

Análisis de colocaciones producidas por aprendices de ELE a partir de un corpus

MARÍA FERNANDA RODRÍGUEZ FARAH

Universidad Autónoma de Baja California

maria.fernanda.rodriguez.farah@uabc.edu.mx

ELEONORA LOZANO BACHIOQUI

Universidad Autónoma de Baja California

eleonoralozano@uabc.edu.mx

ANTONIO REYES PÉREZ

Universidad Autónoma de Querétaro

antonio.reyesp@uaq.mx

Resumen: Este estudio presenta un análisis sobre el aprendizaje y uso comunicativo de colocaciones de aprendices de español como lengua extranjera (ELE). Para efectuar este análisis, se diseñó un instrumento que se conforma de un cuestionario y un corpus denominado Corpus de Aprendices de Colocaciones ELE (CACE), cuyo objetivo es identificar la producción de colocaciones, errores comunes y patrones de uso que muestren evidencia de la verbalización de estos fenómenos. Así, se presenta una fundamentación teórica sobre lingüística de corpus y las colocaciones, así como una clasificación de la taxonomía de colocaciones en materia de ELE. También se describe el proceso que se llevó a cabo para la obtención de datos y la recopilación del corpus, así como la manera en la que se procesaron los datos. Por último, se presenta un análisis y discusión de los datos obtenidos, así como conclusiones sobre lo observado y algunas limitaciones del estudio.

Palabras clave: colocaciones; corpus de aprendices; español como lengua extranjera; lingüística de corpus

Analysis of collocations produced by SFL learners through a corpus

Abstract: This article analyses the learning and communicative use of collocations by learners of Spanish as a Foreign Language (SFL). A multiple-choice test and a corpus were designed to carry out the analysis. The Spanish as a Foreign Language Learner Corpus (CACE, for its acronym in Spanish) aims to identify collocational issues such as use, common errors, and use patterns that provide evidence of how these phenomena are verbalized. Thus, a theoretical foundation on corpus linguistics, collocations, and classification of SFL collocations are presented. The data collection process, the compilation of the corpus, and the way the texts were classified are also described. Finally, an analysis and discussion of the data obtained is presented, as well as observed conclusions and some limitations of the study.

Key words: Collocations; learner corpus; Spanish as a foreign language; corpus linguistics

1. Introducción

Varios investigadores, como Corder (1967), defienden los efectos positivos de observar el desarrollo lingüístico dentro del aula de lenguas extranjeras. Analizar el desarrollo lingüístico de los aprendientes de lenguas llega a ser de gran utilidad para docentes e investigadores, pues permite visualizar el progreso de desempeño del estudiante. En la producción de colocaciones, es común que el estudiante requiera de instrucción específica para su aprendizaje, pues su naturaleza restrictiva por lo regular impide que exista una transferencia entre su lengua nativa (L1) y la lengua extranjera meta (L2).

El presente trabajo tiene como objetivo mostrar un análisis y discusión sobre la producción de colocaciones por parte de aprendices de español como lengua extranjera (ELE), a partir de datos obtenidos de un corpus de aprendices diseñado para esta investigación; el cual se compiló a partir de la recolección de muestras de producción escrita y, además de presentar evidencia empírica sobre el fenómeno, sirve como fuente de información sobre distintos fenómenos en el campo de ELE. Este recurso ha sido denominado el Corpus de Aprendices de Colocaciones de ELE (CACE) (Rodríguez Farah, 2022) mismo que puede resultar de utilidad para docentes y aprendientes de ELE.

1.1. Objetivo

De acuerdo con Naeimi y Foo (2015) el aprendizaje de vocabulario es una de las habilidades más básicas para la adquisición de un idioma. Sin embargo, según los autores, los aprendientes de lenguas extranjeras lo consideran el área de mayor dificultad en su aprendizaje. Para Ellis (1997), el aprendizaje de palabras individuales no es suficiente, dando mayor importancia a las estructuras léxicas y la relación que unas palabras tienen con otras, tal como las frases léxicas y las colocaciones. Así mismo, hace una distinción entre el aprendizaje de vocabulario y del discurso. Mientras que la adquisición de vocabulario involucra secuenciar las propiedades fonológicas de una lengua, el aprendizaje del discurso involucra secuenciar las unidades léxicas de una lengua (como las frases y las colocaciones).

Entre los estudios que se han hecho en materia de vocabulario y, específicamente, en materia de aprendizaje y producción de colocaciones de estudiantes de ELE, Travalia (2006) menciona que el conocimiento colocacional ofrece ventajas a los aprendientes, como el incremento de vocabulario y una mayor fluidez en la lengua meta. De acuerdo con la misma autora, los aprendientes necesitan saber cómo crear sintagmas en ese idioma, y son las colocaciones las que les pueden servir de modelo en esa tarea. Se argumenta que los sintagmas varían entre distintos idiomas, por lo que no pueden recurrir a su L1 fácilmente para comprender estas construcciones.

Desde la semántica, las colocaciones representan una variedad de dificultades en el proceso de enseñanza-aprendizaje de segundas lenguas, que se encuentran en la combinatoria y no tanto en su significado, pues se consideran relativamente transparentes (Corpas Pastor, 1996). Higuera (2006) declara que es esta transparencia

de significado, y el hecho de tratarse de palabras individuales familiares, lo que hace que las colocaciones pasen desapercibidas por muchos estudiantes.

A partir de lo anterior, surge la pregunta de cómo se puede apoyar la producción de colocaciones en el aula de ELE. Por tanto, el objetivo de este trabajo es analizar cómo producen las colocaciones los aprendices de ELE, considerando aspectos de la interlengua, gramaticales, metafóricos, entre otros, y según la taxonomía propuesta por Corpas Pastor (1996), Koike (2001) y Travalía (2006), descrita en el apartado 2.1.

1.2 Lingüística de corpus en la enseñanza de lenguas

En esta sección se describirán algunos de los varios trabajos de lingüística de corpus en función de la enseñanza de lenguas. En particular, aquellos que se centran en el fenómeno de las colocaciones. Según Rica Peromingo (2010), el campo de investigación de la lingüística de corpus se ocupaba, desde sus inicios, tanto de estudiar unidades léxicas basadas en su significado como en la coaparición con otros términos. El mismo autor declara que en los últimos años se ha probado, en el campo de la adquisición de lenguas extranjeras, que el uso de corpus lingüísticos es una herramienta muy eficaz. Además, como sugiere Conrad (2002), estos últimos han sido de gran utilidad para determinar patrones de comportamiento de la lengua, para reconocer selecciones típicas o inusuales de los usuarios, y describir la interacción entre varios fenómenos de producción de la L2.

De acuerdo con Sinclair (1991), un corpus es una colección de textos de lenguaje natural, los cuales han sido seleccionados para caracterizar el estado o la variedad de una lengua. Adicionalmente, Sinclair, en 1996, trabajando con un grupo dedicado a los corpórea textuales llamado *Expert Advisory Group on Language Engineering Standards (EAGLES)*, por sus siglas en inglés) sostiene que un corpus es una colección de muestras de lenguaje seleccionadas y ordenadas de acuerdo con criterios lingüísticos explícitos con la finalidad de ser utilizado como muestra de lenguaje.

Los corpus lingüísticos han jugado un papel importante en los estudios de léxico. La base de estos estudios son las colocaciones y los grupos léxicos (Schmitt, 2000). Por ejemplo, Vincze et al. (2016) llevaron a cabo una investigación para analizar la producción de colocaciones de aprendices de ELE. Dentro de los resultados se observó tanto la cantidad de colocaciones producidas por los aprendices, como un análisis de los errores en dichas colocaciones. Por su parte, De Alba Quiñones (2012) presenta un estudio sobre el subsistema léxico-semántico de aprendices de español como L2. Para ello, se analizaron producciones escritas de interlengua de aprendices alemanes. Así, la autora realizó un primer acercamiento desde el análisis de errores para el estudio de las lexías simples y, posteriormente, desarrolló un análisis de la actuación sobre un tipo de lexía compleja, las colocaciones.

2. Fenómeno de estudio

El término colocación se puede clasificar como polisémico y distintos autores lo han abordado desde distintas disciplinas. A continuación, se describen algunas de las concepciones de este término. J.R. Firth fue uno de los primeros en acuñar el término

para hacer referencia a combinaciones comunes de unidades léxicas (Firth, 1957). El término de colocación es utilizado en distintos sentidos. En ocasiones, el autor hace referencia a combinaciones comunes o probables entre palabras; en otras, se refiere a combinaciones restringidas en donde un lexema requiere la presencia de otro. Para este autor británico, uno de los aspectos del significado de una palabra es el conjunto de todas las otras palabras con las que se combina. Coseriu (1981) distingue las solidaridades léxicas de los virtuales de Pottier (1969), las cuales son combinaciones usuales y frecuentes. Coseriu afirma que contar con dos lexemas que coocurren frecuentemente no es prueba de que exista una colocación.

Para Hausmann (1979), una colocación se define como una colocación restringida y orientada, compuesta por una base y un colocativo. Por ejemplo, en *célibataire endurci* la base es el nombre y el *colocativo*, el adjetivo. Según Hausmann, la primera es autónoma. Sin embargo, la definición del colocativo necesita ser completada por la base. Vincze et al. (2016) coinciden con Hausmann, agregando que la selección del colocativo está condicionada por las restricciones léxicas impuestas por la base. Será bajo esta concepción de colocación que se trabajará la presente investigación.

2.1 Tipología de las colocaciones

Las colocaciones en español se pueden clasificar en dos tipos: léxicas y gramaticales. Las primeras, a su vez, se distinguen por ser de carácter simple o complejo. Mientras que las colocaciones léxicas simples se conforman de dos unidades léxicas sencillas como *preparar el desayuno*, las complejas se conforman por una unidad léxica acompañada de una unidad fraseológica, o locución, como en *dormir como un tronco* (Koike, 2001). Por su parte, Travalia (2006) identifica las colocaciones gramaticales, describiéndolas como frases que consisten en una palabra de categoría abierta acompañados por una palabra de categoría cerrada. A continuación, se presenta una tabla con la taxonomía de colocaciones del español.

Colocaciones léxicas simples
Sustantivo (Sujeto)+ Verbo
Verbo+ Sustantivo (Objeto)
Adjetivo+ Sustantivo
Sustantivo+ Preposición+ Sustantivo
Verbo participio+ Adverbio
Verbo+ preposición+ Sustantivo
Sustantivo+ Adjetivo
Verbo+ Adverbio(mente)
Adverbio+ Adjetivo
Verbo+ Adjetivo
Colocaciones léxicas complejas
Verbo+ locución nominal
Locución verbal+ sustantivo
Verbo+ locución adverbial

 Adjetivo+ locución adverbial

Colocaciones gramaticales

verbo+ preposición

sustantivo+ preposición (+sustantivo/oración subordinada)

(artículo+) sustantivo + preposición (+infinitivo)

preposición (+artículo)+ sustantivo (+preposición)

(verbo)+ adjetivo+ preposición

(verbo + sustantivo)+ preposición+ sustantivo

La tabla 1 muestra la taxonomía se utilizó como base para el análisis de la presente investigación (Corpas Pastor, 1996; Koike, 2001; Travalía, 2006).

Fuente: Adaptado de las taxonomías de colocaciones propuestas por Copras Pastor (1996), Koike (2001), y Travalía (2006)

2.2. Corpus y colocaciones

En cualquier área de investigación en la que se utilice un corpus lingüístico, un aspecto fundamental es, como menciona Conrad (2002), que todo análisis lingüístico basado en corpus debe cumplir con algunos requisitos, que son: (1) que sea empírico y analice los patrones reales de uso en textos naturales; (2) que utilice una amplia colección de textos naturales como base del análisis; (3) que utilice software para efectuar análisis sobre los datos; y (4) que dependa de técnicas analíticas, tanto cuantitativas como cualitativas. Sierra (2015) concuerda con Conrad, agregando que es necesario trabajar con muestras reales de la lengua. Con base en ello, Sierra menciona que un análisis cualitativo permite estudiar las características de una lengua o algún fenómeno que ocurra en ella, mientras que el cuantitativo ofrece información sobre cifras numéricas y frecuencias del fenómeno observado. Así, Sierra propone una tipología para clasificar los distintos tipos de corpus de acuerdo con sus características y para los fines de este trabajo destacan los siguientes (1) según el origen de los datos, (2) según la temporalidad de los textos, (3) según el propósito de estudio, es decir, si son multipropósito o con propósitos específicos; (4) según la lengua, se clasifican en monolingües o multilingües; (5) según la cantidad de texto, ya sean corpus grandes o pequeños; (6) según la representatividad, los cuales se clasifican en corpus representativos u oportunistas (aquellos que no representan toda la lengua pero sí algún fenómeno en específico).

En materia de corpus pequeños, Ghadessy (2001) declara que el análisis de estos corpus es una poderosa herramienta para la enseñanza y aprendizaje del lenguaje natural. De acuerdo con Beaugrande (2001), estos corpus son de especial relevancia para proyectos que buscan analizar fenómenos específicos de la lengua.

En cuanto a los corpus de aprendices, Granger (2008) define un corpus de aprendices como una colección electrónica de textos producidos por aprendices de lenguas. De acuerdo con la autora, estos corpus cuentan con todas las características que regularmente se atribuyen a los corpus lingüísticos. Álvarez y Gil Martínez (2018)

agregan que los corpus de aprendices permiten tener acceso a la interlengua de quienes están en el proceso de aprendizaje del español y, por ello, permiten hacer contrastes sobre las dificultades de aprendizaje y el vocabulario empleado en la producción.

3. Metodología

3.1 Obtención de datos empíricos

Con la finalidad de poder observar el manejo del fenómeno colocacional por parte de los aprendientes de ELE, en primera instancia se creó un cuestionario que permitiera entender el conocimiento fraseológico de los no nativos hablantes del español, en función de la taxonomía presentada en la tabla 1. Y, en segunda instancia, se incluyó una sección de muestras de escritura que permitiera observar el uso de lenguaje natural y que, a la vez, permita conformar un corpus de aprendices. Ambas secciones fueron diseñadas para incluir elementos que proporcionaran datos de corte cuantitativo y cualitativo, de tal suerte que, por un lado, fuera posible medir la frecuencia y magnitud con la que ocurren los fenómenos observados (Hernández Sampieri et al., 2014); por el otro, observar ejemplos de uso sobre el conocimiento colocacional, logrando así observar el fenómeno desde distintas ópticas.

En la primera sección del cuestionario, se diseñaron 40 reactivos de opción múltiple. Cada reactivo se trata de una oración con un espacio en blanco, el cual debe ser llenado escogiendo una de tres opciones. En cada enunciado se proporcionó la base de una colocación y se dejó en blanco el espacio del colocativo para que los participantes escogieran la opción correcta de una lista de tres. A continuación, se presenta una tabla con algunos ejemplos de los reactivos del cuestionario.

Pregunta	Posibles respuestas
Se _____ un rumor falso en la escuela.	esparció dijo corrió
Laura fue invitada a _____ el cargo de director	desempeñar cumplir ejecutar
Escuché la noticia de una fuente _____.	veraz fidedigna verdadera

Nota: las respuestas marcadas en negrillas representan la respuesta correcta.

Tabla 2. Ejemplos de elementos integrados en el cuestionario

Todos los reactivos de esta sección, ejemplificados en la tabla 2, fueron diseñados según la taxonomía de colocaciones léxicas simples y complejas identificadas por Corpas Pastor

(1996) y Koike (2001), así como las gramaticales identificadas por Travalía (2006), mencionadas en el apartado 2.2. En total, se identificaron veinte categorías distintas entre las taxonomías de las autoras, como se muestra en la tabla 1, y se incluyeron dos reactivos por cada una de las categorías identificadas.

En la segunda sección del cuestionario, se proporcionó una pauta de escritura a los participantes, recuperada de la guía del *Certificado de español como lengua adicional* (CELA) de la Universidad Nacional Autónoma de México (UNAM, 2015), en la que se solicita a los sujetos que escriban un correo electrónico a un amigo. La instrucción indica que se debe redactar una disculpa por no haber podido asistir a una fiesta.

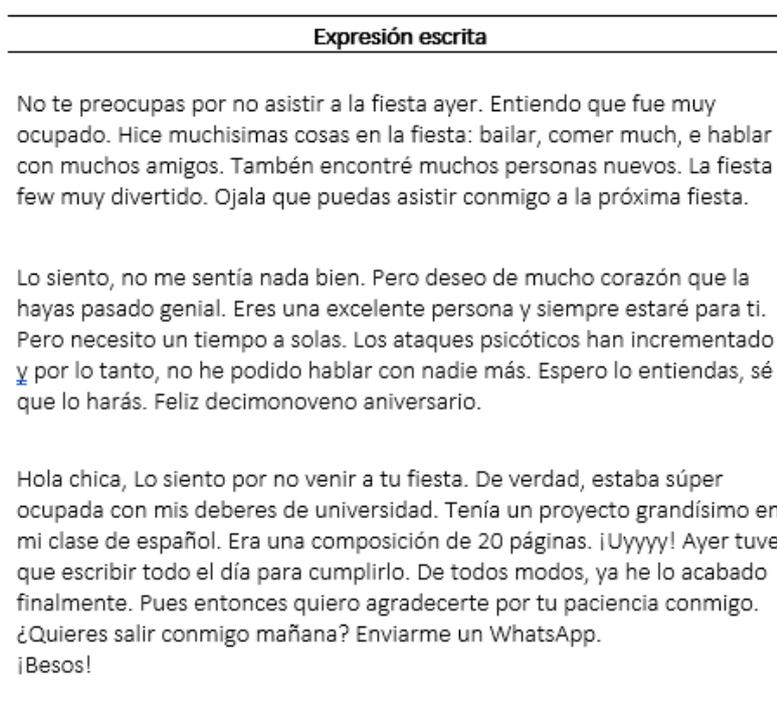


Imagen 1. Ejemplos de muestras de producción escrita

En la imagen 1 se aprecian algunas de las muestras de escritura obtenidas. Y, si bien, hay errores en la lengua, se logra observar el uso de combinaciones comunes en estos escritos, ya sean colocaciones o locuciones. Entre ellas se pueden ver combinaciones comunes como *lo siento* y *todo el día*. También se puede observar en el segundo párrafo una colocación producida con error de *mucho corazón* (en lugar de la expresión más común de todo corazón). Ya que el presente estudio se enfoca en el fenómeno colocacional, no se adentrará en el tema de las locuciones. Sólo se mencionará que las locuciones son unidades léxicas con cierto grado de fijación formal y que, además, forman parte de las colocaciones complejas identificadas por Koike (2001).

La muestra estuvo compuesta por dos grupos de sujetos participantes: el primero de 144 sujetos en redes sociales, estudiantes de ELE, y el segundo también de 144 estudiantes extranjeros que tomaron un curso de ELE en la Universidad Autónoma de

Baja California (UABC). Los sujetos se clasificaron por edad, género, país de origen y nivel de dominio de español.

Finalmente, una vez obtenidas las muestras de producción escrita, se conformó un corpus de aprendices: El CACE (Rodríguez Farah, 2022). Este se creó porque si bien existen otros corpus de aprendices con amplias muestras de lenguaje, tales como el Corpus Escrito del Español L2 (CEDEL2) (Lozano & Mendikoetxea, 2013) o el *Corpus de Aprendices de Español* (CAES) (Instituto Cervantes, n.d.), estos datan de diez años atrás. Para la compilación del CACE, cada muestra de producción escrita se procesó como un texto independiente y, a su vez, se convirtió en texto plano para ser procesado posteriormente en conjunto con el resto de los textos, a través de un gestor de corpus. Asimismo, a cada texto se le asignó una nomenclatura individual con base en los datos estadísticos proporcionados por los individuos participantes, considerando el tipo de sujeto (redes sociales o de la institución), su lengua nativa, género, nivel de dominio del español, y si cuentan con alguna certificación oficial de dominio del español. Así, una muestra de producción escrita podría contar con la siguiente nomenclatura: FB-AL-M-INT-CERT, indicando fue producida por un individuo del grupo de redes sociales (FB), cuya lengua nativa es el alemán (AL), de género masculino (M), con nivel intermedio de dominio de español (INT) y que cuenta con un certificado oficial de dominio del español (CERT). Esta nomenclatura es de utilidad para hacer comparaciones entre los fenómenos colocacionales observados y los individuos que los produjeron.

3.2 Descripción general del CACE

Puesto que esta investigación se centra en el análisis de colocaciones producidas por estudiantes de ELE, a continuación, se presenta una breve descripción del corpus.

El CACE (Rodríguez Farah, 2022) consta de 222 muestras escritas, que corresponden a 12,257 tokens y 2,113 tipos de lemas, producidas por parte de aprendices de español de 33 distintas lenguas nativas. Las muestras de producción escrita están clasificadas en cuatro niveles de dominio de la lengua: básico, intermedio, avanzado y sobresaliente. Los datos recopilados no están estructurados y, en principio, no fueron procesados; es decir, se incorporaron al corpus sin alterar la producción del participante

De acuerdo con la tipología de corpus de Sierra (2015), el CACE se puede clasificar como un corpus textual, especializado, de autoría variada, sincrónico, para propósitos específicos, monolingüe, pequeño y oportunista. Este corpus proporciona una muestra del uso actual que se le está dando a la lengua española por parte de aprendices de ELE con variados antecedentes lingüísticos y, si bien, es un corpus pequeño, permite visualizar cómo se verbaliza el fenómeno colocacional de los aprendices de ELE a través de muestras de lenguaje natural (Ghadessy et al., 2001). El CACE se puede consultar en el siguiente enlace: <https://idiomas.uabc.mx/recursos-electronicos/>.

3.3 Identificación de colocaciones en el CACE

La herramienta de *Sketch Engine* es ampliamente utilizada como apoyo en el proceso de enseñanza y aprendizaje de las lenguas (Kilgarriff et al., 2014). A continuación, se ilustra

cómo se utilizaron algunas de las funciones del software de gestión de corpus para el análisis de colocaciones.

Con el apoyo de este software de gestión de corpus, se generaron listas de todos los sustantivos, verbos, adjetivos y adverbios dentro del CACE, haciendo uso de la función *Wordlist*. Leech (2011) considera valiosa la frecuencia ordinal de las palabras dentro de un corpus. En seguimiento a la declaración del autor de que ‘a mayor frecuencia, mayor importancia de aprender’, se decidió trabajar con las unidades léxicas que tuvieran una frecuencia ≥ 5 en cada una de las listas. Así, cada una de estas unidades léxicas se analizaron en la función *Word sketch*, según la imagen 2, permitiendo observar los elementos con los que coaparece una unidad léxica dentro del CACE.

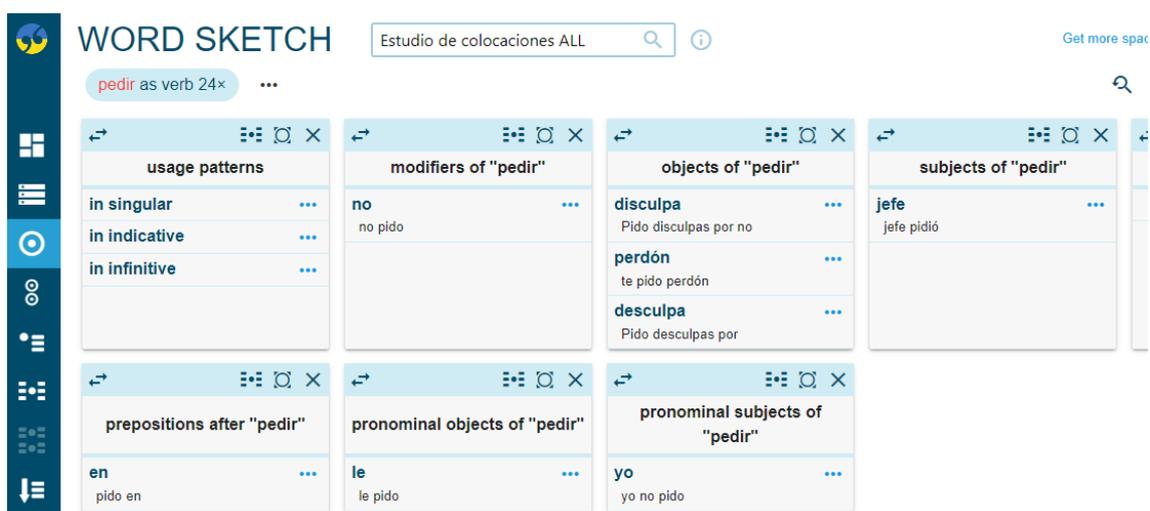


Imagen 2. Ejemplo del CACE en software de gestión de corpus

Nota: Herramienta Word sketch, que muestra las palabras que se relacionan con otra en particular

Como se puede observar en la imagen 2, la función denominada *Word sketch* (Kilgarriff et al., 2014) permite observar los objetos, los modificadores, así como los verbos con los que está colocando una unidad; esto es, los elementos colocacionales que, de manera más frecuente se relacionan con la palabra analizada dentro del corpus. A manera de ejemplo, la imagen 2 muestra el verbo *pedir*.

Firth (1962) considera cualquier par de palabras que coocuran al menos dos veces en un corpus como una colocación potencial. A partir de estas declaraciones, se decidió considerar cualquier par de palabras que coapareciera al menos dos veces dentro del CACE como una colocación potencial. La función de concordancia de *Sketch Engine* se utilizó para este fin.



Imagen 3. Ejemplo de análisis de colocaciones dentro del software de gestión de corpus

Nota: Herramienta Concordance que muestra la concordancia lingüística de la colocación pedir disculpas dentro del corpus

En la imagen 3 se puede observar la función *Concordance*. Esta función fue utilizada para observar la manera en la que una colocación en particular es producida dentro del corpus. Esta función permite observar las frases producidas por aprendices, con énfasis en la colocación buscada. En este caso, se utiliza la colocación léxica simple *pedir + disculpas* (verbo + sustantivo) como ejemplo. Tanto la función de Word sketch, para analizar colocaciones, como la de *Concordance*, para ver concordancias o ejemplos de uso de una palabra, proporcionan además metadatos sobre cada uno de los textos que conforman el corpus. Esto se ilustra en la imagen 4.

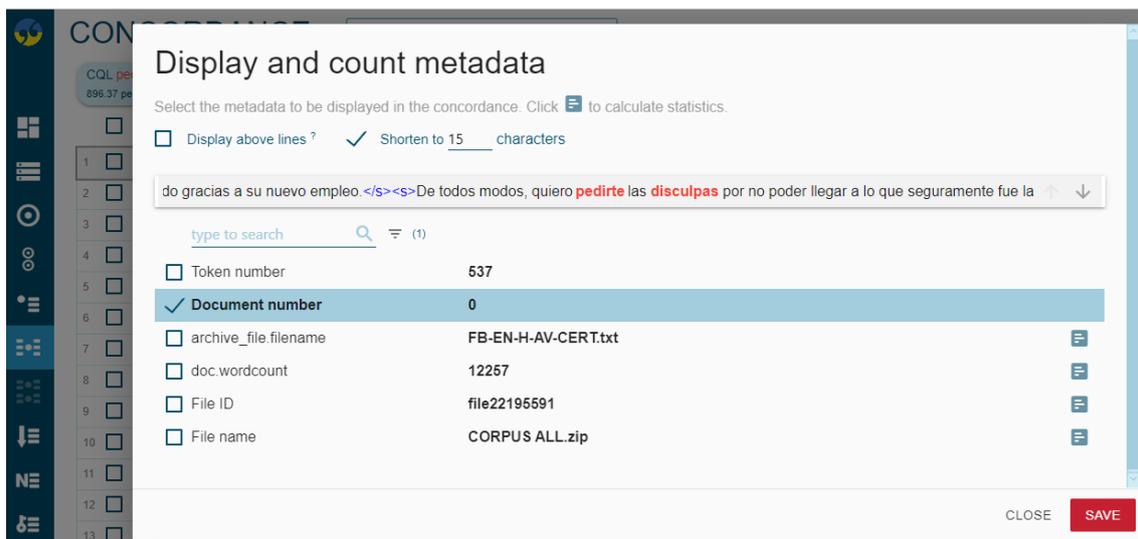


Imagen 4. Metadatos del CACE

La imagen 4 ilustra la función de Sketch Engine que proporciona información sobre cada entrada del CACE. Se puede apreciar que en esta función se visualiza la nomenclatura que se asignó a cada muestra de producción escrita del CACE, de acuerdo al perfil de los sujetos participantes.

Adicionalmente, se consultó el índice de información mutua entre las combinaciones de palabras encontradas en el corpus de aprendices. El índice de información mutua, o *MI score* en inglés, se utiliza para calcular el número de apariciones observadas de una palabra frente al número de veces que se espera que ocurra una palabra (Hunston, 2002). Hunston afirma que la información mutua mide el grado de no aleatoriedad presente en la coocurrencia de dos palabras. Hunston y Francis (2000) afirman que esto proporciona una idea más precisa de la relación entre dos palabras. Además, establecen que la información mutua evalúa la importancia de una colocación y muestra una imagen más clara de la relación entre palabras que la que proporciona una simple lista de colocación por sí sola. Este índice es una medida de atracción bidireccional. La autora afirma que un número mayor a tres debería de considerarse importante. En la imagen 5, se muestra el *MI score* de la colocación *pedir + disculpas*.

The screenshot shows the 'CONCORDANCE' interface for the query 'Estudio de colocaciones ALL'. A blue box indicates 'CQL pedir + disculpa • 13' and '896.37 per million tokens • 0.09%'. Below this, the 'Collocations' section is active, with buttons for 'CHANGE CRITERIA' and 'BACK TO CONCORDANCE'. A table lists the following data:

	Word	Cooccurrences ?	Candidates ?	MI	LogDice ↓
1	<input type="checkbox"/> disculpas	10	20	9.12	13.28 ...
2	<input type="checkbox"/> disculpa	3	8	8.71	12.19 ...

Imagen 5. Índice de información mutua

Al revisar el *MI score*, se puede notar que la palabra disculpa cuenta con *MI score* mayor a 3. Por tanto, la colocación *pedir + disculpa* se considera una colocación válida.

Alonso Ramos (2016) declara que, al analizar datos, es común emplear algún método contrastivo. Por ello, se decidió hacer uso del corpus titulado *Spanish Web 2018*. Se trata de un corpus grande de español general. Es decir, no se trata de un corpus de aprendices. Este corpus se puede consultar dentro de la plataforma de *Sketch Engine*. Dentro del *Spanish Web 2018* se consultaron todas las colocaciones encontradas en el CACE con la finalidad de verificar su validez a partir del *MI Score*, como se muestra en la imagen 6.

The screenshot shows the CONCORDANCE software interface. At the top, there is a search bar with the text 'Spanish Web 2018 (esTenTen18)'. Below the search bar, there is a status bar indicating 'CQL: pedir + disculpa • 93,958' and '4.8 per million tokens • 0.00048%'. The main area displays a list of search results with columns for source, text, and frequency. The results are as follows:

Source	Text	Frequency
deusto.es	llevar a debate las propuestas que vayan saliendo. </s><s>PD – Pido disculpas por mezclar "software privado" y "software privativo",	93,958
deusto.es	salí a recoger el premio, [...] </s><s>Antes que nada, me gustaría pedir disculpas por haber dejado el blog un poco olvidado este mes,	93,958
over-blog.es	in mínimo de decencia deberían todos en bloque en primer lugar pedir disculpas y a renglón seguido presentar todos su dimisión. </s>	93,958
uhu.es	cual estará desactivado el acceso a la Oficina Virtual. </s><s>Le pedimos disculpas por los posibles trastornos que pueda ocasionarle y le	93,958
uhu.es	eso al catálogo de procedimientos hasta nuevo aviso. </s><s>Le pedimos disculpas por los posibles trastornos que pueda ocasionarle y le	93,958
uhu.es	ón Electrónica durante un tiempo estimado de 1 hora. </s><s>Le pedimos disculpas por los posibles trastornos que pueda ocasionarle y le	93,958

Imagen 6. CACE vs. Spanish Web 2018

En esta imagen 6, se puede apreciar que la colocación de *pedir + disculpas* aparece con una frecuencia absoluta de 93,000 veces en el corpus *Spanish Web 2018*. Esta información es nuevamente útil para verificar que las colocaciones identificadas en el CACE son válidas. En el siguiente apartado se presenta un detalle de los hallazgos del CACE.

4. Análisis y discusión de datos

Una vez que se obtuvieron los resultados del cuestionario de los participantes, tanto de redes sociales como de UABC, se llevó a cabo una comparación para determinar similitudes y diferencias entre los datos arrojados por ambos grupos. Como se menciona en el apartado 3.1, cada reactivo proporciona la base de una colocación para que los participantes escojan la opción correcta de una lista de tres opciones.

Así, se concentraron los resultados de ambos grupos para lograr tener una visión más amplia del nivel de reconocimiento colocacional del total de los participantes. La tabla 3 enlista el concentrado de respuestas con mayor grado de reconocimiento correcto y las colocaciones con mayor grado de reconocimiento incorrecto por parte de ambos grupos participantes.

Concentrado de respuestas de todos los participantes				
Categoría	Tipo de colocación	% de respuest correctas	% de respuest incorrectas	% sin respue:
Léxica simple	Sustantivo (Sujeto)+ Verbo	55.9%	41.8%	2.3%
Léxica simple	Verbo+ Adverbio(mente)	39.2%	*57.45%	3.3%
Léxica simple	Verbo+ Sustantivo (Objeto)	37.0%	*60.7%	2.3%
Léxica simple	Adjetivo + Sustantivo	42.2%	*55.4%	2.4%
Léxica simple	Sustantivo+ Preposición+ Sustantivo	44.3%	*53.1%	2.6%
Léxica simple	Verbo participio+ Adverbio	52.8%	45.0%	2.3%
Léxica simple	Verbo+ preposición+ Sustantivo	75.7%	22.6%	1.7%
Léxica simple	Adverbio+ Adjetivo	58.0%	39.8%	2.3%
Léxica simple	Sustantivo+ Adjetivo	54.0%	43.8%	2.3%
Léxica simple	Verbo+ Adjetivo	41.8%	*55.6%	2.6%
Léxica compleja	Verbo+ locución nominal	59.5%	38.0%	2.4%

Léxica compleja	Locución verbal+ sustantivo	40.1%	*57.5%	2.4%
Léxica compleja	Verbo+ locución adverbial	35.6%	*60.9%	3.5%
Léxica compleja	Adjetivo+ locución adverbial	58.2%	38.5%	3.3%
Gramatical	verbo+ preposición	48.6%	*49.1%	2.3%
Gramatical	sustantivo+ preposición (+sustantivo oración subordinada)	71.7%	26.2%	2.1%
Gramatical	(artículo+) sustantivo + preposición (+infinitivo)	55.0%	42.2%	2.8%
Gramatical	preposición (+artículo)+ sustantivo (+preposición)	30.4%	*67.0%	2.6%
Gramatical	(verbo)+ adjetivo+ preposición	56.8%	40.8%	2.4%
Gramatical	(verbo + sustantivo)+ preposición+ sustantivo	59.4%	37.5%	3.1%

Tabla 3. Concentrado de resultados del cuestionario

Nota: Los datos marcados en negrillas representan el tipo de colocación que obtuvo mayoría de respuestas correctas; aquellos marcados con asterisco representan el tipo de colocación que obtuvo mayoría de respuestas incorrectas.

En cuanto a las muestras de escritura que conforman el CACE, se identificaron 764 sustantivos, 361 verbos, 175 adjetivos, y 111 adverbios. A partir del análisis de estas unidades, se identificaron 223 colocaciones producidas por los participantes. Las colocaciones se determinaron a partir de la taxonomía presentada en la tabla 1.

De las colocaciones encontradas en el CACE, se encontró evidencia de producción de colocaciones léxicas simples (Corpas Pastor, 1996; Koike, 2001) y de colocaciones gramaticales (Travalia, 2006). No se encontró evidencia de producción de colocaciones léxicas complejas (Koike, 2001). A continuación, se muestran algunos ejemplos de lo que se puede observar en el CACE.

En la categoría de léxicas simples (Corpas Pastor, 1996; Koike, 2001), se observaron cinco tipos distintos de colocaciones dentro del CACE. La colocación que se emplea con más frecuencia es la de Sustantivo + Adjetivo, mientras que Verbo + Adverbio (mente) es la de menor frecuencia. La tabla 4 muestra algunos ejemplos de las colocaciones léxicas simples encontradas en el CACE.

Tipo de colocación	Frecuencia de aparición en CACE	Ejemplo	
		Base	Colocativo
Sustantivo + Adjetivo	15	amigo	querido
Sustantivo + Preposición + Sustantivo	3	semana	fin

Tabla 4. Ejemplos de colocaciones léxicas simples en el CACE

Nota: De acuerdo a la taxonomía de Corpas Pastor (1996) y Koike (2001)

En la tabla 4, se observa un ejemplo del tipo de colocación que resultó el más frecuente dentro del CACE y que fue producida por participantes de todos los perfiles: *Sustantivo*

+ *Adjetivo*. Algunos ejemplos de la producción de esta colocación son *querido amigo* (*MI score de 9.11*), *correo electrónico* (*MI score de 11.24*), y *mejor amigo* (*IM score de 10.02*). El segundo ejemplo que se aprecia en la tabla 4 resultó el tipo de colocación menos frecuente en el CACE: *Sustantivo + Preposición + Sustantivo*. Ejemplos de este tipo de colocación son: *dolor de cabeza* (*MI score de 11.02*), *periodo de exámenes* (*MI score de 10.5*), y *fin de semana* (*MI score de 10.02*). Este tipo de colocación se produjo por participantes cuya L1 es el inglés y portugués. El nivel, género, y grupo al que pertenecen los participantes es variado.

En cuanto a las colocaciones gramaticales establecidas por Travalía (2006), se identificó el empleo de dos tipos dentro del corpus de aprendices. El tipo más utilizado es el de *verbo+ preposición*, apareciendo en 14 combinaciones. El segundo tipo identificado es el de *(verbo)+ adjetivo+ preposición*, apareciendo dos veces. La tabla 5 muestra las colocaciones gramaticales encontradas dentro del CACE y algunos ejemplos.

Tipo de colocación	Frecuencia de aparición en CACE	Ejemplo Base	Colocativo
verbo+ preposición	14	asistir	a
(verbo)+ adjetivo+ preposición	2	lleno de	estar

Fuente: Travalía (2006).

Tabla 5. Colocaciones gramaticales en el CACE

El tipo de colocación gramatical más frecuente fue *verbo+ preposición* y aparece dentro del CACE con los ejemplos *asistir a* (*MI score de 5.47*), *perdonar por* (*MI score de 6.08*), *creer que* (*MI score de 4.93*), e *invitar a* (*MI score de 5.3*). Al revisar el perfil de los participantes que produjeron estas colocaciones, se encontró que se trata de sujetos con perfiles variados en cuanto a L1, género, nivel, y grupo. El perfil de los participantes indica que esta colocación fue producida por sujetos que declaran tener un nivel de dominio de la lengua básico e intermedio. El resto de los datos estadísticos de los participantes es variado. El otro tipo de colocación gramatical identificado en el CACE es el *(verbo)+ adjetivo+ preposición*. Entre los ejemplos que se produjeron con más frecuencia se encuentran *estar lleno de* (*MI score de 5.38*) y *estar presente en* (*MI score de 9.65*).

4.1 Observaciones sobre el manejo de colocaciones léxicas simples

Las colocaciones léxicas simples (Corpas Pastor, 1996; Koike, 2001) son las que arrojaron resultados más diversos. Por ejemplo, en el caso de la colocación *Sustantivo + Adjetivo*, se encontró que hay evidencia de conocimiento sobre su uso tanto en el cuestionario como en el CACE. Adicionalmente, fue producida por sujetos participantes de todos los perfiles. Por su parte, el tipo *Verbo + Adjetivo* tuvo un bajo nivel de reconocimiento en el cuestionario y no hubo evidencia de su producción dentro del CACE. Por ejemplo, se observa que la colocación *resultar + ileso* no fue reconocida correctamente en el

cuestionario. Se teoriza que esto se debe a la ausencia de esta colocación en la L1, como en el caso de inglés, donde se traduce como *unharmmed*.

El resto de las distintas colocaciones léxicas simples obtuvieron resultados variados, mostrando evidencia de reconocimiento dentro del cuestionario, pero sin haber sido producidos dentro del CACE, o viceversa. Se teoriza que los aprendientes de ELE se beneficiarían de una instrucción enfocada hacia la existencia y el uso de las colocaciones. Como futura línea de investigación, se sugiere profundizar sobre las colocaciones léxicas simples en los aprendientes de ELE, con la finalidad de fomentar la fluidez hablada y escrita.

4.2 Observaciones sobre el manejo de colocaciones léxicas complejas

La categoría de colocaciones léxicas complejas (Koike, 2001), se caracteriza por tratarse de combinaciones que incluyen una palabra y una locución. Esta categoría obtuvo un 50% de reconocimiento de respuestas correctas dentro del cuestionario. Dentro del CACE no se encontró evidencia de producción de ninguno de los tipos de esta categoría. Si bien, se observó evidencia de producción de ciertas locuciones dentro del CACE, éstas no fueron expresadas en forma de colocación léxica compleja.

Se observa que las locuciones y las colocaciones léxicas complejas son combinaciones que parecen dificultarse para los aprendices de ELE. Algunas combinaciones, como *llorar a moco tendido*, de tipo *Verbo + locución adverbial*, contienen elementos metafóricos. Y, por tanto, su significado podría no llegar a ser transparente para los aprendientes. Se observa que este tipo de colocación requiere de instrucción explícita dentro del aula de ELE. En el aula, es necesario primero introducir el tema de las locuciones, para después presentarlas colocaciones léxicas complejas.

4.3 Observaciones sobre el manejo de colocaciones gramaticales

Las colocaciones gramaticales (Travalia, 2006), aparecen 87 veces dentro del CACE. La colocación (*verbo*)+ *adjetivo*+ *preposición* fue reconocida correctamente en el cuestionario, además hay evidencia de su producción dentro del CACE. Algunos ejemplos del CACE son (*estar*)+ *lleno*+ *de* y (*estar*)+ *presente*+ *en*. El resto de los tipos de colocaciones gramaticales obtuvieron resultados variados. Es decir, algunos tipos de colocaciones gramaticales, como la colocación *verbo*+ *preposición*, obtuvieron un bajo grado de reconocimiento dentro del cuestionario, pero fue la colocación gramatical con mayor producción dentro del CACE. Por ejemplo, en el cuestionario no se reconoció la colocación *pensar + en*, pero en el CACE se produjo *perdonar + por*. En el caso de otros tipos, como la colocación (*artículo*)+ *sustantivo + preposición (+infinitivo)*, sucedió lo contrario. Tuvieron alto grado de reconocimiento positivo dentro del cuestionario con combinaciones como (*el*)+ *derecho + a (+tener)*, pero no hubo evidencia de su producción dentro del CACE.

Se teoriza que la colocación *verbo*+ *preposición* podría llegar a ser una colocación común entre las distintas L1 de los participantes; en consecuencia, es posible que suceda un fenómeno de transferencia lingüística entre la L1 y la lengua meta. Por ejemplo, la colocación gramatical *hablar sobre (algo)*, del tipo *verbo*+ *preposición*, es una colocación

similar a decir *fale sobre (algo)* en portugués y *talk about (something)* en inglés. En cambio, no sucede lo mismo con los otros tipos de colocaciones gramaticales. Se teoriza que esto es debido a que las preposiciones se emplean de manera distinta entre las diferentes lenguas. Por ello, en el caso de las colocaciones gramaticales, los aprendices de ELE se beneficiarían de apoyo para el desarrollo de su conocimiento preposicional, así como los términos y palabras con las que colocan. Así mismo, algunos autores como Rodríguez-Fernández et al. (2019), proponen el desarrollo de herramientas para clasificar automáticamente los errores gramaticales de las colocaciones cometidos por aprendices de lenguas.

4.4. Las colocaciones en el aprendizaje de ELE

Las colocaciones no son absolutas o fijas, sino que dependen de la probabilidad de atracción entre palabras. Entonces, ¿cuál es la importancia de las colocaciones en el aprendizaje de idiomas? Por un lado, las investigaciones han demostrado que el conocimiento colocacional es un buen indicador del dominio general de una L2 (Ellis, 1997; Travalía, 2006). Por otro lado, si un aprendiente de lengua tuviera que pensar en palabras aisladas para construir una frase, sería difícil que se lograra fluidez. El aprendizaje de colocaciones integra el conocimiento de palabras de los estudiantes y agiliza su fluidez conversacional (Cheikh-Khamis Cases, 2013).

Esto respalda el hecho de que la enseñanza debería basarse en la idea de que la producción del lenguaje es la unión de unidades ya preparadas para una situación particular. La comprensión de dichas unidades, como son las colocaciones, depende del conocimiento de los patrones de predicción en diferentes situaciones. Por lo tanto, la instrucción debe centrarse en estos patrones y en las formas en que se pueden unir, junto con las formas en que varían y las situaciones en las que ocurren (Nattinger, 1980).

Pero, cada idioma maneja distintas colocaciones. Entonces surge la interrogante de cuáles son las más utilizadas y las más útiles para los aprendientes y cómo se puede descubrir el tipo de problemas que experimentan los alumnos con estas combinaciones. Para encontrar las colocaciones más comunes, se necesitan corpus actualizados del lenguaje y, para saber más sobre el uso de estas por parte de los estudiantes, se requieren corpus de la lengua de los aprendices. Estos corpus pueden proporcionar a investigadores, profesores y estudiantes una gran variedad de ejemplos de elementos lingüísticos específicos. Las colocaciones también pueden ayudar a los aprendices a construir significados y patrones de uso basados en oraciones o fragmentos de discurso recopilados de textos publicados o transcritos (Kolln, 2007).

Existe evidencia empírica que respalda el hecho de que hacer uso de colocaciones para acceder al patrón y uso de las palabras más frecuentes, en conjunto con su patrón de usos, puede mejorar la competencia lingüística de los estudiantes (Belz, 2007). El dominio colocacional puede influir en su desempeño en diferentes habilidades y componentes del lenguaje (Jaefarpour, 2006).

Otro factor importante que debe considerarse para aumentar la competencia colocacional de los alumnos es la disminución de los errores lingüísticos. Se ha sugerido que uno de los errores más comunes en el desempeño de los alumnos son los errores

de colocación (Faghih & Sharafi, 2006) y existe una alta correlación entre la competencia general de los alumnos y su dominio colocacional. Por encima de todo, en la clase de lengua, se debe aumentar la conciencia colocacional de los aprendientes y exponer repetidamente a los alumnos a las colocaciones típicas de la lengua meta. Si se enseña vocabulario, pero no se presta atención a la colocación, se está dando a los aprendices una imagen incompleta de cómo funciona el idioma.

5. Conclusiones

En este trabajo se buscó analizar la producción de colocaciones de aprendientes de ELE, a través de los datos empíricos obtenidos de un cuestionario y del corpus de aprendices denominado CACE. El CACE es un corpus de aprendices que proporciona una muestra del uso actual que se le está dando a la lengua española por parte de aprendientes de ELE. Si bien, el CACE puede ser utilizado para observar distintos fenómenos tal como cuestiones dialectales, fenómenos de interlengua, entre otros, el objetivo principal del CACE es analizar la producción colocacional de estudiantes de ELE. Por tratarse de un corpus pequeño, es relevante mencionar que los datos recopilados del CACE no representan una muestra exhaustiva del fenómeno colocacional, ni del lenguaje de los aprendientes.

De acuerdo a los criterios que Conrad (2002) declara que debe cumplir todo análisis lingüístico, se puede afirmar que en esta investigación se trabajó con patrones reales de uso en textos naturales, puesto que se observó el uso actual que se le da a las colocaciones; se trabajó con muestras de producción escrita naturales; se utilizó el software de Sketch Engine para su procesamiento y posterior análisis; se presentan ejemplos de cómo explotar el corpus para formular observaciones tanto cuantitativas como cualitativas. Además, si bien, se trata de un corpus pequeño, de acuerdo con Ghadessy et al. (2001), el análisis de corpus pequeños es una poderosa herramienta para la enseñanza y aprendizaje del lenguaje natural.

A pesar de algunas limitaciones, como el confinamiento por causa de COVID-19, se logró obtener datos de análisis con la conformación del CACE. Al analizar los datos proporcionados por los participantes, el instrumento de recopilación de datos mostró ser útil para hacer observaciones sobre la producción de colocaciones de los aprendientes. Dentro del análisis colocacional, se observó evidencia de producción de colocaciones léxicas simples y de colocaciones gramaticales. Así mismo, se constató la validez de las colocaciones encontradas en el CACE al contrastar su índice de información mutua con un corpus grande y general del español. Este estudio puede ser de utilidad para reconocer cuáles combinaciones podrían llegar a generar mayor o menor dificultad para los estudiantes de la lengua. Aun así, queda trabajo por hacer. Se espera que la presente investigación sirva de apoyo para profesionales en la enseñanza de lenguas, traductores, investigadores, y aquellos interesados en el fenómeno colocacional.

Bibliografía

ALONSO RAMOS, M. (2016). Spanish learner corpus research: Achievements and challenges. In M. Alonso Ramos (Ed.), *Spanish Learner Corpus Research. Current trends and future perspectives* (pp. 3–32). John Benjamins Publishing Company. <https://doi.org/10.1075/scl.78.01alo>

ÁLVAREZ, I., & GIL MARTÍNEZ, A. (2018). Los corpus de aprendices - un terreno en expansión para la enseñanza de español. As *Línguas Estrangeiras No Ensino Superior: Propostas Didáticas e Casos Em Estudo*, 1, 35–55.

BEAUGRANDE, R. (2001). *Large corpora, small corpora, and the learning of «language»* (E. Tognini-Bonelli & W. Teubert, Eds.; pp. 3–30). John Benjamins Publishing Company.

BELZ, J. A. (2007). The role of computer mediation in instruction and development of L2 pragmatic competence. *Annual Review of Applied Linguistics*, 27, 45–75. <https://doi.org/10.1017/S0267190508070037>

CHEIKH-KHAMIS CASES, F. (2013). *Lingüística cognitiva aplicada a la enseñanza del léxico en segundas lenguas. Las colocaciones en la clase de ELE*. (Vol. 53, Issue 9) [UNED]. <https://doi.org/10.1017/CBO9781107415324.004>

CONRAD, S. (2002). Corpus linguistic approaches for discourse analysis. *Annual Review of Applied Linguistics*, 22, 75–95.

CORDER, S. P. (1967). The significance of learner's errors. *IRAL*, 5, 161–170.

CORPAS PASTOR, G. (1996). *Manual De Fraseología Española*. Editorial Gredos. <https://es.scribd.com/document/406855356/Corpas-Pastor-Gloria-Manual-De-Fraseologia-Espanola-pdf>

COSERIU, E. (1981). *Principios de semántica estructural*. Editorial Gredos.

DE ALBA QUIÑONES, V. (2012). Lexías simples y colocaciones léxicas en la enseñanza de ELE. In *Lexías simples y colocaciones léxicas en la enseñanza de ELE*. Peter Lang D. <https://doi.org/10.3726/978-3-653-01932-2>

ELLIS, N. C. (1997). Vocabulary acquisition: Word structure, collocation, word-class, and meaning. In N. Schmitt & M. McCarthy (Eds.), *Vocabulary: Description, acquisition and pedagogy* (pp. 122–140). Cambridge University Press.

FAGHIH, E., & SHARAFI, M. (2006). The Impact of Collocation on Iranian EFL learners Interlanguage. *Science and Research Quarterly*, 16(58). https://www.researchgate.net/publication/285668229_The_Impact_of_Collocation_on_Iranian_EFL_learners_Interlanguage

FIRTH, J. R. (1957). Modes of Meaning. In *Papers in linguistics, 1934-1951* (pp. 190–215). Oxford University Press.

FIRTH, J. R. (1962). Studies in Linguistic Analysis. *A Synopsis of Linguistic Theory*, 1–32.

GHADESSY, M., HENRY, A., & ROSEBERRY, R. L. (2001). *Small Corpus Studies and ELT* (E. Tognini-Bonelli & W. Teubert, Eds.). John Benjamins Publishing Company.

- GRANGER, S. (2008). Learner Corpora. In A. Lüdeling & M. Kytö (Eds.), *Corpus Linguistics. An International Handbook*. (pp. 259–275). Walter de Gruyter. <https://doi.org/10.1002/9781405198431.wbeal0669>
- HAUSMANN, F. J. (1979). Un dictionnaire des collocations est-il possible? *Travaux de Littérature et de linguistique de RUniversité de Strasbourg*, 17(1), 187–195.
- Hernández Sampieri, R., Fernández Collado, C., & Baptista Lucio, M. del P. (2014). *Metodología de la investigación* (6ta ed.). McGraw Hill.
- Higueras García, M. (2006). *Las colocaciones y su enseñanza en la clase de ELE*. Arco/Libros.
- HUNSTON, S. (2002). *Corpora in Applied Linguistics*. Cambridge University Press. <https://doi.org/10.1017/CBO9781139524773>
- HUNSTON, S., & FRANCIS, G. (2000). Pattern grammar : a corpus-driven approach to the lexical grammar of English. In *Computational Linguistics*. John Benjamins. <https://doi.org/10.1075/scl.4>
- INSTITUTO CERVANTES. (n.d.). *Corpus de Aprendices de Español como Lengua Extranjera (CAES)*.
- JAEFARPOUR, A. AK. M. (2006). Data-driven learning and teaching collocation of preposition. Research on foreign language. *Journal of Faculty of Letters and Humanity*, 49(200).
- KILGARRIFF, A., BAISA, V., BUŠTA, J., JAKUBÍČEK, M., KOVÁR, V., MICHELFEIT, J., RYCHLÝ, P., & SUCHOMEL, V. (2014). The sketch engine: Ten years on. *Lexicography*, 1(1), 7–36. <https://doi.org/10.1007/s40607-014-0009-9>
- KOIKE, K. (2001). *Colocaciones léxicas en el español actual: estudio formal y léxico semántico*. Universidad de Alcalá y Universidad de Takushoku.
- KOLLN, M. (2007). *Rhetorical grammar: Grammatical choices, rhetorical effects*. Pearson.
- LEECH, G. (2011). Frequency, corpora and language learning. In F. Meunier, S. De Cock, G. Gilquin, & M. Paquot (Eds.), *A Taste for Corpora* (pp. 7–27). John Benjamins B.V. <https://doi.org/10.1017/CBO9781107415324.004>
- LOZANO, C., & MENDIKOETXEA, A. (2013). Learner corpora and Second Language Acquisition: the design and collection of CEDEL2.1. *Automatic Treatment and Analysis of Learner Corpus Data, 01584*, 1–30.
- NAEIMI, M., & FOO, T. C. V. (2015). Vocabulary Acquisition through Direct and Indirect Learning Strategies. *English Language Teaching*, 8(10). <https://doi.org/10.5539/elt.v8n10p142>
- NATTINGER, J. R. (1980). A Lexical Phrase Grammar for ESL. *TESOL Quarterly*, 14(3), 337–344.
- POTTIER, B. (1969). *Introduction à l'étude des structures grammaticales fondamentales*. Publications Linguistiques de la Faculté des Lettres et Sciences Humaines de Nancy.

- RICA PEROMINGO, J. P. (2010). Colocaciones gramaticales en la producción escrita de estudiantes universitarios españoles. *Reduca (Filología). Serie Lengua Inglesa*, 1, 1–25.
- RODRÍGUEZ FARAH, M. (2022). *Corpus de Aprendices de Colocaciones ELE (CACE)*. UABC. <https://idiomas.uabc.mx/recursos-electronicos/>
- RODRÍGUEZ-FERNÁNDEZ, S., CARLINI, R., & WANNER, L. (2019). Generation of a Spanish artificial collocation error corpus. *LREC 2018 - 11th International Conference on Language Resources and Evaluation*, 2519–2525.
- SCHMITT, N. (2000). *Vocabulary in Language Teaching*. www.cup.ac.uk
- SIERRA, G. E. (2015). Introducción a los corpus lingüísticos. *Introducción a Los Corpus Lingüísticos*.
- SINCLAIR, J. (1991). *Corpus, Concordance, Collocation*. OUP.
- SINCLAIR, J. (1996). *EAGLES. Preliminary recommendations on corpus typology*. Consiglio Nazionale delle Ricerche. Istituto di Linguistica Computazionale. <http://www.ilc.cnr.it/it/content/eagles-fase-i>
- TRAVALIA, C. (2006). Las colocaciones en la enseñanza del ELE: un acercamiento sintáctico. *Actas Del XXIV Congreso Internacional de AESLA*, 1–13.
- UNAM. (2015). *Certificado De Español Como Lengua Adicional Nivel Independiente-B1* (pp. 1–29). Universidad Nacional Autónoma de México.
- VINCZE, O., GARCÍA-SALIDO, M., OROL, A., & ALONSO RAMOS, M. (2016). A corpus study of Spanish as a foreign language learners' collocation production. In M. Alonso Ramos (Ed.), *Spanish Learner Corpus Research. Current trends and future perspectives* (pp. 299–332). John Benjamins Publishing Company. <https://doi.org/10.1075/scl.78.01alo>