre ambas palabras (Gablasová et al., 2017). El valor medio que suele alcanzar esta puntuación es 10 (Rychlý, 2008), por lo que los valores por encima de 7 se consideran indicativos de colocaciones fuertes (Cuberos et al., 2024). Decidimos calcular la sofisticación de las colocaciones mediante una medida de fuerza de asociación porque la gran mayoría de colocaciones que obtienen puntuaciones altas en estas medidas tienen significados específicos (Ellis et al., 2008;

5. En línea con García-Salido y García (2018) y Cuberos et al. (2024), optamos por calcular la diversidad de las colocaciones considerando los lemas en lugar de las formas, como es habitual en lenguas como el inglés. Esta decisión se justifica tanto por la complejidad morfológica del español, que puede inflar artificialmente la diversidad al contar formas flexionadas, como por los hallazgos en procesamiento de colocaciones que indican que la variación morfológica no inhibe su ventaja de procesamiento (Vilkaitė-Lozdienė, 2022; Pérez-Serrano e Hidalgo-Gallardo, en prensa).

Gablasova et al., 2017). Por ejemplo, mientras que *perro grande* y *perro pequeño* son las combinaciones de sustantivo-adjetivo más frecuentes en el corpus esTenTen23 (Kilgariff y Renau, 2013), las combinaciones *perro guardián* y *perro callejero* son las que obtienen puntuaciones logDice más altas, lo que indica que las colocaciones con puntuaciones más altas tienen una mayor especialización semántica. Además, porque en estudios anteriores (véase, por ejemplo, Paquot, 2019; Vandeweerd et al., 2021; Rubin et al., 2021) estas medidas han resultado especialmente útiles para describir el desarrollo de la competencia fraseológica en L2.

3.4. Análisis de datos

Para abordar los objetivos de investigación, en primer lugar, se realizaron análisis estadísticos descriptivos y exploraciones visuales de los datos con diagramas de cajas. Seguidamente, se realizaron una serie de análisis ANOVA robustos de medidas repetidas con el Nivel de competencia como factor intersujeto (intermedio, avanzado, nativo) y el Tipo de texto como factor intrasujeto (narrativo, expositivo). Los análisis se llevaron a cabo en R (R Core Team, 2021) con la función *bwtrim* del paquete *WSR2* (Mair y Wilcox, 2016). En caso de interacciones significativas, se analizaron efectos simples de Nivel de competencia y de Tipo de texto con las funciones *lincon* y *yuend*, respectivamente. En las comparaciones por pares se aplicó la corrección de Bonferroni para evitar errores de tipo I.

4. RESULTADOS

En este apartado se recoge, en primer lugar, los resultados relativos a la complejidad fraseológica de las colocaciones verbo-sustantivo(objeto). Seguidamente, se incluyen los resultados relativos a la complejidad fraseológica de las colocaciones sustantivo-adjetivo.

4.1. Complejidad fraseológica de las colocaciones verbo-sustantivo(objeto) en textos narrativos y expositivos por nivel de competencia

En la Tabla 3 se muestra la estadística descriptiva de las medidas de complejidad fraseológica analizadas de las colocaciones verbo-sustantivo(objeto) en los textos narrativos y expositivos en L1 y L2. En la Figura 1 se ilustran gráficamente los índices de diversidad y sofisticación por nivel de competencia en ambos tipos de texto. Los resultados descriptivos indican, por una parte, que tanto la diversidad como la sofisticación de estas colocaciones son similares en los dos tipos de textos

en todos los niveles de competencia y, por otra, que ambas aumentan con el nivel de competencia. A juzgar por las desviaciones típicas de las medidas de complejidad fraseológica, la diversidad de las colocaciones presenta una baja variabilidad en todos los grupos analizados, aunque es algo superior en los hablantes de nivel intermedio. En contraste, la sofisticación presenta una alta variabilidad, sobre todo en los grupos de hablantes no nativos.

Medida	Nivel	M (DT)	Narrativo M (DT)	Expositivo M (DT)
Diversidad	Intermedio	1,09 (0,69)	1,17 (0,65)	1,02 (1,05)
	Avanzado	1,55 (0,62)	1,58 (0,69)	1,52 (0,56)
	Nativo	2,54 (0,60)	2,53 (0,74)	2,55 (0,47)
Sofisticación	Intermedio	5,96 (4,38)	7,03 (3,74)	4,89 (4,88)
	Avanzado	7,86 (2,38)	7,81 (2,47)	7,9 (2,35)
	Nativo	8,26 (1,17)	7,91 (1,36)	8,61 (0,88)

Tabla 3. Estadística descriptiva (Media (M) y Desviación típica (DT)) de la diversidad y la sofisticación de las colocaciones verbo-sustantivo en textos narrativos y expositivos por nivel de competencia

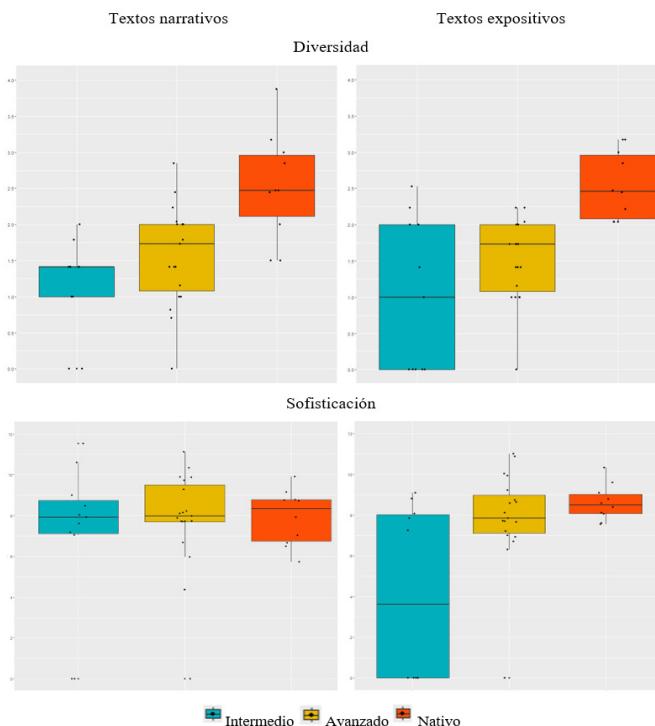


Figura 1. Distribución de la diversidad y la sofisticación de las colocaciones verbo-sustantivo en textos narrativos y expositivos por nivel de competencia

Los resultados del ANOVA factorial para la diversidad indican que se da un efecto simple del Nivel de competencia, $Q_{(2,13,36)} = 19,77, p < ,001$, en el sentido de que a medida que aumenta el nivel de competencia, aumenta la diversidad léxica. Los análisis *post-hoc* indican que se dan diferencias en todas las comparaciones por pares de modo que la diversidad de las colocaciones verbo-sustantivo aumenta progresivamente con el nivel de competencia (en todas las comparaciones $p < ,016$). En el caso de las diferencias en función del tipo de texto, no encontramos un efecto simple del Tipo de texto ni tampoco una interacción significativa entre este factor y el Nivel de competencia, de esta forma los resultados indican que la diversidad de este tipo de colocaciones es comparable en textos narrativos y expositivos.

Por su parte, los resultados del ANOVA para la sofisticación de las colocaciones indican que en este caso también se da un efecto simple del Nivel de competencia, $Q_{(2,74)} = 4,14, p = ,011$, por el cual aumenta la sofisticación léxica de las colocaciones con el nivel de competencia. En este caso, sin embargo, los análisis *post-hoc* indican que se dan diferencias en la comparación del grupo de nivel intermedio con el grupo de nivel avanzado ($p = ,013$) y con el grupo de nativos ($p = ,011$), en el sentido de que los hablantes más expertos incorporan en sus textos colocaciones más sofisticadas, mientras que las diferencias entre los hablantes de nivel avanzado y los nativos no resultan significativas. Tampoco en este caso se da un efecto simple del Tipo de texto ni se observa una interacción significativa entre ambos factores. Esto supone que, de nuevo, tampoco hay diferencias en la sofisticación de estas colocaciones en los textos narrativos y expositivos.

4.2. Complejidad fraseológica de las colocaciones sustantivo-adjetivo en textos narrativos y expositivos por nivel de competencia

La Tabla 4 recoge la estadística descriptiva de las medidas de complejidad fraseológica analizadas de las colocaciones de sustantivo-adjetivo. En la Figura 2 se ilustran gráficamente los índices de diversidad y sofisticación por nivel de competencia en ambos tipos de texto. Al igual que en las colocaciones verbo-sustantivo, la complejidad fraseológica de estas colocaciones tampoco parece depender del tipo de texto, pues los valores de diversidad y sofisticación léxicas son similares en textos narrativos y expositivos. Se observan, eso sí, patrones de desarrollo distintos por nivel de competencia para cada una de las dimensiones analizadas. Así, al igual que en las colocaciones de verbo-sustantivo, la diversidad de estas colocaciones también aumenta en función del nivel de competencia. En cuanto a la sofisticación, los resultados descriptivos muestran diferencias entre hablantes nativos

y no nativos; sin embargo, los índices medios de sofisticación son similares entre los hablantes no nativos. En ambos casos, las desviaciones típicas son más altas que las observadas en las colocaciones de verbo-sustantivo. Esta variabilidad se acentúa sobre todo en la sofisticación de las colocaciones de los hablantes no nativos. Por el contrario, los hablantes nativos muestran un uso más homogéneo, pues los valores de la desviación típica son más bajos que los observados en los otros grupos.

Medida	Nivel	M (DT)	Narrativo M (DT)	Expositivo M (DT)
Diversidad	Intermedio	0,69 (0,63)	0,56 (0,56)	0,83 (0,69)
	Avanzado	0,91 (0,74)	0,83 (0,81)	1 (0,68)
	Nativo	1,53 (0,55)	1,24 (0,55)	1,81 (0,39)
Sofisticación	Intermedio	4,23 (3,83)	3,71 (3,9)	4,75 (3,87)
	Avanzado	4,97 (3,87)	3,95 (3,63)	5,99 (3,93)
	Nativo	6,85 (2,36)	6,62 (3,24)	7,08 (1,09)

Tabla 4. *Estadística descriptiva (Media (M) y Desviación típica (DT) de la diversidad y la sofisticación de las colocaciones verbo-sustantivo en textos narrativos y expositivos por nivel de competencia*

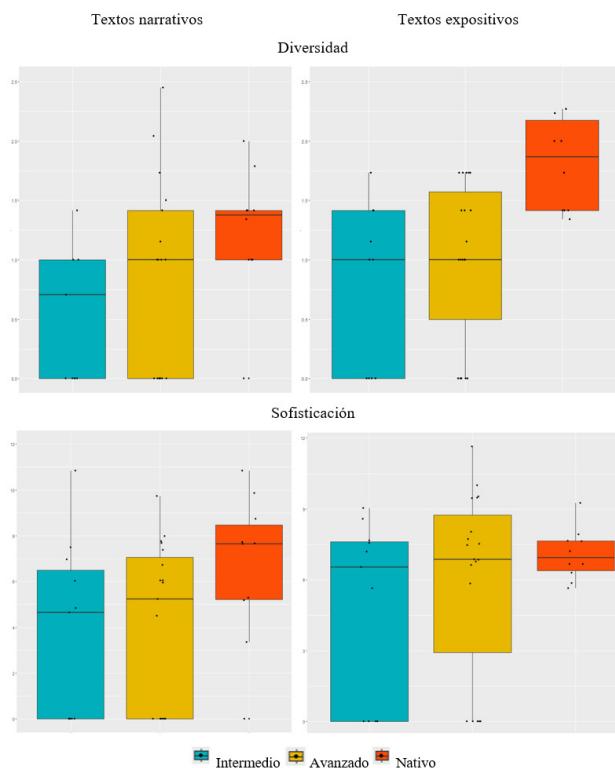


Figura 2. Distribución de la diversidad y la sofisticación de las colocaciones sustantivo-adjetivo en textos narrativos y expositivos por nivel de competencia

Los resultados del ANOVA factorial de la diversidad de las colocaciones de sustantivo-adjetivo indican un efecto simple del Nivel de competencia, $Q_{(2, 12, 94)} = 8,36$, $p = ,004$, en el sentido de que a medida que aumenta el nivel de competencia aumenta también la variedad de colocaciones utilizadas. En este caso, y a diferencia de lo observado en las colocaciones de verbo-sustantivo, los análisis *post-hoc* indican que solo hay diferencias significativas en las comparaciones entre los hablantes nativos y los grupos de nivel intermedio ($p = ,003$) y avanzado ($p = ,007$), mientras que las diferencias entre los dos grupos de no nativos no resultan significativas. Al igual que en las colocaciones de verbo-sustantivo, tampoco se da un efecto simple del Tipo de texto ni una interacción significativa entre ambos factores. De este modo, la diversidad de las colocaciones de sustantivo-adjetivo también se mantiene comparable en los textos narrativos y expositivos.

En cambio, y a diferencia de lo observado para las colocaciones de verbo-sustantivo, los resultados del ANOVA factorial de la sofisticación de estas colocaciones de sustantivo-adjetivo indican que no hay efectos simples del Nivel de competencia ni del Tipo de texto. Tampoco se ha encontrado una interacción significativa entre ambos factores. Estos resultados indican, por tanto, que la sofisticación de las colocaciones sustantivo-adjetivo es similar en todos los niveles de competencia analizados y que se mantiene comparable en los textos narrativos y expositivos.

5. DISCUSIÓN

En este estudio hemos analizado dos propiedades fundamentales de la complejidad fraseológica: la diversidad y la sofisticación léxica de las colocaciones en producciones escritas de estudiantes universitarios que son hablantes nativos y no nativos de español y que, en el caso de estos últimos, presentan diferentes niveles de competencia. Estas propiedades se han examinado en los dos tipos de colocaciones más frecuentes en español: verbo-sustantivo(objeto) y sustantivo-adjetivo, en dos tipos de textos escritos: narrativos y expositivos.

Según nuestros resultados, y en línea con lo observado en estudios previos, la complejidad fraseológica aumenta con el grado de dominio lingüístico, por cuanto los hablantes más experimentados producen una mayor diversidad de colocaciones, las cuales son, además, más sofisticadas. Este desarrollo se da con independencia del propósito comunicativo, pues no encontramos diferencias significativas en función del tipo de texto, y el índice de diversidad y sofisticación de las colocaciones utilizadas se mantienen en los textos narrativos y expositivos en todos los grupos

de hablantes estudiados. Estos resultados proporcionan evidencia adicional de que el uso de colocaciones constituye un indicador fundamental de competencia lingüística en L2 (Cuberos, 2019; Cuberos et al., 2024; Fan, 2009; García-Salido et al., 2019), y esto es así con independencia del género discursivo (Cuberos, 2019). Los patrones de desarrollo identificados sí varían según la estructura sintáctica de las colocaciones. Este resultado, atestiguado en estudios previos (Paquot, 2018; 2019; Vandeweerd et al., 2021), subraya la importancia de considerar la naturaleza sintáctica de las colocaciones para analizar apropiadamente el desarrollo de la competencia fraseológica en L2. A continuación, discutimos en primer lugar los resultados relativos a las colocaciones verbo-sustantivo y, seguidamente, hacemos lo propio con las de sustantivo-adjetivo.

En relación con las colocaciones de verbo-sustantivo, los resultados indican que la complejidad fraseológica aumenta con el nivel de competencia si bien se observan patrones de desarrollo distintos para cada una de las dimensiones analizadas. Así, la diversidad de las colocaciones utilizadas muestra un aumento progresivo con diferencias significativas en todas las comparaciones por pares mientras que, en el caso de la sofisticación, se observan diferencias significativas en la comparación entre los hablantes no nativos de niveles intermedios con los de nivel avanzado y los nativos, entre los cuales no hay diferencias. Un primer análisis pormenorizado de estos resultados nos lleva a hipotetizar que algunas de estas colocaciones incluyen lo que la literatura previa (véase, por ejemplo, García Salido, 2014; 2016; Pérez-Serrano y Mavrou, 2022) denomina construcciones con verbo de apoyo o verbo ligero, como *hacer fotos*, *dar ejemplo* o *tener confianza*, es decir, colocaciones en las que el verbo está deslexicalizado y el contenido semántico lo aporta principalmente el sustantivo. Ejemplos de este tipo de construcciones son *hacer fotos*, *dar ejemplo* o *tener confianza*. Estudios previos señalan que los aprendices de L2 tienden a utilizar estas construcciones con verbo de apoyo con más frecuencia que los nativos; estos, por su parte, disponen de un repertorio más variado de este tipo de formas (García Salido, 2014). Asimismo, en un trabajo anterior en el mismo corpus analizado en este estudio (Cuberos et al., 2024), observamos que hablantes nativos y no nativos de distintos niveles de competencia utilizan frecuentemente estas construcciones, pero que los aprendices más experimentados incorporan de forma recurrente otras combinaciones con verbos con significado pleno (p. ejemplo, *acarrear consecuencias*). Dado el papel central que parecen desempeñar estas construcciones en el desarrollo de la competencia fraseológica en L2, estas merecen, sin duda, un análisis pormenorizado en próximos estudios.

En el caso de las colocaciones sustantivo-adjetivo encontramos un escenario ligeramente distinto. Si bien los nativos utilizan un repertorio más variado de este tipo de colocaciones que los no nativos de distintos niveles de competencia, entre los que no se dan diferencias significativas, la sofisticación de estas construcciones es similar en todos los grupos analizados, pues no hay un efecto de nivel de lengua. Estos resultados sugieren que, a diferencia de lo que ocurre con las colocaciones de verbo-sustantivo, el nivel de competencia parece jugar aquí un papel menos relevante. Este comportamiento diferenciado según la estructura sintáctica, y atestiguado también en otras lenguas como inglés (Paquot, 2019) y francés (Vandeweerd et al., 2021), puede deberse, por una parte, a que se trata de colocaciones menos frecuentes en español (Cuberos et al., 2024; Koike, 2001), lo que podría condicionar su desarrollo incluso en niveles avanzados. Por otra parte, la propia naturaleza sintáctica de estas colocaciones podría también ser responsable de la ausencia de diferencias significativas entre los grupos.

En las colocaciones sustantivo-adjetivo, la coaparación de ambos constituyentes parece tener, en algunos casos, un nivel de fijación menor de tal manera que presentan, de algún modo, mayor opcionalidad que las combinaciones verbo-sustantivo, especialmente las de verbo de apoyo. Así, mientras el sustantivo y el verbo se encuentran aquí en una relación de interdependencia en virtud de la cual, por así decirlo, el sustantivo depende gramaticalmente del verbo y el verbo necesita semánticamente al sustantivo, en las colocaciones sustantivo-adjetivo, este último aporta, por definición, información adicional. Este carácter opcional otorgaría a los adjetivos una mayor flexibilidad en su uso. En futuros trabajos sería, sin duda, conveniente profundizar en la naturaleza semántica y el comportamiento discursivo de este tipo de colocaciones.

Tomados en conjunto, los resultados de este estudio subrayan, por una parte, la naturaleza multifacética de las colocaciones y apuntan a que las propiedades definitorias de las mismas interactúan entre sí para describir el desempeño lingüístico en L2 (Paquot, 2019). Además, confirman la idoneidad de considerar la diversidad de las colocaciones utilizadas en un texto para analizar y caracterizar el desarrollo lingüístico en L2 (Cuberos et al., 2024) y apuntan, como novedad, a la relevancia que supone examinar la estructura sintáctica de estos elementos. Concretamente, el análisis de las colocaciones verbo-sustantivo resulta especialmente relevante para capturar diferencias en el desarrollo de la competencia fraseológica como parte de la competencia léxica de los aprendices. Por último, nos parece importante señalar cómo, en futuros trabajos, sería pertinente analizar su comportamiento en

relación con otros indicadores de desarrollo lingüístico, tales como la complejidad sintáctica o la densidad discursiva (Paquot, 2019). Igualmente, se podrían estudiar otras formas de medir la complejidad fraseológica, por ejemplo, explorando otros índices de sofisticación.

DECLARACIÓN DE CONTRIBUCIÓN DE AUTORÍA

Las autoras han contribuido de manera equitativa y significativa a la elaboración del artículo. Las tareas principales realizadas por cada autora se listan a continuación:

Rocío Cuberos Vicente: Conceptualización, Investigación, Metodología, Diseño de instrumento de identificación de colocaciones, Validación del instrumento de identificación de colocaciones, Análisis formal de datos, Redacción–borrador original, Redacción–revisión y edición.

Elisa Rosado Villegas: Conceptualización, Investigación, Metodología, Diseño de instrumento de identificación de colocaciones, Validación del instrumento de identificación de colocaciones, Supervisión, Validación, Redacción–revisión y edición.

Ambas autoras han participado en la revisión crítica del documento, aprobaron la versión final del manuscrito y asumen responsabilidad por su contenido.

FINANCIACIÓN

La recogida de los datos utilizados en esta investigación ha sido financiada por el proyecto «El desarrollo de la calidad textual en castellano y catalán L2. Indicadores léxicos, morfosintácticos y discursivos» (I.P.: J. Perera, Universitat de Barcelona) (EDU2012-34394, Ministerio de Economía y Competitividad). El tratamiento y análisis de los datos ha sido financiado por la beca postdoctoral Margarita Salas otorgada a R. Cuberos por la Universitat de Barcelona (Next Generation EU y el Ministerio de Universidades).

REFERENCIAS

- Aparici, M. (2010). *El desarrollo de la conectividad discursiva en diferentes géneros y modalidades de producción* [Tesis doctoral inédita, Universitat de Barcelona].
- Bardovi-Harlig, K., y Stringer, D. (2017). Unconventional expressions: Productive syntax in the L2 acquisition of formulaic language. *Second Language Research*, 33(1), 61-90. doi: <https://doi.org/10.1177/02676583166417>
- Berman, R. y Verhoeven, L. (2002). Cross-linguistic perspectives on the development of text production abilities: Speech and writing. *Written Language and Literacy*, 5(2), 1-43. doi: <https://doi.org/10.1075/wll.5.1.02ber>
- Bosque, I. (2001). Sobre el concepto de ‘colocación’ y sus límites. *Lingüística Española Actual*, 23(1), 9-40.
- Bosque, I. (2006). *Diccionario combinatorio práctico del español contemporáneo*. SM.

- Bosque, I. (2011). Deducing collocations. En I. Boguslavsky y L. Wanner (Eds.), *Proceedings of the 5th International Conference on Meaning-Text Theory* (pp. 6-13).
- Bosque, I. (2017). On the conceptual bases of collocations: Restricted adverbs and lexical selection. En S. Torner y E. Bernal (Eds.), *Collocations and other lexical combinations in Spanish: Theoretical, lexicographical and applied perspectives* (pp. 9-20). Routledge.
- Boone, G., y Eyckmans, J. (2023). Exploring collocation development in L2 German from students' perspective: A contrasting case study. *Studies in Second Language Learning and Teaching*, 13(3), 571-599. doi: <https://doi.org/10.14746/ssllt.32539>
- Corpas, G. (1996). *Manual de fraseología española*. Gredos.
- Corpas, G. (2001). Apuntes para el estudio de la colocación. *Lingüística Española Actual*, 23(1), 41-56.
- Cowie, A. (1998). *Phraseology: Theory, analysis and applications*. Oxford University Press.
- Crossley, S., Cai, Z. y McNamara (2012). Syntagmatic, paradigmatic, and automatic N-gram approaches to assessing essay quality. En *Proceedings of the 25th International Florida Artificial Intelligence Research Society Conference, FLAIRS-25*, 214-219.
- Cuberos, R. (2019). *Indicadores léxicos de calidad textual en español nativo y no nativo* [Tesis doctoral, Universitat de Barcelona]. doi: <https://deposit.ub.edu/dspace/handle/2445/178686>
- Cuberos, R., Rosado, E. y Mañas, I. (2024). Towards a graded lexical inventory of multi-word combinations. The distribution of collocations across Spanish CEFR-levels. *International Journal of Applied Linguistics*, 175(1), 46-76. doi: <https://doi.org/10.1075/itl.22021.cub>
- Durrant, P. y Schmitt, N. (2009). To what extent do native and non-native writers make use of collocations? *International Review of Applied Linguistics in Language Teaching*, 47, 157-177. doi: <https://doi.org/10.1515/iral.2009.007>
- Ellis, N.C., Simpson-Vlach, R., Romer, U., Brook O'Donnell, M. y Wulff, S. (2015). Learner corpora and formulaic language in second language acquisition. In S. Granger, G. Gilquin y F. Meunier (Eds.), *The Cambridge handbook of learner corpus research* (pp. 357-378). Cambridge University Press. doi: <https://doi.org/10.1017/CBO9781139649414.016>
- Fan, M. (2009). An exploratory study of collocational use by ESL students. A task based approach. *System*, 37, 110-123. doi: <https://doi.org/10.1016/j.system.2008.06.004>
- Firth, J. (1957). *Papers in Linguistics 1934-1951*. Oxford University Press.
- Gablasova, D., Brezina, V. y McEnery, T. (2017). Exploring learner language through corpora: Comparing and interpreting corpus frequency information. *Language Learning*, 67(21), 130-154.
- García-Salido, M. (2016). Error analysis of support verb constructions in written Spanish learner corpora. *The Modern Language Journal*, 100, 362-376. doi: <https://doi.org/10.1111/modl.12320>
- García-Salido, M., García, M. y Alonso-Ramos, M. (2019). Towards a graded dictionary of Spanish collocations. En I. Kosem, T. Zingano, M. Correira, J.P. Ferreira, M.

- Jansen, I. Pereira, J. Kallas, M. Jakubíček, S. Krek y C. Tiberius (Eds.), *Proceedings of eLex 2019 conference* (pp. 849–864).
- García-Salido, M., Lorente, P. y Basanta, A. (2019). Las construcciones con verbos de apoyo del español en la producción escrita de aprendices francófonos. *Journal of Spanish Language Teaching*, 6(1), 32–48.
doi: <https://doi.org/10.1080/23247797.2019.1590022>
- Gitsaki, C. (1996). *The development of ESL collocational knowledge* [Doctoral dissertation, The University of Queensland]. doi: <https://doi.org/10.14264/205374>
- Granger, S. y Paquot, M. (2008). Disentangling the phraseological web. En S. Granger, y F. Meunier (Ed.), *Phraseology: An interdisciplinary perspective* (pp. 28–49). John Benjamins. doi: <https://doi.org/10.1075/z.139.07gra>
- Halliday, M. A. (1961). *Categories of the theory of grammar. Department of English Language and General Linguistic Monographs* (pp. 241–292). Edinburgh University Press.
- Hausmann, F. (1979). Un dictionnaire des collocations est-il possible? *Travaux de Linguistique et de Littérature*, 17(1), 187–195
- Hausmann, F. (1998). O diccionario de colocacións: criterios de organización. In X. Ferro Ruibal (Ed.), *Actas do I Coloquio Galego de Fraseoloxía* (pp. 63–81). Centro Ramón Piñeiro, Xunta de Galicia.
- Higueras, M. (2006). *Las colocaciones y su enseñanza en la clase de ELE*. Arco Libros.
- Howarth, P. (1996). *Phraseology in English academic writing: Some implications for language learning and dictionary making*. Max Niemeyer Verlag.
doi: <https://doi.org/10.1515/9783110937923>
- Hyland, K. (2008). As can be seen: Lexical bundles and disciplinary variation. *English for Specific Purposes*, 27(1), 4–21. doi: <https://doi.org/10.1016/j.esp.2007.06.001>
- Kecskes, I. (2007). Formulaic language in English Lingua Franca. En I. Kecskes y L. R. Horn (Eds.), *Explorations in Pragmatics* (pp. 191–218). Mouton de Gruyter.
doi: <https://doi.org/10.1515/9783110198843.3.191>
- Kilgarriff, A. (2005). Language is never, ever, ever, random. *Corpus Linguistics and Linguistic Theory*, 1(2), 263–276. doi: <https://doi.org/10.1515/cllt.2005.1.2.263>
- Kilgarriff, A. y Renau, I. (2013). esTenTen, a vast web corpus of Peninsular and American Spanish. *Procedia-Social and Behavioral Sciences*, 95, 12–19.
doi: <https://doi.org/10.1016/j.sbspro.2013.10.617>
- Kilgarriff, A., Baisa, V., Busta, J., Jakubicek, M., Kovar, V., Michelfet, J., Rychlý, P. y Suchomel, V. (2014). The Sketch Engine: ten years on. *Lexicography ASIALEX*, 1, 7–36.
- Koike, K. (2001). *Colocaciones léxicas en el español actual: Estudio formal y léxicosemántico*. Universitat de Alicante y Takushoku University.
- Laso, N. (2002). *A corpus-based study of phraseological behaviour of abstract nouns in medical English: a needs analysis of a Spanish medical community*. [Tesis doctoral, Universitat de Barcelona]. doi: <http://hdl.handle.net/10803/1671>
- MacWhinney, B. (2000). *The CHILDES Project: Tools for analyzing talk* (3rd Ed.). Lawrence Erlbaum.
- Mair, P. y Wilcox, R. (2016). *WRS2: A collection of robust statistical methods*.
doi: <https://CRAN.R-project.org/package=WRS2>

- Mel'čuk, I. (2012). Phraseology in the language, in the dictionary, and in the computer. *Yearbook of Phraseology*, 3, 31–56. doi: <https://doi.org/10.1515/phras-2012-0003>
- Mittmann, B. (2013). New tendencies in the treatment of collocations. En R. H. Gouws, U. Heid, W. Schweickard y H. E. Wiegand (Eds.), *Supplementary volume dictionaries. An International Encyclopedia of Lexicography* (pp. 500–509). De Gruyter Mouton. doi: <https://doi.org/10.1515/9783110238136.500>
- Nesselhauf, N. (2004). What are collocations? En D. Allerton, N. Nesselhauf y P. Skandera (Eds.), *Phraseological units: Basic concepts and their application* (pp. 1–21). Schwabe.
- Paquot, M. (2018). Phraseological competence: A missing component in university entrance language tests? Insights from a study of EFL learners' use of statistical collocations. *Language Assessment Quarterly*, 15(1), 29–43. doi: <https://doi.org/10.1080/15434303.2017.1405421>
- Paquot, M. (2019). The phraseological dimension in interlanguage complexity research. *Second Language Research*, 35(1), 121–145. doi: <https://doi.org/10.1177/0267658317694221>
- Pérez-Serrano, M. y Mavrou, I. (2022). La incidencia del tipo de ejercicio y de la capacidad de la memoria operativa en el aprendizaje de las construcciones con verbo de apoyo en español como segunda lengua. *Onomázein*, 58, 47–68. doi: <https://doi.org/10.7764/onomazein.58.03>
- Pérez-Serrano, M. y Hidalgo-Gallardo, M. (en prensa). ¿Son *da un paseo* y *daba un paseo* variantes de la misma colocación o colocaciones distintas? Variación flexiva y procesamiento de las colocaciones En J. Amenós Pons, C. Chacón García y M. Pérez Serrano (Eds.) *Léxico y sintaxis en la adquisición de español como segunda lengua: perspectivas actuales*. Ediciones Complutense.
- Schmitt, N. (2004). *Formulaic sequences: Acquisition, processing and use*. John Benjamins.
- Sinclair, J. (1987). *Looking up – An account of the COBUILD Project in lexical computing*. Harper-Collins.
- Sinclair, J. (1991). *Corpus, concordance, collocation*. Oxford University Press.
- R Core Team (2021). *R: A language environment for statistical computing*. Version 4.1.2.
- Ravid, D. y Tolchinsky, L. (2002). Developing linguistic literacy: a comprehensive model. *Journal of Child Language*, 29(2), 417–447. doi: <https://doi.org/10.1017/S0305000902005111>
- Real Academia Española. (2014). *Diccionario de la lengua española* (23rd Ed.).
- Rosado, E., Salas, N., Aparici, M. y Tolchinsky, L. (2014). Production of linguistic devices for attaining a detached stance in Spanish and Catalan. *Journal of Pragmatics*, 60, 36–53. doi: <https://doi.org/10.1016/j.pragma.2013.10.004>
- Rubin, R., Housen, A. y Paquot, M. (2021). Phraseological complexity as an index of L2 Dutch writing proficiency: A partial replication study. En Granger, S. (Ed.), *Perspectives on the second language phrasicon: The view from learner corpora* (pp. 101–125). Multilingual Matters. doi: <https://doi.org/10.21832/9781788924863-006>

- Rychlý, P. (2008). A lexicographer-friendly association score. En P. Sojka y A. Horák (Eds.) *Proceedings of Recent Advances in Slavonic Natural Language Processing, RASLAN 2008* (pp. 6 - 9). University de Masaryk.
- Tolchinsky, L. y Aparici, M. (2015). *Spanish GRERLI-ES1 corpus* [Conjunto de datos]. CHILDES Talkbank. doi: <https://doi.org/10.21415/T53G63>
- Tolchinsky, L., Rosado, E., Aparici, M. y Cubero, R. (2023). *Spanish GRERLI-ES2 Corpus* [Conjunto de datos]. CHILDES Talkbank.
doi: <https://doi.org/10.21415/2DJF-JH34>
- Vandeweerd, N., Housen, A. y Paquot, M. (2021). Applying phraseological complexity measures to L2 French: A partial replication study. *International Journal of Learner Corpus Research*, 7(2), 197-229.
- Vilkaitė-Lozdienė, L. (2022). Do different morphological forms of collocations show comparable processing facilitation? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 48(9): 1328-1347. doi: <https://doi.org/10.1037/xlm0001130>
- Wood, D. (2020). Classifying and identifying formulaic language. En S. Webb (Ed.), *The Routledge Handbook of Vocabulary Studies* (pp. 30-45).
<https://doi.org/10.4324/9780429291586>
- Wray, A. (2002). *Formulaic language and the lexicon*. Cambridge University Press.
doi: <https://doi.org/10.1017/CBO9780511519772>

