

Universidad Autónoma de Baja California
Facultad de Idiomas
Doctorado en Ciencias del Lenguaje



Análisis de colocaciones producidas por aprendices de ELE

TESIS

Que para obtener el grado de
Doctor en Ciencias del Lenguaje

Presenta

María Fernanda Rodríguez Farah

Director: Dra. Eleonora Lozano Bachioqui

Mexicali, Baja California, junio de 2025.

Universidad Autónoma de Baja California

Facultad de Idiomas

Doctorado en Ciencias del Lenguaje



Análisis de colocaciones producidas por aprendices de ELE

TESIS

Que para obtener el grado de

Doctor en Ciencias del Lenguaje

Presenta

María Fernanda Rodríguez Farah

APROBADA POR

Dra. Eleonora Lozano Bachioqui
Director de tesis

Dr. Rafael Saldivar Arreola
Sinodal

Dr. Antonio Reyes Pérez
Sinodal

Dra. Carmen Mellado Blanco
Sinodal

Dr. Ignacio Rodríguez Sánchez
Sinodal

Mexicali, Baja California, junio de 2025.

Índice

Índice.....	3
Índice de Tablas	6
Índice de Figuras.....	7
Resumen.....	8
Abstract	9
Capítulo I. Planteamiento del problema.....	10
1.1. Antecedentes	10
1.1.1. La lingüística de corpus y los corpus de aprendices.....	11
1.1.1.2. El empleo de corpus en secuencias formulaicas y colocaciones en ELE	13
1.1.1.3. La lingüística de corpus, el interlenguaje y el análisis del error	16
1.1.2. Colocaciones en la educación.....	18
1.1.3. Colocaciones en los diccionarios.....	20
1.2. Planteamiento del problema	21
1.3. Objetivos	24
1.3.1 Objetivo General.....	24
1.3.2 Objetivos Específicos	24
1.4. Justificación y uso de resultados	24
Capítulo II. Fundamentación Teórica	27
2.1. Definiendo el término <i>colocación</i> y sus componentes.....	27
2.1.1. Características de las colocaciones.....	33
2.1.1.1. Formales.....	33
2.1.1.2. Semánticas	35
2.1.2. La colocación con relación a las combinaciones libres y las locuciones	36

2.1.3. Clasificación de categorías y tipos de colocaciones	38
2.1.3.4 La taxonomía de las colocaciones.....	40
2.2. Aprendizaje de vocabulario y léxico en las segundas lenguas.....	41
2.2.1. La lingüística cognitiva y la enseñanza de colocaciones.....	44
2.2.2. El enfoque léxico y la frecuencia	46
2.3. Corpus y su tipología.....	49
2.3.1. Los corpus de aprendices.....	52
2.3.2. La frecuencia en los corpus	54
Capítulo III. Metodología	57
3.1. Tipo y enfoque de la investigación	57
3.2. Técnicas de recolección y análisis de la información	58
3.2.1. Primera sección: El cuestionario	59
3.2.2. Segunda sección: producción escrita.....	62
3.3. Procedimiento.....	63
3.3.1. Pruebas piloto	63
3.3.2. Instrumento: Estudio de colocaciones ELE.....	66
3.3.2.1 Compilación, preparación y arquitectura del corpus	70
3.3.2.2 Procesamiento a través de software de gestión de corpus	73
3.4. Población y muestra	78
Capítulo IV. Análisis y discusión de resultados	84
4.1. Resultados sobre el conocimiento colocacional de los aprendices de ELE a partir del cuestionario	84
4.1.1. El conocimiento colocacional del grupo de redes	84
4.1.2. El conocimiento colocacional del grupo de UABC.....	86

4.1.3. Síntesis de los hallazgos del conocimiento colocacional a partir del cuestionario	88
4.1.3.1 Síntesis de los hallazgos del conocimiento de las colocaciones léxicas simples...	93
4.1.3.2 Síntesis de los hallazgos del conocimiento de las colocaciones léxicas complejas.	97
4.1.3.3 Síntesis de los hallazgos del conocimiento de las colocaciones gramaticales.....	100
4.1.4. Análisis estadístico y ejemplos ilustrativos	102
4.2. El Corpus de Aprendices de Colocaciones de ELE (CACE)	105
4.2.1. Resultados de la producción de colocaciones a partir del CACE.....	106
4.2.1.1. La producción de colocaciones léxicas en el CACE	107
4.2.1.2. La producción colocaciones gramaticales en el CACE	111
4.2.2. Análisis y discusión de la producción de colocaciones léxicas en el CACE	112
4.2.3. Análisis y discusión de la producción de colocaciones gramaticales en el CACE ...	114
4.3. Análisis y discusión del conocimiento colocacional vs. la producción colocacional	115
4.3.1. El conocimiento colocacional vs. producción de las colocaciones léxicas simples..	116
4.3.2. El conocimiento colocacional vs. producción de las colocaciones léxicas complejas	122
4.3.3. El conocimiento colocacional vs. producción de las colocaciones gramaticales	124
Capítulo V. Conclusiones	128
Referencias.....	135
Anexo 1. Perfil de participante	148
Anexo 2. El cuestionario.....	150
Anexo 3. Colocaciones encontradas en el CACE	158

Índice de Tablas

Tabla 1. Taxonomía de las colocaciones	40
Tabla 2. Características de la muestra.....	59
Tabla 3. Ejemplos de elementos integrados en el cuestionario	62
Tabla 4. Prueba piloto 2	65
Tabla 5. Nomenclatura del CACE	71
Tabla 6. Edad de los participantes	79
Tabla 7. Nivel de dominio de español	81
Tabla 8. L1 de los participantes	82
Tabla 9. Resultados del cuestionario de redes sociales.....	84
Tabla 10. Resultados del cuestionario de UABC.....	86
Tabla 11. Concentrado de resultados del cuestionario.....	88
Tabla 12 Resultados del análisis de chi-cuadrado sobre la producción de colocaciones por grupo de aprendices.....	102
Tabla 13. Colocaciones léxicas simples en el CACE	107
Tabla 14. Colocaciones gramaticales en el CACE	111

Índice de Figuras

Figura 1 Ejemplo de reactivo de la prueba piloto 2	67
Figura 2. Ejemplo de reactivo del instrumento final Estudio de Colocaciones ELE 2.....	67
Figura 3. Ejemplos de producción escrita.....	69
Figura 4. Ejemplo del CACE en software de gestión de corpus.....	73
Figura 5. Ejemplo de análisis de colocaciones dentro del software de gestión de corpu	74
Figura 6. Metadatos del CACE	75
Figura 7. Índice de información mutua.....	77
Figura 8. CACE vs. Spanish Web 2018.....	78

Resumen

El presente estudio analiza la producción de colocaciones por parte de estudiantes de español como lengua extranjera (ELE). Para ello, se aplicó un cuestionario, con el fin de recopilar datos sobre su conocimiento colocacional de aprendices, y se creó un corpus propio con las producciones de estos estudiantes. El análisis se enfocó en identificar fenómenos de uso y patrones recurrentes que nos den evidencia de la manera en la que se verbalizan estos fenómenos a diferentes niveles, considerando aspectos tales como la interlengua, errores léxicos y gramaticales, el uso de metáforas, entre otros. Todo ello se examinó con base en función la taxonomía de las colocaciones en español propuesta por Corpas Pastor (1996), Koike (2001) y Travalia (2006), con el fin de apoyar el proceso de enseñanza-aprendizaje de la lengua. Este estudio se centra en el aprendizaje de español como lengua extranjera (ELE). Con el fin de analizar la producción de colocaciones de los aprendices de ELE, como parte de la metodología, se diseñó un instrumento de recopilación de datos que nos permitiera visualizar el conocimiento colocacional de aprendices de ELE a través de un cuestionario de opción múltiple y una sección de producción escrita. La sección de producción escrita nos permitió compilar un corpus de aprendices propio, titulado Corpus de Aprendices de Colocaciones de ELE (CACE). El análisis se basó en las taxonomías mencionadas previamente, clasificando las colocaciones en léxicas simples, léxicas complejas y gramaticales. A partir de ello, se ofrece un análisis y discusión en torno a los resultados del cuestionario y del CACE. Entre los hallazgos principales, se identificó una limitada producción de colocaciones léxicas complejas por parte de los aprendices, destacando la necesidad de fortalecer su competencia colocacional. Así, se plantean algunas conclusiones, limitaciones del estudio, y futuras líneas de investigación.

Abstract

The present study analyzes the production of collocations by learners of Spanish as a Foreign Language (SFL). To this end, a questionnaire was administered to collect data on learners' collocational knowledge, and a learner corpus was created based on their written productions. The analysis focused on identifying usage phenomena and recurrent patterns that provide evidence of how these phenomena are verbalized at different levels, considering aspects such as interlanguage, lexical and grammatical errors, the use of metaphors, among others. All of this was examined based on the taxonomy of collocations in Spanish proposed by Corpas Pastor (1996), Koike (2001), and Travalia (2006), with the aim of supporting the language teaching and learning process.

This study centers on the learning of Spanish as a Foreign Language (SFL). In order to analyze learners' production of collocations, a data collection instrument was designed as part of the methodology, allowing us to observe the collocational knowledge of SFL learners through a multiple-choice questionnaire and a written production section. The written production section allowed us to compile a learner corpus, titled *Corpus de Aprendices de colocaciones de ELE* (CACE).

The analysis was based on the taxonomies proposed by Corpas Pastor (1996), Koike (2001), and Travalia (2006), classifying collocations into simple lexical, complex lexical, and grammatical types. Based on this, an analysis and discussion of the results from the questionnaire and the CACE are presented. Among the main findings, a limited production of complex lexical collocations by learners was identified, highlighting the need to strengthen their collocational

competence. Accordingly, some conclusions, study limitations, and future lines of research are proposed.

Capítulo I. Planteamiento del problema

Este capítulo comienza con una revisión de los antecedentes que resultan relevantes para la presente investigación y se incluyen aquellas investigaciones sobre lingüística de corpus, corpus de aprendices, el empleo de secuencias formulaicas y colocaciones en el aprendizaje de español como lengua extranjera (ELE), así como sobre colocaciones en la educación. Así mismo, en este capítulo se aborda el planteamiento del problema, los objetivos, la justificación y el uso de resultados.

1.1. Antecedentes

Se puede argumentar que el estudio de una lengua extranjera es uno de los retos cognitivos de mayor dificultad que puede emprender un adulto. Al respecto, algunos autores mencionan que el aprendizaje de vocabulario de una lengua extranjera (LE) es, probablemente, uno de los aspectos más retadores para lograr una buena competencia lingüística (Meara, 1980; Milton, 2009).

La adquisición de vocabulario es un proceso continuo y está estrechamente condicionado por la experiencia del estudiante. Entre las características que lo distinguen de otras habilidades lingüísticas destaca el hecho de que su aprendizaje es potencialmente ilimitado (Swain y Carroll, 1987). Diversos autores, como Teubert (2004), afirman que son las colocaciones, y no las palabras individuales, las que constituyen el verdadero vocabulario de una lengua. Si bien, esta

declaración podría considerarse algo extrema, se calcula que las secuencias formulaicas, como las colocaciones, representan el 20% de los textos escritos y el 30% de la lengua oral (Leech, 2011). A continuación, se presenta un análisis de algunos estudios que se han hecho en materia de colocaciones en el español como lengua extranjera (ELE). Así mismo, se incluyen investigaciones sobre la lingüística de corpus y su papel en el análisis de producción de colocaciones.

1.1.1. La lingüística de corpus y los corpus de aprendices

Un corpus lingüístico se puede definir como una colección de textos escritos u orales que permite realizar una variedad de análisis sobre la lengua en cuestión. Para efectos de la presente investigación, se trabajará con corpus de aprendices, una colección de textos, tanto escritos como orales, producidos por aprendices de una L2, que se emplean para el estudio de temas sobre la adquisición y la enseñanza de lenguas. Los corpus de aprendices regularmente se utilizan con la finalidad de analizar frecuencias, secuencias formulaicas, y errores producidos por los aprendices de lenguas.

De acuerdo con Alonso Ramos (2016), los corpus de aprendices surgen a partir de la intersección de tres disciplinas: la lingüística de corpus, la adquisición de una segunda lengua y la enseñanza-aprendizaje de lenguas extranjeras. La autora argumenta que existe una gran necesidad de generar recursos de aprendizaje basados en evidencia proporcionada por estos corpus.

Un ejemplo de ello es *Avalingua*, una herramienta desarrollada para el gallego con el propósito de apoyar el aprendizaje de lenguas. Alonso Ramos considera que podría ser beneficioso para el español pues proporciona retroalimentación al aprendiz a partir de la

detección de errores. En esta misma línea, Tetreault y Leacock (2014), destacan la necesidad de desarrollar recursos para la detección de errores de aprendices de lenguas distintas al inglés. De manera similar, una de las metas futuras de la presente investigación es hacer uso de los datos obtenidos para retroalimentar el proceso de enseñanza-aprendizaje de ELE.

Aunque las investigaciones de corpus tienen sus inicios a finales de los años ochenta, no fue hasta el año 2014 que el Instituto Cervantes publicó el corpus de aprendices de español. De acuerdo con la página de internet “*Learner corpora around the world*” de la Universidad de Louvain, los corpus de aprendices de inglés representan un 60% del total de los corpus existentes (Université catholique de Louvain, Centre for English Corpus Linguistics, n.d.). Por su parte, Alonso Ramos (2016), señala que los corpus en español representan un 20% del total. Sin embargo, si se consideran exclusivamente los corpus de aprendices, el español representa apenas un 8%, una cifra notablemente baja considerando que es una de las lenguas con mayor número de hablantes en el mundo.

Para comprender el desarrollo de los corpus de aprendices de español, es importante contextualizar el estatus del español como lengua extranjera (ELE). El español es la segunda lengua nativa con el mayor número de hablantes después del chino-mandarín. El Instituto Cervantes reporta que existen más de 21 millones de aprendices de ELE, aunque se estima que esta cifra puede ser hasta 25% mayor a lo reportado (Instituto Cervantes, n.d.).

A pesar de que sería difícil hablar de todos los corpus de español en existencia, Alonso Ramos (2016) se dio a la tarea de recopilar un compendio de aquellos corpus orales que han generado gran número de investigaciones; entre los que destacan están: el *Spanish Corpus*

Proficiency Level Training (D. Koike, 2011), un corpus multimodal diseñado para apoyar a los docentes de español a determinar los niveles de dominio de sus estudiantes, el *Spanish Learner Language Oral Corpora* (Mitchell et al., 2008), el cual recolecta datos de estudiantes de español cuya L1 es inglés, así como datos de nativo-hablantes del español, el corpus LANGSNAP de aprendices de español cuya L1 es inglés (Trac-Ventura et al., 2016), y el corpus CORELE (2018) que incorpora aprendices de español que son hablantes de nueve distintas L1.

Según la misma autora (Alonso Ramos, 2016), la categoría de corpus escritos es la más amplia. Entre ellos se encuentra el *Corpus para el Análisis de errores de aprendices de E/LE* (CORANE), el cual ha sido utilizado por estudiantes de lingüística aplicada para la enseñanza de lenguas en la Universidad de Alcalá (Cestero et al., 2002). Otro corpus, publicado por el Instituto Cervantes, es el *Corpus de Aprendices de Español como lengua extranjera* (CAES), el cual incluye ejemplos de escritos de aprendices de español con seis distintas L1 y se encuentra lematizado y anotado con datos sobre categorías gramaticales (Parodi, 2015).

Otro de los corpus de aprendices más prolíferos es el Corpus escrito del español como L2 (CEDEL2), el cual incluye escritos de aprendices de ELE de todos los niveles con inglés como L1 (CEDEL2, n.d.). Algunos otros corpus fueron creados para el estudio de construcciones específicas, como el *Anglia Polytechnic University Learner Spanish Corpus* que se ha utilizado para el estudio de la adquisición de los verbos *ser*, *estar* y *haber* (Ife y Peña Calvo, n.d.).

Los trabajos que se presentan en este apartado muestran el trabajo que se ha realizado sobre corpus de aprendices de ELE enfocados tanto en producción escrita como en el estudio de construcciones específicas. Éstos son relevantes para la presente investigación pues son

evidencia de la relevancia del tema del estudio del uso de la lengua por parte de aprendientes de ELE.

1.1.1.2. El empleo de corpus en secuencias formulaicas y colocaciones en ELE

En cuanto a investigaciones sobre colocaciones ELE en España se refiere, Martos Eliche y Contreras Izquierdo (2018) destacan las secuencias formulaicas de la competencia léxica. Estas se definen como una combinación de palabras almacenadas en la memoria, que pueden ser recuperadas como un todo al momento de uso. Según los autores, estas secuencias formulaicas (SF), también conocidas como *chunks*, constituyen aproximadamente la mitad del discurso nativo.

Martos Eliche y Contreras Izquierdo (2018) proponen el enfoque léxico para la enseñanza de secuencias formulaicas, basándose en el enfoque de Michael Lewis (1997) quien sostiene que la lengua consiste en léxico gramaticalizado, y no en gramática lexicalizada. Según los autores, los estudiantes de ELE no sólo deben conocer el significado del vocabulario, sino también el uso de las unidades léxicas y sus restricciones.

Asimismo, destacan que los corpus son recursos que apoyan el aprendizaje de aspectos pragmáticos, culturales y discursivos, ya que permiten contextualizar la L2 y exponer al alumno a muestras auténticas de la lengua por parte de nativos. No obstante, en esta investigación, no se pretende exponer a los aprendices de ELE a muestras auténticas, sino analizar las combinaciones de unidades léxicas o colocaciones que ellos producen.

La selección del léxico a analizar en el corpus debe fundamentarse en textos auténticos, tanto escritos como orales. Martos Eliche y Contreras Izquierdo (2018) señalan que la selección debe centrarse en las palabras de uso más frecuentes, según principios léxico-estadísticos.

Entre sus beneficios, Cruz Piñol (2017) destaca que los corpus son útiles para proporcionar concordancias y coapariciones, listas de frecuencias, corpus lematizados y analizados. Por su parte, Villayandre y Llanos (2017) recopilaron maneras de beneficiarse de estas herramientas, como la contextualización de palabras y expresiones, frecuencias de palabras, patrones de uso de sufijos y prefijos, y el análisis de la combinatoria de las palabras.

A pesar de estas ventajas, Martos Eliche y Contreras Izquierdo (2018) afirman que el uso de corpus en la enseñanza de ELE sigue siendo limitado. En particular, la escasa aplicación de los corpus en la enseñanza de colocaciones de ELE es una de las razones que motivaron el desarrollo de la presente investigación.

En cuanto a investigaciones de corpus de aprendices y el análisis de colocaciones en Latinoamérica, destaca el estudio de San Martín et al. (2019) de Chile. Su objetivo era determinar las colocaciones léxicas simples verbo-nominales de nivel A2 y B1 en un sub-corpus del corpus CAELE3 utilizando los procedimientos de Corpus de Aprendices de Lenguas en Formato Electrónico. En este estudio, se clasificaron las colocaciones identificadas de acuerdo con la taxonomía de colocaciones léxicas, al igual que en la presente investigación. Sin embargo, su enfoque se centra únicamente en las colocaciones de carácter verbo-nominales y emplea como criterio de frecuencia el propuesto por Pérez Serrano (2014). A diferencia de esta metodología, la presente investigación se enfoca en una taxonomía más amplia de las colocaciones y se basa

en la frecuencia según Firth (1962). Para mayores detalles sobre las taxonomías y las frecuencias consideradas en esta investigación, referirse a los apartados 2.1.3.4 y 2.2.2., respectivamente.

Por su parte, Hincapié (2018) también presentó el proceso de diseño y construcción del Corpus de Aprendices de Español como Lengua Extranjera y Segunda Lengua (caele/2) en su componente escrito. Se trata del primer corpus de aprendices de español como lengua extranjera y segunda lengua en Colombia. Tiene como objetivo ofrecer a docentes e investigadores muestras reales de lengua de aprendices extranjeros y sordos, con el fin de llevar a cabo análisis y propuestas didácticas que beneficien a estas poblaciones. Si bien, el autor se enfoca en el proceso del diseño y construcción de su corpus de aprendices, la presente investigación se ocupa del análisis de la información colocacional derivada del diseño y construcción del CACE.

En cuanto a los estudios llevados a cabo en México, destaca el de Oportus Torres y Ferreira Cabrera (2020). Dicho estudio se enfoca en la frecuencia de colocaciones gramaticales verbo + preposición en aprendices anglófonos de ELE de nivel A2 y B1 bajo el modelo de Análisis Contrastivo de la Interlengua (ACI). Para ello, los autores implementaron una tipología colocacional sustentada en un criterio de fijación intermedia de estas unidades. Los autores registraron las diferencias encontradas entre patrones sintáctico-semántico, variables de competencia, y otros factores.

Si bien existen algunos estudios en materia de corpus de aprendices y colocaciones, como los mencionados en este apartado, y que comparten el interés de la presente investigación, no se encontraron estudios en México que se enfoquen específicamente en la creación de corpus de aprendices de ELE para el análisis de producción escrita de colocaciones en general.

1.1.1.3. La lingüística de corpus, el interlenguaje y el análisis del error

Hasselgård y Johansson (2011) realizaron un trabajo de investigación sobre corpus de aprendices y el análisis de la interlengua en países escandinavos. El estudio de la interlengua de aprendices de lenguas extranjeras se ha convertido en una herramienta clave para identificar tanto las diferencias explícitas como sutiles entre su producción lingüística y la de los hablantes nativos.

Como señalan Hasselgård y Johansson (2011), cuando se encuentran problemas de variación, el corpus de los aprendices debe compararse con un corpus de lengua nativa (L1) que contemple variantes regionales, tipos de texto, y otras categorías. La comparación de datos entre un corpus de aprendices y uno de L1 permite identificar fenómenos como el sobreuso, el uso insuficiente y el mal uso de la lengua. En la presente investigación, se lleva a cabo el contraste de un corpus propio de aprendices de ELE, con uno de español como lengua nativa, con la finalidad de validar resultados.

En contraste con lo planteado por Hasselgård y Johansson (2011), distintos autores han puesto en duda la pertinencia de tomar el desempeño de hablantes nativos como el único punto de referencia en los estudios sobre adquisición de segundas lenguas (Cook, 2007; Davies, 2003). Desde la perspectiva de la multicompetencia propuesta por Cook (2007), el hablante de una L2 no debe entenderse como un “nativo incompleto”, sino como un usuario legítimo de la lengua, con un sistema lingüístico propio en constante desarrollo. Bajo esta visión, los corpus de aprendices adquieren una función central, ya que permiten observar patrones reales de uso lingüístico y analizar el avance de la competencia léxica a partir de producciones auténticas (Alonso Ramos, 2016). Jarvis (2017) subraya que este tipo de corpus ofrece un espacio propicio

para examinar la relación entre las formas de uso del vocabulario y diversos aspectos del conocimiento y la competencia lingüística. Alonso Ramos argumenta que estos recursos resultan especialmente valiosos en el ámbito de la enseñanza de ELE, pues facilitan la identificación de necesidades léxicas y contribuyen al diseño de estrategias didácticas basadas en evidencia empírica.

Entre otros estudios relevantes, se puede mencionar aquel de Rodríguez-Fernández et al. (2015), quienes destacan que las combinaciones arbitrarias de palabras recurrentes, o colocaciones, son un elemento clave en el aprendizaje de idiomas. De acuerdo con los autores, incluso los estudiantes avanzados tienen dificultades al usar las colocaciones; por lo que contar con herramientas eficientes para su aprendizaje resultaría de gran ayuda. Sin embargo, las herramientas existentes para verificar colocaciones todavía presentan deficiencias en la detección de errores, ya que se centran únicamente en los errores léxicos, dejando de lado los gramaticales.

De acuerdo con estos autores, los errores léxicos atraen más la atención, pero los gramaticales también deben ser considerados si el objetivo es desarrollar una herramienta integral para el aprendizaje de colocaciones que permita corregir ambos tipos de errores. En este sentido, proponen un enfoque para clasificar automáticamente los errores gramaticales de colocación cometidos por estudiantes de español de EE. UU. como base para el diseño de estrategias de corrección específicas para cada tipo de error.

Si bien, el objetivo de la presente investigación no es el análisis del error en las colocaciones, se espera que el corpus de aprendices sirva como herramienta útil para quienes deseen llevar a cabo este tipo de estudios en el futuro.

1.1.2. Colocaciones en la educación

Vincze et al. (2016) llevaron a cabo una investigación para analizar la producción de colocaciones de aprendices de español como lengua extranjera (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.

De acuerdo con Vincze et al. (2016), los estudios basados en corpus, en materia de colocaciones, ayudan a identificar información esencial de dos tipos. Primeramente, el uso de colocaciones de aprendices y nativo-hablantes puede ser comparado de manera cuantitativa para descubrir patrones de sobreuso y uso insuficiente. Además, el uso de colocaciones de los aprendices se puede analizar en términos de exactitud a través de un estudio detallado del error. Es dicha información sobre la producción de errores lo que se busca identificar en esta investigación presente.

En su investigación, Vincze et al. (2016) se plantearon la siguiente pregunta de investigación: ¿cuáles son las características de los errores producidos por aprendices de ELE? Para responder a esta pregunta, consideraron los siguientes criterios: (1) el elemento de la colocación afectado por el error (base o colocativo), (2) la naturaleza del error (gramatical, léxica, de registro), y (3) la posible fuente del error (interlingual o intralingual). En términos generales, la investigación de Vincze et al. se llevó a cabo haciendo anotaciones manuales en una porción del corpus de aprendices CEDEL2, el cual incluyó ensayos escritos tanto por aprendices de ELE como por nativo-hablantes de español

Por su parte, De Alba Quiñones (2012) llevó a cabo 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. A estas unidades léxicas plurimembres les dedicó una introducción teórica y una justificación de las características que las definen para poder desentrañar su naturaleza y localizarlas en el discurso. Finalmente, como el trabajo tiene una vocación didáctica, propuso diferentes actividades para corregir los errores más frecuentes detectados en el análisis del error; asimismo, se diseñaron distintos ejercicios para desarrollar la competencia colocacional del alumnado de ELE.

Por último, Sampedro (2022) realizó una investigación basada en lingüística de corpus y en corpus de aprendices con el objetivo de hacer una propuesta didáctica en el ámbito de ELE. Su estudio se enfocó específicamente en el acto de habla *aconsejar*. Su estudio tuvo como finalidad mostrar la importancia de los corpus de aprendices en el aula de ELE. De acuerdo con la autora, tanto el profesorado como los aprendices pueden beneficiarse al familiarizarse con las variaciones discursivas de la L2, las cuales pueden observarse con el apoyo de los corpus de aprendices.

Así, los trabajos presentados en esta sección establecen una base metodológica para el estudio de las colocaciones en la producción escrita de aprendices de ELE. Asimismo, muestran que son pocos los estudios que existen sobre la producción de colocaciones por parte de estudiantes de ELE, sobre todo en Latinoamérica y en México. Sin embargo, estos estudios han aprovechado significativamente las herramientas de la lingüística de corpus y los corpus de aprendices. En contraste con estos trabajos previos, la presente investigación se enfocará en la identificación de colocaciones y en su producción escrita por parte de aprendices de ELE.

1.1.3. Colocaciones en los diccionarios

Mellado Blanco (2008), en su estudio sobre fraseología en los diccionarios, incluye un bloque dedicado al tema de colocaciones y aspectos sintáctico funcionales, y otro que se enfoca en aspectos fraseográficos. En su investigación, presenta una concepción amplia de la fraseología, discutiendo tanto colocaciones y construcciones con verbos de apoyo y refranes.

Uno de los aspectos clave de su estudio es el contraste español-alemán en relación con las denominadas “colocaciones en un sentido estricto”, es decir, construcciones con verbo de apoyo, (Mellado Blanco, 2008). En este análisis, se menciona la existencia de una valencia lógico-semántica para algunos verbos colocativos. Así mismo, se señala que estos verbos de apoyo no están completamente vacíos de significado, lo que permite determinar clases de sustantivos con los que pueden combinarse.

Por su parte, Travalia (2001) propone una clasificación de las colocaciones marcadas diafásicamente como pertenecientes al registro coloquial. De acuerdo con la autora, los trabajos previos sobre colocaciones, en general, suelen centrarse en aquellas que forman parte de la lengua estándar, pasando por alto las del registro familiar. Algunos autores han destacado la dimensión del registro en las colocaciones. Koike (2001), por ejemplo, identifica colocaciones que forman parte del registro elevado y del registro coloquial o informal. Sin embargo, autores como Koike no ofrecen una delimitación precisa del concepto de colocación coloquial. A partir de ello, Travalia se dio a la tarea de delimitar con mayor precisión la noción de colocación coloquial y establecer criterios concretos para su clasificación.

Dentro del mismo tema de colocaciones, Travalia también aporta una clasificación de las colocaciones en dos grupos: léxicas y gramaticales. Haciendo una comparación interlingüística entre inglés y español, la autora encuentra que, en las colocaciones gramaticales, el verbo español admite más preposiciones que el inglés.

En el marco de la presente investigación, se analiza el uso de colocaciones con el apoyo de software de análisis lingüístico. Si bien, no se trabaja con diccionarios de colocaciones, se reconoce su valor como una herramienta útil para el análisis, clasificación y aprendizaje de colocaciones en español.

1.2. Planteamiento del problema

De acuerdo con Naeimi y Foo (2015), el aprendizaje de vocabulario es una de las competencias fundamentales para el aprendizaje de una lengua. Sin embargo, según los autores, los aprendices de lenguas extranjeras lo consideran el área de mayor dificultad en su proceso de aprendizaje. Estos autores afirman que un conocimiento sólido del vocabulario es crucial para desarrollar la competencia comunicativa.

Por su parte, Ellis (1997) sostiene que el aprendizaje y la comprensión de una lengua deben incluir un análisis gramatical de los bloques léxicos de una lengua. Según el autor, el aprendizaje de palabras individuales no es suficiente, ya que es necesario considerar las estructuras léxicas y las relaciones entre palabras, tales como las frases léxicas y las colocaciones. Así mismo, distingue entre el aprendizaje de vocabulario y del discurso. Mientras que el primero implica la secuenciación de las propiedades fonológicas de una lengua, el segundo considera la secuenciación de unidades léxicas de una lengua, como las frases y las

colocaciones (Ellis, 1997). Si bien se han realizado estudios en materia de vocabulario y, específicamente, de aprendizaje y producción de colocaciones de estudiantes de ELE, aún queda una brecha por cubrir en este campo.

Travalia (2006) señala que el aprendizaje de colocaciones ofrece diversas ventajas a los aprendices, entre ellas, el incremento de su vocabulario y una mayor fluidez en la lengua meta. Además, establece que los aprendices de lenguas necesitan saber cómo crear sintagmas en la nueva lengua, y son precisamente las colocaciones las que pueden servirles para construir dichos sintagmas. También, destaca que los sintagmas varían entre las distintas lenguas, lo que dificulta que los aprendices recurran a su lengua materna (L1) para comprender estas construcciones

Desde una perspectiva semántica, las colocaciones presentan una serie de dificultades en el proceso de enseñanza-aprendizaje de segundas lenguas. No obstante, dichas dificultades se encuentran en la combinatoria y no tanto en su significado, pues se consideran relativamente transparentes (Corpas Pastor, 1996). Higuera (2006) explica que esta transparencia de significado y el hecho de estar compuestas por palabras familiares hacen que las colocaciones pasen desapercibidas por muchos estudiantes.

En la actualidad, el dominio de una lengua extranjera se ha convertido en un requisito obligatorio en el ámbito profesional, debido al impacto de la globalización. En países de habla inglesa, se ha visto en los últimos años un cambio cultural que resalta la importancia del bilingüismo. En el caso del español en los Estados Unidos, aproximadamente una quinta parte de la población es hispanohablante, lo que lo convierte en la segunda lengua con mayor afluencia en el país (Vazquez, 2021). En el Reino Unido, el español ha sido posicionado como el idioma con mayor influencia para el país, tanto por razones económicas como no mercantiles (Tinsley y

Board, 2017). De hecho, Irwin y Szurmuk (2009) señalan que a partir de los años noventa, el español ha sido la lengua extranjera con mayor número de estudiantes inscritos en los Estados Unidos. La Asociación de Lenguas Modernas incluso hace la distinción entre español y “otras lenguas modernas” al momento de reportar estadísticas sobre el estudio de lenguas (Furman et al., 2009). Estas cifras, sumadas a los alcances que ha tenido la globalización, explican el creciente interés de los adultos por aprender español.

Las ventajas de la enseñanza y el aprendizaje de colocaciones de español como lengua extranjera han sido reconocidas por parte de docentes y diseñadores de material didáctico (Pérez Serrano, 2014). Sin embargo, existe una gran discrepancia entre la cantidad de material didáctico existente para el aprendizaje de colocaciones en ELE y los materiales existentes para otras lenguas, como el inglés. Son pocos los estudios que identifican las colocaciones más comunes y útiles en español (Pérez Serrano, 2014). En este contexto, la presente investigación tiene como objetivo analizar la producción de colocaciones por parte de aprendices de ELE.

Para ello, se utilizará un instrumento diseñado para recopilar datos sobre el conocimiento y la producción escrita de colocaciones, así como con el apoyo de un corpus de aprendices. La detección y el análisis de producción de colocaciones ofrecen información valiosa para mejorar el proceso de aprendizaje y enseñanza de ELE. Como señalan Alonso Ramos et al. (2010), la identificación y clasificación de la producción de colocaciones es crucial para apoyar de manera efectiva a los aprendices de lengua. De acuerdo con los autores, un corpus de aprendices permite identificar patrones de producción recurrentes, lo que facilita el desarrollo de propuestas para la clase de ELE.

A partir de todo lo anteriormente expuesto en este apartado, surge la pregunta central de esta investigación: ¿cómo es la producción de colocaciones de aprendices de ELE considerando aspectos de interlengua, factores gramaticales y metafóricos, así como el grado de reconocimiento y producción colocacional? En este sentido, el presente estudio, se centra en la producción de colocaciones de aprendices de ELE.

1.3. Objetivos

Para responder la pregunta de investigación, este trabajo se ha planteado los siguientes objetivos:

1.3.1 Objetivo General

Analizar la producción de colocaciones de aprendices de español como lengua extranjera (ELE), con diversas L1 y con edad mayor a 15 años.

1.3.2 Objetivos Específicos

Derivados del objetivo general, se establecen los siguientes objetivos específicos:

1. Determinar el conocimiento colocacional en aprendices de ELE.
2. Compilar un corpus de producciones escritas por aprendices de ELE.
3. Observar la producción de colocaciones en textos escritos de los aprendices.
4. Clasificar y categorizar los fenómenos observados en la producción de colocaciones.

1.4. Justificación y uso de resultados

De acuerdo con Vincze et al. (2016), el aprendizaje de colocaciones en español puede resultar especialmente complejo pues, a pesar de sus restricciones léxicas, estas pueden variar

entre las distintas variantes del español. Sin embargo, los cursos tradicionales de enseñanza de español como lengua extranjera (ELE), no suelen enfocarse en la enseñanza explícita de secuencias formulaicas, como las colocaciones. En su mayoría, la enseñanza de vocabulario se centra en el aprendizaje de palabras individuales, dejando de lado la importancia de las combinaciones léxicas.

Para que un aprendiz de ELE desarrolle una mayor fluidez, es fundamental que conozca las relaciones entre palabras y sus combinaciones más frecuentes. De este modo, al entablar una conversación en la lengua meta, podrá contar con un repertorio de frases y unidades léxicas, que le permitan transmitir su mensaje de manera clara. En esta línea, Cheikh-Khamis Cases (2013) destaca que el profesor debe contribuir al desarrollo de la competencia léxica del estudiante a través de imágenes conceptuales compartidas entre la L1 y la L2, así como facilitando la comprensión de aquellas que no lo sean. Por ello, la enseñanza de vocabulario no debería limitarse al aprendizaje de palabras aisladas, sino abordar también las unidades léxicas en su conjunto.

La falta de conocimiento sobre las asociaciones entre palabras por parte de aprendices de ELE puede impedir la transmisión de significado. En ocasiones, los errores pueden ser de índole gramatical, pero en otras, la dificultad radica en el uso inadecuado de una palabra dentro de un contexto determinado, más que un error gramatical per se.

Dado que es importante que los aprendices de ELE hagan uso correcto de las colocaciones en su discurso, un enfoque basado en su aprendizaje podría contribuir significativamente a mejorar los procesos de enseñanza de ELE. Así pues, el análisis de la producción de colocaciones por parte de los aprendices de ELE permite identificar los desafíos

específicos que enfrentan y, a partir de ello, desarrollar estrategias pedagógicas efectivas, que les permita mejorar sus habilidades comunicativas.

Por último, la presente investigación busca proporcionar un marco de referencia útil tanto para profesionales en la enseñanza de lenguas, como para traductores, investigadores, y otros interesados en el fenómeno colocacional.

Para cerrar este primer capítulo, se ofrece un resumen de los puntos clave abordados. En primer lugar, se presentaron los antecedentes relevantes para el desarrollo de este trabajo. Se destacó que, si bien existen trabajos sobre la producción escrita y el uso de colocaciones, hasta ahora no se ha desarrollado un estudio basado en corpus específicamente diseñado para su estudio en ELE. Posteriormente, se presentó el planteamiento del problema, subrayando la necesidad de estudiar las colocaciones desde la perspectiva del conocimiento y la producción de los aprendices de ELE. Finalmente, se expusieron los objetivos, la justificación y el uso de resultados, con el propósito de explicar la contribución de este estudio al campo de la enseñanza y adquisición de colocaciones en ELE.

Capítulo II. Fundamentación Teórica

En este capítulo se presenta la fundamentación teórica de la investigación. Para ello, se abordan tres temas principales. En primer lugar, se define el término *colocación*, según distintos autores, así como los componentes que conforman una colocación. En segundo lugar, se examina el aprendizaje y la adquisición de léxico en segundas lenguas, y su relación con las colocaciones. Por último, se explora el papel de la lingüística de corpus en el estudio de las colocaciones.

2.1. Definiendo el término *colocación* y sus componentes.

El término *colocación* es polisémico y ha sido abordado desde distintas perspectivas y disciplinas. A continuación, se presentan algunas de las concepciones más representativas.

Se atribuye a J. R. Firth (1957) la introducción del término *collocation* para referirse a combinaciones frecuentes de unidades léxicas dentro del discurso. Sin embargo, el concepto ha sido interpretado de diversas maneras. En algunos casos, se utiliza para describir asociaciones comunes o probables entre palabras. En otros casos, se restringe a aquellas combinaciones en las que un elemento léxico depende de otro para su aparición.

De acuerdo con Firth (1957), una parte del significado de una palabra se manifiesta en el conjunto de palabras con las que tiende a combinarse. El autor sostiene que el significado lingüístico sólo puede comprenderse a partir de su uso real, por lo que el contexto resulta esencial para determinarlo. En su modelo, distingue distintos niveles de análisis: fonético-fonológico, gramatical, léxico y semántico; y sitúa el estudio de las colocaciones en el plano léxico, al identificar los lexemas que tienden a coocurrir de manera recurrente.

Estas ideas marcaron un punto de partida importante para el estudio del léxico y sentaron las bases de la lingüística de corpus, al reconocer que el significado y el uso de las palabras se explican mejor a partir de los patrones que se repiten en el discurso real (McEnery y Hardie, 2012).

Lyons (1966) compara la propuesta de Firth con la teoría distribucional del significado, aunque señala que el propio Firth rechazó una interpretación estrictamente distribucionalista de su enfoque. Según Lyons, la potencialidad de colocación puede considerarse parte del significado de un lexema, pero no hasta el punto de sostener que las colocaciones estén determinadas exclusivamente por ese lexema, ni tampoco en el sentido contrario de definir su significado únicamente a partir del conjunto de sus colocaciones. En otras palabras, la relación entre colocación y significado debe entenderse como una interacción recíproca, no como una dependencia absoluta en ninguna de las dos direcciones.

Por su parte, Halliday (1966) propuso la necesidad de desarrollar una teoría del léxico independiente de la teoría gramatical e introdujo la noción de lexicalidad. Planteó la cuestión de si las relaciones entre palabras debían estudiarse dentro o fuera de la gramática, y argumentó que ambas dimensiones son interdependientes. Desde su perspectiva, los elementos que pertenecen a una misma familia léxica pueden tratarse como una sola unidad, y no existen estructuras exclusivamente léxicas o exclusivamente gramaticales, ya que toda forma lingüística participa simultáneamente de ambas. Más adelante, Halliday y Hasan (1976) retomaron el concepto de colocación para describir la tendencia de ciertos elementos léxicos a aparecer en contextos similares, lo que contribuye a la cohesión y coherencia del texto.

Por otro lado, de Hausmann (1979) define la colocación como una combinación restringida y orientada. En este tipo de relación, uno de los elementos, es decir, la *base*, posee independencia semántica; mientras que el otro, el *colocativo*, depende de la *base* para adquirir sentido. Por ejemplo, en la colocación *célibataire endurci*, el sustantivo *célibataire* funciona como *base* y el adjetivo *endurci* como *colocativo*. Hausmann subraya que la orientación entre ambos componentes no es recíproca: el colocativo necesita de la base, pero no al contrario. Además, destaca que el aprendizaje de los colocativos sólo puede entenderse dentro de su contexto de aparición, es decir, en la colocación misma.

Coseriu (1981) distingue las solidaridades léxicas de los virtúemas propuestos por Pottier (1969). Los virtúemas hacen referencia a asociaciones habituales entre palabras que reflejan relaciones de significado compartidas por los hablantes de una lengua. En cambio, las solidaridades léxicas describen vínculos más estables dentro del sistema lingüístico, que condicionan la selección de ciertos elementos en contextos específicos. Así, Coseriu señala que la coocurrencia frecuente de dos lexemas no basta para considerar que forman una colocación, ya que esta debe estar motivada por restricciones propias de la lengua y no solo por factores de conocimiento del mundo. Por ejemplo, la combinación *gaviota blanca* ilustra una unión libre basada en una característica perceptible del referente (Coseriu, 1981), mientras *que lluvia torrencial* representa una colocación léxica estable propia del español (Corpas Pastor, 1996).

Benson et al. (1986) definen las colocaciones como combinaciones fijas y reconocibles, aunque no idiomáticas. Consideran que las colocaciones son combinaciones recurrentes que se repiten de forma predecible en el uso de la lengua. Los autores distinguen dos grandes tipos: las *colocaciones gramaticales* y las *colocaciones léxicas*. Las primeras están formadas por una

palabra dominante, tal como sustantivo, adjetivo o adverbio, combinada con una preposición o una estructura sintáctica (por ejemplo, *depend on*, *afraid of*). Las segundas se constituyen por combinaciones de nombres, adjetivos, verbos y adverbios (como *make a decision* o *strong tea*). Esta clasificación ha fungido como influencia en los estudios posteriores sobre colocaciones y en la enseñanza del vocabulario en lenguas extranjeras (Corpas Pastor, 1996).

En esta misma línea, Cumming (1986) sostiene que las colocaciones son combinaciones composicionales, cuyo significado puede resultar transparente para el oyente, aunque su correcta producción exige un conocimiento específico por parte del hablante. De este modo, una colocación se entiende sin dificultad, pero no siempre se puede generar espontáneamente si no se domina su forma estable. El autor distingue las colocaciones de las expresiones idiomáticas, las cuales define como secuencias fijas cuyo significado no es composicional ni predecible a partir del sentido de sus componentes. Esta distinción permite delimitar con mayor precisión el carácter intermedio de las colocaciones dentro del continuo entre combinaciones libres e idiomáticas.

Por otro lado, Cruse (1986) define las restricciones de colocación como limitaciones de coocurrencia que no afectan al valor de verdad del enunciado. El autor distingue tres tipos de restricciones: sistemáticas, semi-sistemáticas e idiosincrásicas. En las sistemáticas, la restricción puede explicarse a partir del significado; por ejemplo, en inglés se dice *grill a hamburger* y *toast some bread*, pero resulta poco natural intercambiar los verbos en esas combinaciones. Las restricciones semi-sistemáticas comparten cierta motivación semántica, aunque presentan excepciones. En cambio, las idiosincrásicas son arbitrarias y no responden a una relación de significado clara.

Tanto Cruse (1986) como Cumming (1986) coinciden en que, en las colocaciones, cada elemento mantiene su sentido propio, aunque el colocativo adquiere un matiz particular en presencia de la base. En este sentido, las relaciones colocacionales no se definen tanto por representaciones de significado, sino por las asociaciones que se establecen entre unidades léxicas (Nirenburg y Nirenburg, 1988).

Por su parte, Heid y Raab (1989) proponen organizar el conocimiento colocacional dentro del léxico de un sistema de generación multilingüe. Siguiendo los planteamientos de Hausmann(1979) y Mel'čuk (1984), distinguen tres niveles de análisis: el semántico, el léxico y el sintáctico. Los autores coinciden en que la selección de una colocación está determinada por el hecho de que una base no admite todas las combinaciones posibles según las condiciones generales de compatibilidad semántica, sino solo un subconjunto limitado. Así, en francés son posibles combinaciones como *retenir son admiration*, *retenir sa haine* o *retenir sa joie*, pero no *retenir son désespoir*.

Alonso Ramos (1994) menciona que el criterio para definir un conjunto léxico es la potencialidad de aparición. La presencia de una unidad en un entorno colocacional solo puede describirse en términos de probabilidad. Desde esta perspectiva, las colocaciones se evalúan según su grado de lexicalidad: ante una determinada combinación, el hablante nativo no la acepta o rechaza de manera categórica, sino que la percibe como más o menos natural. Para la autora, los conjuntos léxicos no deben entenderse como constructos teóricos cerrados, sino como agrupaciones con distintos niveles de generalidad.

Además, Alonso Ramos (1994) señala que existen combinaciones restringidas en las que, teniendo un lexema como base, no todas las asociaciones posibles resultan aceptables para

expresar un significado concreto. Estas restricciones se evidencian con claridad al intentar traducir literalmente ciertas combinaciones a otras lenguas. Por ejemplo, mientras en español se dice *dar un paseo*, en francés no es posible *donner une promenade*; el verbo adecuado es *faire*. En las colocaciones, cada elemento contribuye con una parte del significado a la expresión conjunta.

La autora también afirma que las colocaciones presentan direccionalidad: uno de los lexemas actúa como seleccionador del otro. En las estructuras nombre-adjetivo, el nombre tiende a seleccionar un adjetivo determinado. Así, en inglés, *a pregnant student* constituye una coocurrencia libre determinada semánticamente por los significados de ambos lexemas. En cambio, *impeccable taste* representa una coocurrencia restringida, ya que el sustantivo *taste* selecciona el adjetivo *impeccable*, y otras opciones como *immaculate* resultarían poco naturales. Según Alonso Ramos (1994), en las colocaciones formadas por sustantivos o nombres, el elemento seleccionador es siempre el sustantivo, tanto en construcciones nombre-adjetivo como verbo-nombre.

En relación con la direccionalidad, Cruse (1986) también aborda la idea de selección léxica, pero desde la perspectiva del vínculo entre núcleo y modificador. Para este autor, el modificador actúa como seleccionador, y la impropiedad ocurre cuando las presuposiciones del seleccionador no son satisfechas por el elemento seleccionado.

Evert (2009) interpreta las colocaciones, en el sentido Firthiano, como declaraciones empíricas sobre la predictibilidad de la coocurrencia entre palabras. Según el autor, la fraseología concibe las colocaciones como combinaciones de unidades léxicas determinadas de manera parcialmente composicional y léxica.

Después de revisar algunas definiciones del concepto, en este trabajo se entenderá por *colocación* una combinación restringida de dos elementos léxicos: una base, que posee mayor peso semántico y se elige libremente en la producción lingüística, y un colocativo, cuya selección está condicionada por las restricciones léxicas impuestas por la base (Vincze et al., 2016). Esta definición se considera la más adecuada para este estudio, ya que recoge los elementos centrales de las propuestas anteriores y contempla tanto la dependencia léxica entre los constituyentes como el nivel de restricción que las distingue.

2.1.1. Características de las colocaciones

A continuación, se presentan algunas consideraciones en relación al concepto de colocación, con especial atención en las distintas características particulares desde un punto de vista tanto semántico como formal.

2.1.1.1. Formales

Según Corpas Pastor (1996), en relación al sistema, las colocaciones no se distinguen de asociaciones libres de palabras que provienen del uso creativo de la lengua. Las diferencias se observan únicamente desde el punto de vista de la norma, ya que las colocaciones están sujetas a una serie de restricciones y preferencias combinatorias impuestas en ellas por el uso.

Un aspecto formal altamente característico de las colocaciones es la frecuencia de coocurrencia entre dos unidades léxicas. No obstante, es importante señalar que no es una característica exclusiva de las colocaciones, ya que no todas las coocurrencias de alta frecuencia son colocaciones. Para ilustrar lo anterior, Coseriu (1981, citado en Corpas Pastor, 1996) distingue las colocaciones de las combinaciones libres en la combinación *gaviota blanca*. En este

caso, el hecho de que el adjetivo *blanca* se combine frecuentemente con el sustantivo *gaviota*, depende de las características físicas de las gaviotas y no un hecho de una lengua en particular.

Se debe mencionar que cuando se habla de la coocurrencia de dos unidades léxicas, no necesariamente se encontrarán una enseguida de la otra. Es posible que aparezcan una o dos palabras entre las unidades léxicas de la colocación. Se le llama distancia colocacional o *collocational span* a este fenómeno. Corpas Pastor (1996) establece que la teoría de colocaciones no especifica la distancia con la que dos unidades léxicas aún se pueden considerar parte de la misma colocación. Aun así, distintos autores, como Sinclair (1991) y Howarth (1996), que han estudiado este fenómeno coinciden en que el rango se extiende entre tres a cinco posiciones hacia la derecha o la izquierda de la base de la colocación. Según Koike (2001), se puede dar la situación en la que dos unidades léxicas conformen una colocación, pero estén separadas por más de cinco posiciones.

Otra característica formal de las colocaciones son las restricciones de combinación que le son impuestas por el uso. Por ejemplo, un nombre o sustantivo podrá requerir de la presencia de un verbo específico, aún si desde una perspectiva sintáctico-semántica se pudiera seleccionar algún otro (K. Koike, 2001). Para ejemplificar este fenómeno, se puede mencionar el nombre *bandera* el cual, en español, requiere de la presencia del verbo *izar* debido a las reglas impuestas por el uso. Juntas, estas unidades léxicas forma la colocación *izar la bandera*. Sin embargo, se podrían seleccionar verbos distintos, como *levantar* o *alzar*, desde un punto vista sintáctico y semántico, sin producir una frase completamente errónea. Este ejemplo ayuda a visualizar cómo el grado de fijación de las colocaciones es menor que el de las locuciones, ya que la fijación de las colocaciones proviene de la norma de uso y no de las reglas del sistema.

Por último, es importante resaltar la composicionalidad de las colocaciones, lo que les permite presentar características similares a las de las combinaciones libres, entre las cuales existe la posibilidad de la modificación (K. Koike, 2001). Koike establece que las colocaciones se caracterizan por ser composicionales y también por ser variables combinatoriamente y morfológicamente. Esto se debe a que presentan cierto grado de flexibilidad que les permite transformaciones sintácticas como la nominalización, la transformación a la voz pasiva, la sustitución de una unidad léxica, y las modificaciones adjetivales, entre otras. Cada colocación tiene su propio grado de flexibilidad, y no todas las transformaciones mencionadas aplican a todas las colocaciones.

En resumen, las características formales de las colocaciones indican que presentan una serie de restricciones y preferencias combinatorias impuestas por el uso. Es decir, la costumbre del habla impone su uso (Corpas Pastor, 1996). Así mismo, Corpas Pastor (1996) y Koike (2001) coinciden en que es posible que exista un espacio colocacional entre las unidades léxicas de la colocación. Adicionalmente, Koike señala que toda colocación tiene cierto grado de flexibilidad. Por ejemplo, en la colocación *poner (algo) en práctica*, la unidad léxica entre paréntesis podría ser sustituida por distintos nombres.

2.1.1.2. Semánticas

Las colocaciones cuentan con una serie de características semánticas relevantes, como los lazos entre dos lexemas y la relación que existe entre ellos, así como la precisión semántica que representan (K. Koike, 2001). Los lazos léxicos existentes en una colocación pueden extenderse a otros tipos de colocaciones, lo que produce una serie de cambios gramaticales que afectan el posicionamiento de cada uno de sus miembros (Navajas Algaba, 2006). Esta característica se

puede observar en algunas colocaciones formadas por un verbo y un adverbio con terminación en *-mente*. Estos mismos, a su vez, corresponden a una colocación de sustantivo-adjetivo, como es el caso de la colocación *llorar amargamente*. Esto se debe a que los adverbios con terminación *-mente* se derivan de adjetivos (K. Koike, 2001).

Otra característica de las colocaciones es la relación existente entre las dos unidades léxicas que las conforman. Siguiendo un ejemplo de Koike (2001), combinaciones como *tocar la guitarra* y *rasguear la guitarra* se consideran colocaciones mientras que *limpiar la guitarra* y *guardar la guitarra* no lo son, ya que *limpiar* y *guardar* no son verbos con una relación semántica típica con el sustantivo *guitarra*.

Así mismo, es relevante mencionar la precisión semántica entre los dos componentes de una colocación, los cuales se pueden emplear como definidores. En otras palabras, se utilizan para definir unidades léxicas dentro de los diccionarios (K. Koike, 2001). Por ejemplo, en el diccionario de la RAE, la primera definición ofrecida para el verbo *esperar* es *tener esperanza*, colocación de tipo sustantivo-verbo (Navajas Algaba, 2006)

En resumen, las características semánticas de las colocaciones indican que existe una relación entre los lexemas que la forman, como en el ejemplo de *llorar amargamente* o *tocar la guitarra* (K. Koike, 2001). Además, los componentes de las colocaciones pueden ayudar a definir un término como sería el caso del verbo *enfermarse*, el cual se podría también expresar como la colocación *estar enfermo* (Navajas Algaba, 2006).

2.1.2. La colocación con relación a las combinaciones libres y las locuciones

Como se comenta en el apartado 2.1. Definiendo el término *colocación* y sus componentes., no todas las combinaciones frecuentes de palabras se consideran una colocación. Existen algunas características que permiten diferenciar a las colocaciones de las combinaciones libres de palabras y de las locuciones.

Según Corpas Pastor (1996) la diferencia entre colocaciones y combinaciones libres de palabras no afectan el sistema de la lengua. Más bien, la diferencia radica en la norma de uso ya que las colocaciones muestran preferencias de combinación, así como restricciones impuestas por el uso, en comparación con las combinaciones libres que no demuestran preferencia.

Una de las diferencias entre colocaciones, combinaciones libres y locuciones se relaciona con el grado de restricción combinatoria impuesta por el uso. Este grado de fijación es menor en las combinaciones libres en comparación con las colocaciones. Koike (2001) señala que las combinaciones libres son más flexibles en su combinatoria, morfología, y sintaxis en contraste con las colocaciones. A su vez, las colocaciones no representan el mismo grado de fijación que las locuciones (Navajas Algaba, 2006) y además suelen ser más flexibles que las locuciones (K. Koike, 2001).

Por su parte, las colocaciones y las locuciones son un poco más complicadas de diferenciar. En este apartado se aborda el tema de las locuciones, no sólo en función de distinguir este concepto de aquel de colocación, sino también por el papel que juegan en las colocaciones léxicas complejas (conformadas por una unidad léxica y una locución), identificadas por Koike (2001), y las cuales se abordan más adelante, en el apartado 2.1.3.

Corpas Pastor (1996) propone que una locución es una unidad fraseológica que cuenta con “fijación interna, unidad de significado y fijación externa pasemática” (p.88). No se trata de enunciados completos y funcionan como elementos oracionales. Por su parte, Koike (2001) coincide en que las locuciones tienen una estructura fija y se caracterizan por ser idiomáticas. Según Koike, las locuciones no suelen permitir transformaciones sintácticas como la nominalización, la modificación, o la transformación pasiva, mientras que estas operaciones suceden con frecuencia en las colocaciones.

Existen, sin embargo, instancias en las que las locuciones son a la vez colocaciones. Koike (2001) afirma que las locuciones en ocasiones son el resultado de la metaforización de sus vínculos colocacionales. Un ejemplo es *ganar la lotería*, en donde su significado literal se considera una colocación (*Me gané la lotería este año*), pero en el sentido metafórico se considera una locución (*Me gané la lotería con mi esposa*).

2.1.3. Clasificación de categorías y tipos de colocaciones

Las colocaciones en español se pueden clasificar en dos tipos: las colocaciones léxicas (Corpas Pastor, 1996; K. Koike, 2001) y las colocaciones gramaticales (Travalia, 2006). Las colocaciones léxicas pueden ser simples o complejas. Las colocaciones léxicas simples se conforman de dos unidades léxicas simples como en *preparar el desayuno*. En cambio, las complejas consisten en una unidad léxica acompañada de una unidad fraseológica, o locución, como en *dormir como un tronco* (K. Koike, 2001).

Por su parte, Benson (1986) y Travalia (2006) identifican las colocaciones gramaticales, describiéndolas como frases que consisten en una palabra de categoría abierta acompañados por una palabra de categoría cerrada. A continuación, se presentan los diferentes tipos de

colocaciones dentro de las categorías de colocaciones léxicas simples, léxicas complejas y gramaticales.

Las colocaciones léxicas simples han sido identificadas y clasificadas tanto por Corpas Pastor (1996) como por Koike (2001). Ambos autores ofrecen su propia clasificación. Para Corpas Pastor (1996), se pueden distinguir seis tipos de colocaciones simples según su función gramatical y la relación entre colocados. La autora clasifica las colocaciones léxicas simples en seis tipos: (1) sustantivo (sujeto) + verbo: *correr un rumor*; (2) verbo + sustantivo (objeto): *desempeñar un cargo*; (3) sustantivo + adjetivo/ sustantivo: *fuelle fidedigna, paquete bomba*, (4) sustantivo + preposición + sustantivo: *una rebanada de pan*, (5) verbo + adverbio: *negar rotundamente*; y (6) adverbio + adjetivo: *profundamente dormido*.

Por su parte, Koike (2001) presenta su propia clasificación de las colocaciones léxicas simples, argumentando que los dos primeros tipos de Corpas Pastor (1996) se pueden agrupar en una sola categoría *sujeto + verbo*. Sin embargo, Koike subdivide esta categoría en tres grupos: *sustantivo (sujeto) + verbo*, *verbo + sustantivo (complemento directo)*, y *verbo + preposición + sustantivo*. Adicionalmente, Koike propone la categoría de *verbo + adjetivo* como en la colocación *salir ileso*. Así, Koike clasifica las colocaciones léxicas simples como (1) sustantivo + verbo, (1.1) sustantivo (sujeto) + verbo: *rumiar {la vaca}*, (1.2) verbo + sustantivo (CD): *cometer homicidio*, (1.3) verbo + preposición + sustantivo *poner (algo) en práctica*; (2) sustantivo + adjetivo: *lluvia torrencial*; (3) sustantivo + de + sustantivo: *rebanada de pan*; (4) verbo + adverbio: *cerrar herméticamente*; (5) adverbio + adjetivo/ participio: *visiblemente afectado*; y (6) verbo + adjetivo: *resultar ileso*.

Además, de las colocaciones léxicas simples Koike (2001) identifica las colocaciones léxicas complejas y son: (1) verbo + locución nominal: *dar un golpe de Estado*; (2)) locución verbal + sustantivo: *llevar a cabo un proyecto*; (3) sustantivo + locución adjetival: *dinero contante y sonante*; (4) verbo + locución adverbial: *llorar a moco tendido*; y (5) adjetivo + locución adverbial: *loco de remate*.

En cuanto a la categoría de colocaciones gramaticales, Travalía (2006) argumenta que la mayoría de los estudios en español sobre colocaciones se han concentrado en las colocaciones léxicas. Son pocos los estudios que existen en materia de colocaciones gramaticales.

Así, Travalía (2006) propone su propia clasificación de las colocaciones gramaticales, que son: (1) verbo + preposición (*hablar de / sobre / acerca de*); (2) sustantivo + preposición (+ sustantivo / oración subordinada) (*Por – piedad por, amor por*); (3) artículo + sustantivo + preposición + infinitivo (*De – el placer de, la responsabilidad de*); (4) preposición + artículo + sustantivo + preposición (*en comparación con*); (5) verbo + adjetivo + preposición (*estar seguro de*); y (6) verbo + sustantivo + preposición + sustantivo (*Oír / ver [algo] en la radio / la televisión*).

2.1.3.4 La taxonomía de las colocaciones.

A continuación, se presenta un resumen de los distintos tipos de colocaciones presentados en el apartado anterior.

Tabla 1.

Taxonomía de las colocaciones

Colocaciones léxicas simples	Ejemplo de colocación
Sustantivo (Sujeto) + Verbo	correr un rumor
Verbo + Sustantivo (Objeto)	desempeñar un cargo / una función

Adjetivo + Sustantivo	relación estrecha
Sustantivo + Preposición + Sustantivo	ciclo de conferencias
Verbo participio + Adverbio	negar rotundamente
Verbo + Preposición + Sustantivo	poner (algo) en práctica
Verbo + Adverbio (mente)	llorar amargamente
Adverbio + Adjetivo	visiblemente afectado
Verbo + Adjetivo	resultar ileso
Colocaciones léxicas complejas	Ejemplo de colocación
Verbo + locución nominal	dar un golpe de estado
Locución verbal + sustantivo	llevar a cabo un proyecto
Verbo + locución adverbial	estar en ayunas
Adjetivo + locución adverbial	claro como el agua
Colocaciones gramaticales	Ejemplo de colocación
verbo + preposición	pensar en
sustantivo + preposición (+ sustantivo / oración subordinada)	miedo de
(artículo +) sustantivo + preposición (+ infinitivo)	el placer de, la responsabilidad de
preposición (+ artículo) + sustantivo (+ preposición)	en comparación con
(verbo) + adjetivo + preposición	(Estar) convencido de
(verbo + sustantivo) + preposición + sustantivo	(Oír / ver) (algo) en la radio

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

La tabla 1 resume los tipos de colocaciones de la lengua española establecidos por Corpas Pastor (1996), Koike (2001), y Travalía (2006). Se incluye un ejemplo para cada una de las categorías de las colocaciones léxicas simples, léxicas complejas y gramaticales.

2.2. Aprendizaje de vocabulario y léxico en las segundas lenguas

El conocimiento del vocabulario va más allá que simplemente conocer el significado de una palabra determinada de forma aislada, también implica conocer las palabras que tienden a coexistir entre ellas (Celce-Murcia et al., 1993). En este apartado, se aborda el tema de aprendizaje de vocabulario y léxico general de una segunda lengua, para luego discutir el aprendizaje de colocaciones.

De acuerdo con Richards y Schmidt (2013), una palabra es una unidad lingüística mínima semánticamente independiente que puede aparecer tanto en el discurso oral como en el escrito. Los autores definen *vocabulario* como el conjunto de lexemas que incluyen palabras simples, compuestas y expresiones idiomáticas. En su planteamiento, el término léxico se entiende como el vocabulario de una lengua, en contraste con su gramática. No obstante, desde una perspectiva más reciente, se reconoce que el léxico y la gramática conforman un continuo funcional, lo que se denomina como *lexicogrammar*, en el que ambos aspectos interactúan en la construcción del significado (Hallida et al., 2014).

Por su parte, Read (2000) propone una definición más sencilla y establece que las palabras son los componentes básicos del lenguaje, las unidades de medida con las cuales se forman estructuras mayores, como enunciados, párrafos y textos completos. Este autor también identifica dos categorías generales para clasificar palabras: las de contenido y las funcionales. Las de contenido son aquellas que se consideran semánticamente independientes y regularmente se utilizan para evaluar exámenes de vocabulario. El autor afirma que estas palabras incluyen sustantivos, verbos, adjetivos y adverbios. En cambio, las palabras funcionales, en cualquier

lengua, juegan un papel gramatical. Estas funcionan como enlace entre enunciados y expresiones, y no tienen mayor significado cuando se utilizan por sí solas. Algunos ejemplos de estas palabras en español son los artículos y las preposiciones como *el, la, que, por, su, mi*, etc.

Además de palabras de contenido y las funcionales, también llamadas casos, varios autores (Read, 2000; Schmitt y Schmitt, 2020) han identificado grupos de casos que funcionan como una sola unidad con un significado y función específica. Estos grupos de palabras se conocen como frases léxicas o lexemas. De acuerdo con Folse (2004), el término frases fijas también se utiliza frecuentemente para describir estos grupos de palabras. Sin embargo, estos lexemas en ocasiones son considerados como una única unidad de significado, independientemente de la cantidad de palabras que contenga (Schmitt y Schmitt, 2020). Así, los lexemas difieren de las colocaciones ya que estas últimas se pueden distinguir de otros tipos de frases fijas por su significado compositivo, falta de lexicalización y cumplimiento de reglas sintácticas.

Richards (1976) propuso ocho suposiciones acerca del conocimiento de vocabulario, las cuales han sido ampliamente citadas en trabajos acerca de competencia léxica. La primera suposición de Richards es que el crecimiento de vocabulario, a diferencia de la gramática, nunca se detiene; continúa aumentando durante toda la vida. Por ello, el autor argumenta que conocer bien una palabra significa conocer cuál es la probabilidad de encontrar esa palabra en un discurso oral o escrito (frecuencia); conocer cuáles son las limitantes de uso de dicha palabra; conocer el comportamiento sintáctico asociado con ella; conocer las raíces de la palabra y sus derivados; conocer las relaciones que existen entre esta palabra y otras, tal como es el caso de las colocaciones; conocer el valor semántico de la palabra, y reconocer los distintos significados

asociados con la palabra. Se presume que lo que causa dificultades en el aprendizaje de grandes cantidades de vocabulario para una L2 es el tiempo y esfuerzo que toma desarrollar todo este conocimiento léxico y sus usos.

Las propuestas de Richards (1976) han sido un punto de referencia en el estudio del conocimiento del vocabulario, al distinguir entre el conocimiento receptivo y productivo y al describir distintos aspectos que intervienen en el dominio léxico. Sin embargo, su planteamiento se centra principalmente en la descripción de estos componentes, sin profundizar en los procesos de adquisición ni en las implicaciones pedagógicas que derivan de ellos.

Estudios posteriores, como el de Nation (2001), amplían esta perspectiva al integrar dimensiones relacionadas con la forma, el significado y el uso de las palabras, así como factores de frecuencia, cobertura y estrategias de aprendizaje. Este enfoque ofrece una visión más completa del aprendizaje del vocabulario en una segunda lengua y permite vincular el conocimiento léxico con su aplicación en contextos de enseñanza y uso reales.

2.2.1. La lingüística cognitiva y la enseñanza de colocaciones en las segundas lenguas

La lingüística cognitiva desempeña un papel importante en la enseñanza de unidades léxicas y las colocaciones en el ámbito de ELE (Cheikh-Khamis Cases, 2013b). Cheikh-Khamis Cases señala la escasa atención que han recibido las colocaciones en manuales de enseñanza de español como lengua extranjera (ELE), a pesar de su relevancia para lograr una competencia léxica completa. La autora argumenta que al enseñar una nueva unidad léxica, el estudiante también debe aprender las relaciones que dicha unidad tiene con otras palabras y cuáles son aquellas que comúnmente la acompañan. Además, Cheikh-Khamis Cases destaca la importancia de las colocaciones, ya que están determinadas por el uso propio de la lengua y sus

idiosincrasias; por lo que un no-nativo hablante podría enfrentar dificultades para comprender estas construcciones si no coinciden con aquellas que existen en su primera lengua. Por ello, se hace necesario una instrucción específica para que los aprendices comiencen a comprender cómo se relacionan las unidades léxicas entre sí y así desarrollar su competencia metafórica.

Cheikh-Khamis Cases (2013) también hace referencia a la teoría de marcos y prototipos. En este sentido, subraya que el pensamiento y el lenguaje están íntimamente relacionados, lo que implica que la comprensión semántica no se limita exclusivamente a la lingüística, sino que también se relaciona con los conceptos que el aprendiz posee. En esta perspectiva, los marcos se refieren al significado contextual del lenguaje, que va más allá de lo literal. Por lo tanto, un elemento lingüístico puede adquirir diferentes significados según el marco de sus distintos contextos. Basándose en este principio, se puede afirmar que ningún término es autónomo; más bien, se ajusta al conocimiento que el aprendiz tiene del mundo, por lo que es necesario aprender las relaciones entre palabras. La misma autora sostiene que la lingüística cognitiva facilita el aprendizaje de lenguas y permite al estudiante relacionar nuevas formas y nuevos conceptos de tal manera que el uso de la lengua meta se acerque al uso empleado por un nativo hablante. Por ejemplo, puede facilitar el traslado de conceptos metafóricos de la L1 a las L2.

En el tema de enseñanza y aprendizaje de colocaciones, Cheikh-Khamis Cases (2013) explora el concepto de estrategias cognitivas de interferencia, las cuales se presentan en tres grupos: intralinguales, relacionadas con el conocimiento de la lengua materna; interlinguales, que implican una comparación entre la L1 y la L2, y extralinguales, que se relacionan con la percepción del mundo por parte del aprendiz. La autora enfatiza la importancia de inculcar en los estudiantes competencias para saber experimentar con el uso de unidades léxicas, observarlas y

formular hipótesis acerca de la lengua. Además, sostiene que el aprendizaje de colocaciones facilita la producción del lenguaje real, a través de un enfoque que se basa en el uso integrando el conocimiento de palabras de los estudiantes y mejorando su fluidez conversacional. Así mismo, puede contribuir a evitar la transferencia de la lengua materna al aprender cuáles son las combinaciones posibles o imposibles entre palabras de la lengua meta (Cheikh-Khamis Cases, 2013b).

En cuanto al componente metafórico de algunas colocaciones, la autora argumenta que el profesor debe ayudar al estudiante a desarrollar su competencia metafórica mediante la utilización de imágenes conceptuales compartidas entre la L1 y la L2, así como facilitar las que no lo sean. Por lo tanto, el docente debe capacitar al alumno para reconocerlas y aprender a través de unidades léxicas, no sólo con palabras individuales.

De acuerdo con Rica Peromingo (2010), se ha demostrado que, en la adquisición de primeras y segundas lenguas, las estructuras prefabricadas son esenciales para la fluidez de la producción escrita. En esta categoría entran las colocaciones, el conocimiento de las mismas facilita la fluidez en la producción escrita de aprendices, ya que las colocaciones permiten pensar en términos de unidades fraseológicas y no sólo en términos de palabras individuales.

Por su parte, Deveci (2004) afirma que el conocimiento de colocaciones es crucial para los aprendices de una L2, pues representan patrones de uso común en la lengua meta. El autor argumenta que los hablantes nativos de una lengua hacen predicciones inconscientes sobre lo que se va a decir basándose en el contexto de una conversación. El hablante no nativo puede hacer uso de patrones del habla para facilitar su comunicación y compensar otros problemas de comunicación, como podría ser el caso de la pronunciación. De igual manera, Conzett (2000)

argumenta que una instrucción específica en vocabulario es de suma importancia para los aprendices de una L2, y el conocimiento de colocaciones proporcionará al aprendiz las herramientas necesarias para una comunicación más efectiva en la lengua meta.

Adicionalmente, Ibarretxe Antuñano (2010) señala que la lingüística cognitiva trata de integrar diferentes áreas de conocimiento lingüístico, como las teorías de la Gramática de Construcciones, de la Metáfora y Metonimia, la Gramática Cognitiva, entre otras. Así mismo, afirma que las principales hipótesis de la lingüística cognitiva son que el lenguaje no es una facultad cognitiva autónoma, que la gramática debe ser conceptualizada, y que el conocimiento de una lengua sucede a través de su uso. La conceptualización de la gramática implica que, al aprender una nueva palabra, también se aprende el contexto en el que se utiliza, las palabras que suele acompañarla, así como las construcciones lingüísticas en las que se emplea.

2.2.2. El enfoque léxico y la frecuencia

De acuerdo con Leech (2011), la manera más valiosa de sintetizar la relación entre el léxico y la gramática dentro de un mismo marco léxico-gramatical es mediante el uso de técnicas de la lingüística de corpus. Según el autor, uno de los mayores beneficios que ha aportado la lingüística de corpus es información sobre la frecuencia. La frecuencia es un tema recurrente en el aprendizaje de lenguas, por lo que Leech afirma que la relación entre frecuencia, corpus y el aprendizaje de lenguas merece mayor atención.

Leech (2011) identifica tres categorías de frecuencia. La primera, la frecuencia normal, es simplemente la cantidad de veces que ocurre un fenómeno lingüístico dentro de un corpus, texto, o colección de textos. La segunda es la frecuencia normalizada, también conocida como

frecuencia relativa, que se expresa en relación con una unidad de medida, como la cantidad de veces que una unidad léxica aparece por cada millón de palabras. Por último, Leech menciona la frecuencia ordinal, que indica la frecuencia de un fenómeno lingüístico en comparación con otros. Es decir, se trata de una lista de rangos de frecuencia, en la que las palabras están ordenadas de más a menos frecuente.

De acuerdo con el autor, la lengua se comporta de tal forma que las 50 palabras más frecuentes representan el 40% de instancias en un corpus escrito; las 3,000 palabras más frecuentes representan el 85% de instancias, y las 10,000 palabras más frecuentes representan el 92% de las instancias. Es esta frecuencia ordinal la que Leech considera más valiosa en materia de aprendizaje de lenguas, ya que, según el autor, la regla de ‘a mayor frecuencia, mayor importancia de aprender’ parece ser acertada.

En cuanto al enfoque léxico, existen distintas premisas que se han llegado a consolidar. Entre algunas de estas, se puede mencionar la importancia del aprendizaje de léxico al adquirir una L2; el hecho de que un ítem léxico es más que su forma y significado, ya que cuenta con profundidad léxica; el léxico de una lengua consta tanto de palabras individuales como bloques léxicos (incluidas las colocaciones); la importancia de priorizar el aprendizaje del léxico más frecuente; este aprendizaje puede llegar a ser más eficaz si se realiza una comparación entre la L1 y L2, y que el aprendizaje de léxico se lleva a cabo de manera gradual (Jiménez Calderón y Rufat, 2019)(Jiménez Calderón y Rufat, 2019). Jiménez Calderón y Rufat proponen que, aunque las premisas mencionadas anteriormente son preceptos generalmente conocidos, se deben aplicar estos principios a la realidad lingüística del aprendizaje de ELE.

En relación a lo anterior, Jiménez Calderón y Rufat (2019) declaran que las restricciones que imponen los verbos pueden ser de dos tipos: extralingüísticas e intralingüísticas. Las extralingüísticas son aquellas con las que se puede construir su forma de acuerdo a nuestros conocimientos del mundo. Por otro lado, si las restricciones no están basadas en nociones físicas, entonces habrá que recurrir al conocimiento de la lengua para identificar las piezas léxicas que acompañan al verbo.

Basándose en el modelo de Lakoff y Johnson (1999), Jiménez Calderón y Rufat (2019) señalan que la semántica cognitiva proporciona explicaciones para la variedad de significados. Como ejemplo, proponen el verbo *dar*. En este verbo, concurren dos tipos de modelos cognitivos. Por un lado, la extensión del verbo se puede construir a través de una estructura preconceptual. Por ejemplo, *dar algo* (como un regalo u otro sustantivo) a alguien, que es algo que está en nuestra experiencia cotidiana. Por otro lado, el segundo modelo es más complejo pues implica proyecciones metafóricas, como la construcción de *dar pie a algo*. Este modelo facilita la comprensión de aspectos complejos de la realidad, en relación a otros más familiares y básicos. Otro verbo propuesto por los autores es *tomar* y sus distintas colocaciones (*tomar una decisión, tomar nota, tomar un café, tomar ventaja, etc.*). Con estos ejemplos se muestra la dificultad que enfrentan los aprendices al intentar conocer todas las combinaciones posibles de un verbo, especialmente si se toma en cuenta que estas combinaciones son distintas según las lenguas.

2.3. Corpus y su tipología

Según Sinclair (2004), uno de los principios fundamentos de la lingüística de corpus es de identificar aquello que es central y típico en la lengua. De acuerdo con McEnery y Hardie (2012), la lingüística de corpus es un área que se enfoca en una serie de procedimientos, o métodos, para el estudio de la lengua. Desde sus inicios, el campo de investigación de la lingüística de corpus fue tanto estudiar unidades léxicas basadas en su significado como en la coaparición con otras palabras (Rica Peromingo, 2010). De acuerdo con Rica Peromingo, 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. Y, como sugiere Conrad (2002), la lingüística de corpus ha sido de gran utilidad para determinar patrones de comportamiento de la lengua, identificar usos recurrentes o inusuales de la lengua por parte de los usuarios, y describir la interacción entre varios fenómenos de producción de la lengua.

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 y Schmitt, 2020). De acuerdo con Conrad (2002), un aspecto fundamental de todo análisis lingüístico basado en corpus debe cumplir los con algunos requisitos, que incluyen: (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 tal análisis; y (4) que dependa de técnicas analíticas tanto cuantitativas como cualitativas.

La lingüística de corpus ha tenido una influencia significativa en el campo de la lexicología, así como en la creación de diccionarios (Rica Peromingo, 2010). Entre sus

aportaciones más significativas en las que los corpus lingüísticos en diccionarios con un enfoque especial en colocaciones, frecuencia, y fraseología (Hunston, 2002).

Adicionalmente, Francis (1967) define un corpus lingüístico como una colección de textos que se asume son representativos de cierta lengua, dialecto, o cualquier subconjunto la lengua, la cual se empleará para el análisis lingüístico. Según 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. Por otra parte, el grupo de trabajo dedicado a los corpórea textuales del *Expert Advisory Group on Language Engineering Standards* (EAGLES) (Sinclair, 1996) 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. En este mismo sentido, Alonso Ramos (2016) afirma que un corpus lingüístico es una colección de textos orales, escritos, o archivos multimodales con los cuales se puede extraer información, tanto de distribución como estadística, de ciertos fenómenos lingüísticos. De acuerdo con Viera (2019), dichos corpus cuentan con ciertas características: son representativos, son auténticos, se codifican de manera estandarizada, se pueden procesar con el apoyo de herramientas diseñadas con ese fin, y tienden a contar con anotaciones.

Por su parte, Sierra (2015) menciona que los corpus constituyen un modelo de la realidad lingüística de algún fenómeno que se desea observar. El autor establece que, tanto para análisis cuantitativos como cualitativos, es necesario trabajar con muestras reales de la lengua. Al respecto establece que el 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. El mismo autor propone una tipología para clasificar los distintos tipos de corpus de acuerdo con sus características. El autor los clasifica de

la siguiente manera: (1) según el origen de los datos, es decir, si se trata de corpus textuales u orales; (2) según la espontaneidad del habla, en el caso de corpus orales; (3) según la codificación y anotación; (4) según la especificidad de los elementos, donde un corpus especializado reúne datos para la descripción de un tema en particular; (5) según la autoría de los textos, los cuales pueden ser clasificados como de autoría variada; (6) según la temporalidad de los textos, los cuales pueden ser diacrónicos o sincrónicos; (7) según el propósito de estudio, es decir, si son multipropósito o con propósitos específicos; (8) según la lengua, se clasifican en monolingües o multilingües; (9) según la cantidad de texto, ya sean corpus grandes o pequeños; (10) según la distribución del tipo de texto; (11) según la accesibilidad, ya sea que se trate de un corpus de dominio público o de uso privado; (12) según la documentación; y (13) según la representatividad, los cuales se clasifican en corpus representativos o corpus oportunistas (aquellos que no representan toda la lengua pero sí algún fenómeno en específico).

Además de la propuesta de Sierra, otros autores proponen otros tipos de corpus. Por ejemplo, Ghadessy (2001), haciendo referencia a corpus pequeños, declara que el interés por los estudios analíticos de dichos corpus pequeños ha ido en aumento debido a los descubrimientos lingüísticos que estos han aportado, los cuales se han llegado a considerar tan importantes como aquellos derivados de estudios de otros corpus de gran tamaño. De acuerdo con el autor, el análisis de corpus pequeños es una poderosa herramienta para la enseñanza y aprendizaje del lenguaje natural.

Por su parte, Beaugrande (2001) distingue dos tipos de corpus pequeños que son de interés en área de aprendizaje de lenguas: los corpus de aprendices y los corpus especializados. Estos últimos se determinan por su registro, el dominio discursivo, o un área de estudio en particular. De acuerdo con Beaugrande, los corpus pequeños son de especial relevancia, y sobre

todo en materia de aprendizaje de lenguas, para proyectos que buscan analizar fenómenos específicos de la lengua. En relación con los tipos de corpus, McEnery y Hardie (2012) agregan que algunos corpus se pueden clasificar como oportunistas cuando representan nada más y nada menos que los datos que fueron posibles recolectar para una tarea específica. Los autores afirman que pueden llegar a surgir restricciones que previenen la recolección en gran volumen. Así, para efectos de este trabajo, se ha recopilado un corpus pequeño y oportunista enfocado hacia la observación del comportamiento del fenómeno colocacional por parte de aprendices de ELE.

2.3.1. Los corpus de aprendices

De acuerdo con Granger (2008), las investigaciones sobre corpus de aprendices son relativamente nuevas. Sin embargo, la autora señala que el tema de corpus de aprendices es una rama dinámica de la lingüística de corpus y surgió como una disciplina propia entre las décadas de 1980 y 1990. La autora concuerda con Nesselhauf (2005) al definir un corpus de aprendices como una colección electrónica de textos producidos por aprendices de lenguas. Según la autora, los corpus de aprendices cuentan con todas las características que regularmente se atribuyen a los corpus lingüísticos. La única diferencia es que los datos provienen de aprendices de lenguas.

Los corpus de aprendices han sido utilizados generalmente para dos funciones distintas, pero relacionadas entre sí (Granger, 2008). Por una parte, pueden contribuir a enriquecer teorías sobre la adquisición de segundas lenguas al proporcionar una mejor descripción del fenómeno de interlenguaje y una mejor comprensión de los factores que lo influyen. Por otra, pueden ser utilizados para desarrollar herramientas y métodos pedagógicos que aborden de manera más acertada las necesidades de los aprendices de lengua.

En cuanto al tamaño de los corpus de aprendices, Granger (2008) declara que éstos pueden llegar a contener millones de palabras, ya que la mayoría se almacenan electrónicamente y los datos se pueden recolectar de manera relativamente rápida. Sin embargo, de acuerdo con Ragan (2001), los corpus de aprendices pequeños también pueden llegar a ser de valor. De acuerdo con el autor, el tamaño de la muestra es menos importante que la preparación y la adaptación del producto lingüístico y su aplicación subsecuente en el corpus con la finalidad de poner un foco en un fenómeno particular de la lengua.

De acuerdo con Sinclair (1996), uno de los desafíos en la creación de un corpus de aprendices es la autenticidad de los textos. El autor afirma que los datos producidos (o muestras de lenguaje) por los aprendices pueden mostrar una variedad de grados de naturalidad, dependiendo del método utilizado para recolectar los datos. Actividades como leer en voz alta o ejercicios de llenar el blanco suelen producir muestras de baja naturalidad de la lengua, mientras que otras actividades como entrevistas y composiciones tienden a ser más naturales. Sinclair menciona que la noción de un ‘texto continuo’ es fundamental en la creación de cualquier corpus, y que una serie de palabras o de frases descontextualizadas producidas por aprendices de lenguas pueden considerarse datos lingüísticos válidos, aunque no así para un corpus de aprendices. Adicionalmente, Sinclair menciona que, idealmente, un corpus de aprendices debe conformarse con datos que provengan de actividades en las que permitan a los aprendices seleccionar sus propias pautas, y no de aquellas actividades que sean restrictivas.

Por su parte, Álvarez y Gil Martínez (2018) declaran que los corpus de aprendices permiten acceder a la interlengua de quienes están en el proceso de aprendizaje del español y por ello, son una herramienta de gran interés para el área de enseñanza-aprendizaje de lenguas, ya que estos corpus permiten hacer contrastes sobre las dificultades de aprendizaje y el vocabulario

empleado en la producción. Además, Núñez (2019) destaca que el estudio de los errores presentes en las producciones de los estudiantes constituye una de las principales áreas abordadas por la investigación en corpus de aprendices en materia de ELE.

De acuerdo con Calero et al. (2020), los corpus de aprendices de ELE deben tomar en cuenta diversos metadatos, incluidos los de tipo sociolingüístico, los cuales deben ser codificados para permitir filtrar la información en búsquedas automáticas. En esta misma línea, investigaciones recientes en lingüística de corpus han destacado la importancia de contar con esquemas de metadatos estandarizados que favorezcan la interoperabilidad y la comparabilidad entre distintos corpus de aprendices (Paquot et al., 2024). Estos metadatos suelen incluir variables adscritas y sociolingüísticas, tales como género, edad y lengua materna. Asimismo, deben incluir variables relacionadas con la experiencia de aprendizaje, tales como el nivel de dominio de ELE, los años de estudio de la lengua, las estancias en países hispanohablantes y el tipo de instrucción recibida. Un ejemplo destacado de buenas prácticas en este sentido es el corpus CEDEL2 (Lozano y Mendikoetxea, 2013), que incorpora una codificación detallada de metadatos sociolingüísticos y pedagógicos, lo que permite realizar búsquedas automáticas complejas y contrastar patrones de producción entre aprendices con diferentes perfiles lingüísticos. Este tipo de estructura y codificación de la información resulta esencial para examinar de manera más precisa los factores que inciden en la adquisición y el uso de colocaciones en ELE., el nivel de dominio de ELE, los años de estudio del español, las estancias en países hispanohablantes, entre otros factores.

2.3.2. La frecuencia en los corpus

De acuerdo con Leech (2001), previo a la existencia de corpus lingüísticos asistidos por computadora, se solían generar listas de frecuencia de palabras en inglés contando y calculando las instancias a mano. El autor declara que, para estándares modernos, dichos corpus son lamentablemente pequeños y su elección de textos incluían ejemplares de poca utilidad para aprendices de la lengua.

Por su parte, Zipf propone un estudio matemático en torno a frecuencias de palabras. La ley de Zipf de (1935) y (1949) mantiene que la frecuencia de cualquier palabra es inversamente proporcional a su rango en listas de frecuencia. Esta ley de Zipf otorgó mayor importancia a las palabras más frecuentemente utilizadas en relación a lo esperado en una distribución regular.

Algunos lingüistas como Sinclair (1970) consideran que aquellas listas de frecuencia de palabras son de poca utilidad, argumentando que considerar cada instancia de palabras independientes es inadecuado. Este autor considera de mayor relevancia observar la frecuencia de las combinaciones de palabras. Esto se basa en el principio idiomático, en el que se observa que los textos se construyen en términos de frases semi-construídas. Este principio idiomático del autor ha sido adoptado por diversos lingüistas de corpus para quienes las unidades multi-palabras, como las colocaciones y grupos léxicos, son esenciales para la construcción y el aprendizaje de las lenguas. Sinclair, a su vez, estaba siguiendo los pasos de Halliday (1966) y J.R. Firth (1957), quienes por primera vez identificaron la importancia del concepto de co-frecuencia de las colocaciones. Halliday (1966) declaró que los niveles de léxico, incluyendo las colocaciones, debían de considerarse como un nivel específico de la descripción lingüística, a la vez proponiendo que los niveles de gramática y léxico se interrelacionan. Para Halliday, los

niveles gramaticales y léxicos constituyen un solo nivel léxico-gramatical que comprende la estructura formal de la lengua.

Entre los estudios realizados acerca del análisis de las colocaciones tanto para las coocurrencias léxicas como las gramaticales, se puede mencionar el de patrones gramaticales de Hunston y Francis (2000), el cual se describe un enfoque basado en corpus de la gramática léxica del idioma inglés. Similar a la propuesta de Hunston y Francis, en esta investigación, se trabajará bajo los patrones de colocaciones en español, presentado en la sección 2.1.3.4.

Stefanowitsch y Gries (2003) trabajaron una medición estadística del grado de atracción o rechazo entre palabras y construcciones, el cual llamaron *Collostructions*. Por su parte, Cheng et al. (2006) utilizaron el programa ConcGram para generar colocaciones de palabras con variaciones en posición y distancia de tal forma que, por ejemplo, *interpretar un papel*, *interpretar un papel importante*, e *interpretar un papel clave* pueden ser todas agrupadas en la misma categoría de concordancia.

Según Leech (2011), la importancia de estas investigaciones yace en la exploración estadística de la relación entre el léxico y la gramática, la cual previamente había sido mayoritariamente ignorada. Siguiendo los pasos de dichos autores, se consideró la distancia y posición de colocaciones dentro del análisis colocacional que se llevó a cabo en la presente investigación. Para mayores detalles al respecto, se puede referir a los capítulos Capítulo III y Capítulo IV.

Capítulo III. Metodología

A lo largo del Capítulo III se presenta el diseño conceptual de la metodología empleada para cumplir con los objetivos de la presente investigación. El capítulo comienza con una descripción del tipo y enfoque de la investigación que se desarrolló. A través de los apartados del presente capítulo, se ofrece una descripción del proceso de toma de decisión para diseñar el instrumento de recopilación de datos, así como un detalle del proceso para la obtención de los datos. Se describe la estructura del instrumento y las etapas de prueba y rediseño del mismo. Todo esto con la finalidad de verificar que los datos recabados fuesen relevantes y válidos para esta investigación. Así, se ofrece una descripción del instrumento final. Se detallan los criterios de selección de datos, así como la preparación del corpus y la clasificación de los textos. Adicionalmente, se presenta el procesamiento que se le dio a los datos a través de un software de gestión de corpus. Por último, se ofrece una descripción de la población participante.

3.1. Tipo y enfoque de la investigación

La presente investigación se enmarca dentro de un diseño transversal, descriptivo y aplicado, con el propósito de explorar el conocimiento colocacional de los aprendices de ELE. Se trata de una investigación de carácter mixto, ya que se trabajó tanto con datos tanto cuantitativos como cualitativos para examinar la frecuencia de uso de colocaciones y la calidad de la producción escrita de los mismos.

Con la finalidad de poder observar el manejo del fenómeno colocacional por parte de los aprendices de ELE, se diseñó un instrumento que, en primera instancia, proporcione datos cuantitativos permitiendo así medir la frecuencia y magnitud de los fenómenos observados

(Hernández Sampieri et al., 2014). Y, en segunda instancia, que el instrumento aporte datos cualitativos. Esto con la finalidad de hacer un contraste con los datos cuantitativos obtenidos, lo que facilita la comprensión del fenómeno colocacional desde distintas ópticas. Una descripción a detalle del instrumento se ofrece en la siguiente sección. Por último, cabe mencionar que se trabajó con un corpus de corte oportunista, como se menciona en el apartado 2.3., ya que se trabajará solamente con aquellos datos que fueron posibles recopilar para efectos de la presente investigación (McEnery y Hardie, 2012).

3.2. Técnicas de recolección y análisis de la información

Dado el enfoque mixto de la presente investigación, se emplearon tanto técnicas cuantitativas como cualitativas para recopilar información sobre la producción de colocaciones en aprendices de ELE. Se buscó obtener datos cuantitativos para realizar generalizaciones sobre las colocaciones observadas, así como datos cualitativos sobre la producción de esas colocaciones para ampliar y profundizar nuestras observaciones más allá de la información cuantitativa (Hernández Sampieri et al., 2014).

De acuerdo con Ñaupas Paitán et al. (2014), algunos de los instrumentos principales para la recopilación de datos son la observación, los grupos focales, la recopilación documental y el cuestionario. En este estudio, se optó por utilizar el cuestionario como instrumento de investigación, ya que permite recopilar datos de grupos grandes de personas de una manera estandarizada. El cuestionario hace posible recopilar datos a través de una serie de preguntas focalizadas en el conocimiento del fenómeno colocacional. De esta manera, los datos obtenidos

pueden ser controlados y también pueden ser contabilizados. Así mismo, el cuestionario proporciona información sobre el uso natural que se le da a la lengua por parte de aprendices.

El cuestionario diseñado para esta investigación incluyó tanto preguntas cerradas de opción múltiple para obtener datos cuantitativos, así como una pregunta de escritura libre para recopilar datos cualitativos sobre el fenómeno a investigar. En la siguiente sección se presenta una descripción del cuestionario diseñado para la obtención de datos de esta investigación.

3.2.1. Primera sección: El cuestionario

Así mismo, como se comenta en el apartado anterior, se diseñó un cuestionario con la finalidad de medir el conocimiento colocacional de los participantes. Este instrumento se dividió en dos partes. La primera se compone de preguntas destinadas a recabar información n sobre los estudiantes, lo que permite relacionar su competencia colocacional con variables sociolingüísticas.

De acuerdo con lo señalado en el apartado 2.3.1, Calero et al. (2020) sostienen que los metadatos de un corpus de aprendices deben incluir información de tipo sociolingüístico. En concordancia con ello, estas preguntas recogen datos sobre edad, experiencia con el estudio de la lengua española, lengua nativa, y país de origen, entre otros. La tabla 2 presenta las categorías correspondientes a estas variables incluidas en el cuestionario.

Tabla 2

Características de la muestra

Preguntas de Estadística	Rangos de respuesta
País de origen	Cualquiera

Rango de Edad	15-20 / 21-25 / 26-30 / 31-35 /36-40 / 41-45 / 46-50 / 51-55 / 56-60 / 60 en adelante
Género	Mujer / Hombre / Prefiero no decir
Nivel de español	Básico / Intermedio / Avanzado / Sobresaliente
¿Dónde estudiaste español?	Escuela / Otro
Certificaciones de la lengua	No requerida
¿Has practicado español en el extranjero?	Sí / No
¿Cuántos años has estudiado español?	Cualquiera
¿Cuál es tu lengua nativa?	Cualquiera

Nota: Ver Anexo 1. para consultar la sección estadística completa del cuestionario.

Fuente: Datos tomados a partir del cuestionario.

La segunda parte del instrumento tiene la finalidad de medir el conocimiento colocacional de los participantes a través de preguntas de opción múltiple, también conocidas como politómicas. Este tipo de preguntas, al estar formuladas de manera sistemática, permite verificar los objetivos de una investigación, así como abordar las variables implicadas (Ñaupas Paitán et al., 2014).

Diversos estudios han empleado este formato para evaluar el conocimiento de colocaciones léxicas en aprendices del inglés. Investigadores como Hussein (1990), Marton (1977), Bahns (1993), y Biskup (1992), utilizaron preguntas cerradas en sus estudios de aprendices de inglés. De igual manera, Deveci (2004) destaca que este tipo de ejercicios permite medir el nivel de conocimiento de, mientras que Conzett (2000), por su parte, propone actividades de completar oraciones o seleccionar el colocativo adecuado dentro de una lista como estrategias útiles para la evaluación colocaciones

Las preguntas del cuestionario requieren que los participantes elijan entre opciones de respuesta para completar una oración con el término correcto y formar una colocación. El diseño del cuestionario se basa en la taxonomía de colocaciones presentada en la tabla 1, propuesta por las autoras Corpas Pastor (2001), Koike (2001) y Travalía (2006). Esta taxonomía proporciona una estructura para clasificar y comprender las colocaciones comunes en el idioma español.

Si bien este tipo de preguntas de opción múltiple son fáciles de codificar y permiten obtener datos cuantitativos, se reconoce que pueden limitar la información proporcionada por los participantes, ya que solo indican su nivel de reconocimiento colocacional y no ofrecen evidencia directa de su producción de colocaciones. Por lo tanto, se incluyó una pregunta abierta en el cuestionario para obtener muestras naturales del uso de colocaciones por parte de los participantes y formar un corpus de aprendices. Aunque las respuestas abiertas son más difíciles de codificar, permiten una apreciación cualitativa de los datos proporcionados. A continuación, se ofrece una descripción de la estructura del cuestionario diseñado para esta investigación

El cuestionario consta de 40 preguntas de opción múltiple en la primera sección, donde cada pregunta presenta una oración con un espacio en blanco que los participantes deben completar seleccionando una de tres opciones proporcionadas. Como se menciona anteriormente, estas preguntas de opción múltiple se basan en las taxonomías de colocaciones propuestas por las autoras Corpas Pastor (2001), Koike (2001) y Travalía (2006). A continuación, se presenta una tabla con ejemplos de las preguntas del cuestionario. Para ver el cuestionario completo, consulte el anexo 1.

Este tipo de ejercicios de opción múltiple también permite explorar la profundidad del conocimiento léxico de los aprendices, entendida como la habilidad para reconocer y relacionar

unidades léxicas en combinaciones naturales de la lengua. De acuerdo con Qian y Schedl (2004), la profundidad del conocimiento del vocabulario constituye un componente esencial de la competencia lingüística, pues implica comprender las asociaciones semánticas y combinatorias que una palabra mantiene con otras en el discurso. Así, la inclusión de este tipo de reactivos no solo permite medir la amplitud del vocabulario de los participantes, sino también su conocimiento colocacional y su capacidad para identificar combinaciones léxicas típicas del español.

Tabla 3

Ejemplos de elementos integrados en el cuestionario

Pregunta	Posibles respuestas
Se _____ falso en la escuela.	esparció un rumor dijo un rumor corrió un rumor
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.

Fuente: elaboración propia.

Basándonos en la taxonomía mencionada previamente, en total, se identificaron veinte categorías distintas y se incluyeron dos reactivos por cada una de las categorías identificadas.

3.2.2. Segunda sección: producción escrita

Como se menciona en el apartado 2.3.1., los textos de aprendices enfrentan el reto de la autenticidad de los textos (Sinclair, 1991). Es por ello que se decidió observar el fenómeno

colocacional a través de muestras de lenguaje natural. Así, para efectos de la presente investigación, se optó por crear un corpus propio.

Con la finalidad de garantizar la validez de los datos recopilados, se consultaron algunos recursos de producción escrita de aprendices de ELE. En particular, se tomaron en cuenta algunos corpus de aprendices como CEDEL2 (Lozano y Mendikoetxea, 2013), así como algunas guías de examen para la certificación del español, entre ellas la del CELA (UNAM, 2015). El diseño final de esta sección se describe en el apartado 3.3.2.

3.3. Procedimiento

En esta sección se presenta la toma de decisiones que se llevó a cabo en torno al instrumento de recopilación de datos. Así mismo, se proporciona una descripción detallada de las distintas etapas de diseño y prueba, así como los elementos que lo conforman.

Los métodos de validación de pruebas se encuentran al centro de la investigación las lenguas (Xi, 2008). De acuerdo con Xi, la forma en que se conceptualiza la validez de una investigación determina el alcance y la naturaleza de la misma y, por tanto, los métodos para reunir pruebas. Los marcos de validación especifican el proceso utilizado para priorizar, integrar y evaluar la evidencia recopilada utilizando varios métodos. Con la finalidad de diseñar un instrumento para recopilación de datos que proporcionara información válida, se llevó a cabo un proceso de tres etapas en las que se probó, ajustó, e implementó el instrumento.

Para el diseño del instrumento, denominado Estudio de colocaciones ELE, fue necesario hacer tres pruebas piloto. En esta sección se describirán dichas etapas de desarrollo y diseño del cuestionario, las cuales serán denominadas como Prueba piloto 1, Prueba piloto 2 y, por último,

Estudio de colocaciones ELE. Así mismo, se describe el proceso de pilotaje cada uno de los tres cuestionarios diseñados y los elementos que influyeron en la toma de decisiones para cada rediseño.

3.3.1. Pruebas piloto

La prueba piloto 1 primero solicita datos generales para correlacionar el conocimiento de colocaciones con el perfil de los aprendices de ELE. Este apartado del cuestionario se puede apreciar en su totalidad en el Anexo 1. Este apartado se mantuvo sin cambios en la versión final del instrumento.

Para la primera sección de esta primera prueba, se diseñó un cuestionario de opción múltiple enfocado en la competencia colocacional, basado en una serie de errores recurrentes identificados en aprendices de ELE por Alonso Ramos et al. (2010). Sin embargo, en esta primera versión, se observó que la mayoría de los participantes respondieron correctamente a los reactivos, independientemente de su perfil de aprendizaje. Además, al no seguir un patrón estandarizado, los reactivos no proporcionaban información suficiente para medir la competencia colocacional.

Para la segunda sección de la prueba piloto 1, orientada a la producción escrita de colocaciones, se consultó una variedad de corpus de aprendices del español. Como resultado, para la sección de producción escrita de la prueba piloto 1 se decidió trabajar con una de las pautas del corpus CEDEL2 (Lozano y Mendikoetxea, 2013), la cual solicitaba que los participantes escriban acerca de su película o libro favorito. Sin embargo, se observó que muchas respuestas recopiladas no eran adecuadas para el objetivo de observar el fenómeno colocacional. Por ejemplo, algunos participantes sólo escribieron el nombre o título de la película o libro, otras

respuestas fueron redactadas en inglés, y algunos parecían haber copiado reseñas de internet, utilizando un lenguaje estilizado propio de un escritor profesional.

Ante estas observaciones, se decidió tanto ajustar el cuestionario en sus preguntas de opción múltiple, como rediseñar la sección de producción escrita, optando por una pauta que incentivara respuestas más espontáneas y personales, reduciendo así la posibilidad de que copiaran de algún sitio de internet.

Para la prueba piloto 2, los reactivos de opción múltiple se reformularon basados en las taxonomías de colocaciones en español identificadas por Corpas Pastor (1996), Koike (2001), y Travalía (2006), descritas en el apartado 2.3.1. y presentadas en la tabla 1. Se diseñaron 40 reactivos, incluyendo dos preguntas por cada una de las 20 taxonomías identificadas.

En la sección de producción escrita se propuso una pauta de escritura que permitiera a los participantes dar una opinión personal sobre un tema, con el fin de obtener muestra de lenguaje más natural en comparación con las respuestas obtenidas en la prueba piloto 1. En esta versión, se solicitó a los participantes dar consejos sobre cómo continuar estudiando español durante el confinamiento por COVID-19. Esta pauta fue diseñada específicamente para esta investigación.

Una vez finalizado el rediseño de la prueba piloto 2, se aplicó nuevamente el cuestionario en grupos de redes sociales. En esta ocasión, se observó que la distribución de preguntas fue más equilibrada en la sección de opción múltiple. Estos resultados, en conjunto con la distribución de la taxonomía de colocaciones empleada, indicaron que la versión ajustada del cuestionario proporcionaba datos más adecuados para el análisis del fenómeno colocacional. En la tabla 4 se ofrecen dos ejemplos de reactivos.

Tabla 4*Prueba piloto 2*

Pregunta	Posibles respuestas
Escuché la noticia de una fuente _____.	Veraz fidedigna verdadera
El niño comió _____.	una tajada de pan un corte de pan una rebanada de pan

Nota: Las palabras en negrillas indican la opción de respuesta correcta

El primer reactivo de la tabla 4 ilustra una colocación léxica simple del tipo *Adjetivo + Sustantivo*, mientras que el segundo reactivo corresponde a una colocación léxica simple del tipo *Sustantivo + Preposición + Sustantivo*.

Sin embargo, se observó que los reactivos no estaban estandarizados en cuanto a la colocación a evaluar. Como se observa en la tabla 4, algunas preguntas requerían identificar la colocación completa, mientras que otras solicitaban únicamente el colocativo. Con base en esto, se decidió ajustar la estructura de los reactivos.

En cuanto a la segunda sección de producción escrita, un análisis de los datos obtenidos reveló que la pauta proporcionada no generó las muestras de lenguaje necesarias para un análisis colocacional adecuado. Algunas respuestas consistieron en una sola palabra o dos, en su mayoría sustantivos o verbos en gerundio como: “leyendo, viendo YouTube, online, series, etc.”.

Si bien, estas respuestas son válidas en términos generales, no proporcionaban combinaciones léxicas que permitieran observar el fenómeno colocacional. Por lo tanto, se

determinó que la pauta se debía ajustar para incentivar la producción de textos más extensos, lo que facilitaría la identificación de colocaciones en los datos recopilados.

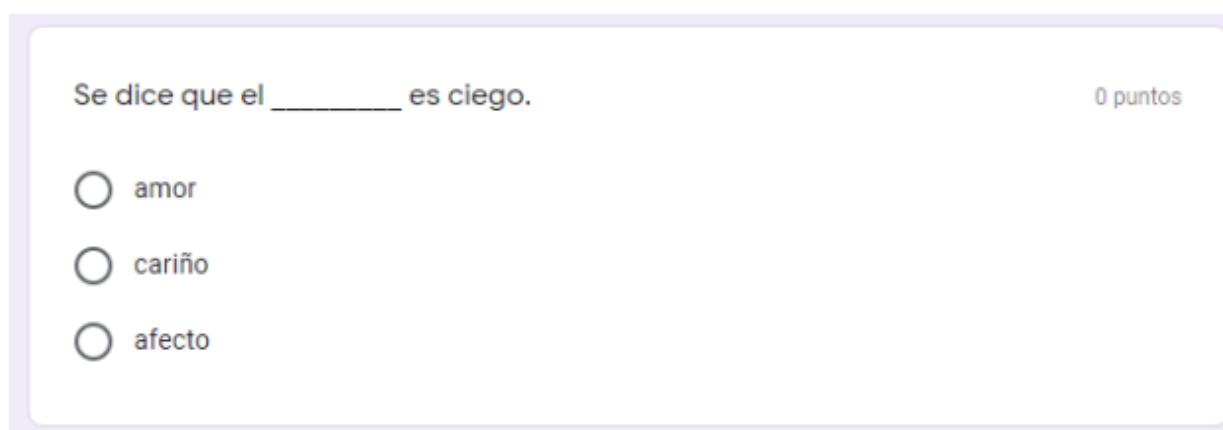
3.3.2. Instrumento: Estudio de colocaciones ELE

Con base en los resultados de las pruebas piloto 1 y 2, se hicieron ajustes al instrumento, lo que llevó al diseño final del Estudio de colocaciones ELE, el cual conserva la estructura general de las pruebas piloto 1 y 2, con mejoras en la homogeneización de los reactivos y en la pauta de producción escrita.

El Estudio de colocaciones de ELE inicia con la misma sección de preguntas del perfil del estudiante utilizadas en las pruebas piloto 1 y 2. Posteriormente, la primera sección mantiene la estructura de las pruebas anteriores, compuesta por 40 reactivos de opción múltiple, diseñados para evaluar la competencia colocacional de los participantes. Estos reactivos se basan en las taxonomías previamente identificadas en el apartado 2.1.3. y presentadas en la tabla 1. No obstante, se realizaron ajustes a los reactivos de la prueba piloto 2 con la finalidad de homogenizar las opciones de respuesta. La figura 1 y figura 2 se presentan ejemplos de estos ajustes en los reactivos de opción múltiple. Para mayor detalle sobre las preguntas de opción múltiple, referirse al anexo 2.

Figura 1

Ejemplo de reactivo de la prueba piloto 2



Se dice que el _____ es ciego.

0 puntos

- ☐ amor
- ☐ cariño
- ☐ afecto

Figura 2

Ejemplo de reactivo del instrumento final Estudio de Colocaciones ELE



Se dice que el amor es _____.

0 puntos

- ☐ intenso
- ☐ caprichoso
- ☐ ciego

Fuente: Elaboración propia

En la figura 1, correspondiente a la prueba piloto 2, se presenta un reactivo de opción múltiple de la colocación “amor ciego”. Esta es una colocación de tipo *Sustantivo + Adjetivo*. En esta versión, se solicitaba a los participantes identificar la base de la colocación (amor). En la figura 2, correspondiente al diseño final, se utilizó el mismo reactivo, pero en esta ocasión, se proporcionó la base de la colocación, dejando en blanco el colocativo para que los participantes lo identificarán. Se concluyó que todos los reactivos debían presentar la base de la colocación para que los participantes seleccionaran el colocativo.

Una vez que se recopilaron las repuestas de los participantes, se clasificaron las respuestas de los 40 reactivos en correctas e incorrectas. A su vez, estos resultados fueron cuantificados y categorizados según el tipo de colocación de acuerdo con la taxonomía de colocaciones del español (Corpas Pastor, 1996; K. Koike, 2001; Travalía, 2006).

Para la segunda sección, enfocada en la producción escrita, fue necesario encontrar una nueva pauta de escritura que incentivara a los participantes a producir textos más extensos que en las pruebas piloto anteriores. Con este objetivo, se consultaron diversas guías para exámenes de certificación del español, incluyendo el Certificado de español como lengua adicional (CELA) de UNAM (2015). En la sección de práctica para la producción escrita de esta guía, se propone la redacción de un correo electrónico como modelo de tarea. Basándose en esta pauta, en la segunda sección del Estudio de colocaciones ELE, se instruyó a los participantes redactar un correo electrónico dirigido a un amigo, disculpándose por no haber asistido a una fiesta. Esta tarea fue diseñada para fomentar respuestas más naturales y espontáneas, evitando la posibilidad de respuestas breves o copiadas de fuentes externas. Para consultar el diseño completo del Estudio de colocaciones ELE, favor de referirse al anexo 3.

Una vez realizados los ajustes en la primera sección del cuestionario, y definido una nueva pauta de escritura para la segunda sección, se puso a prueba esta tercera versión del instrumento.

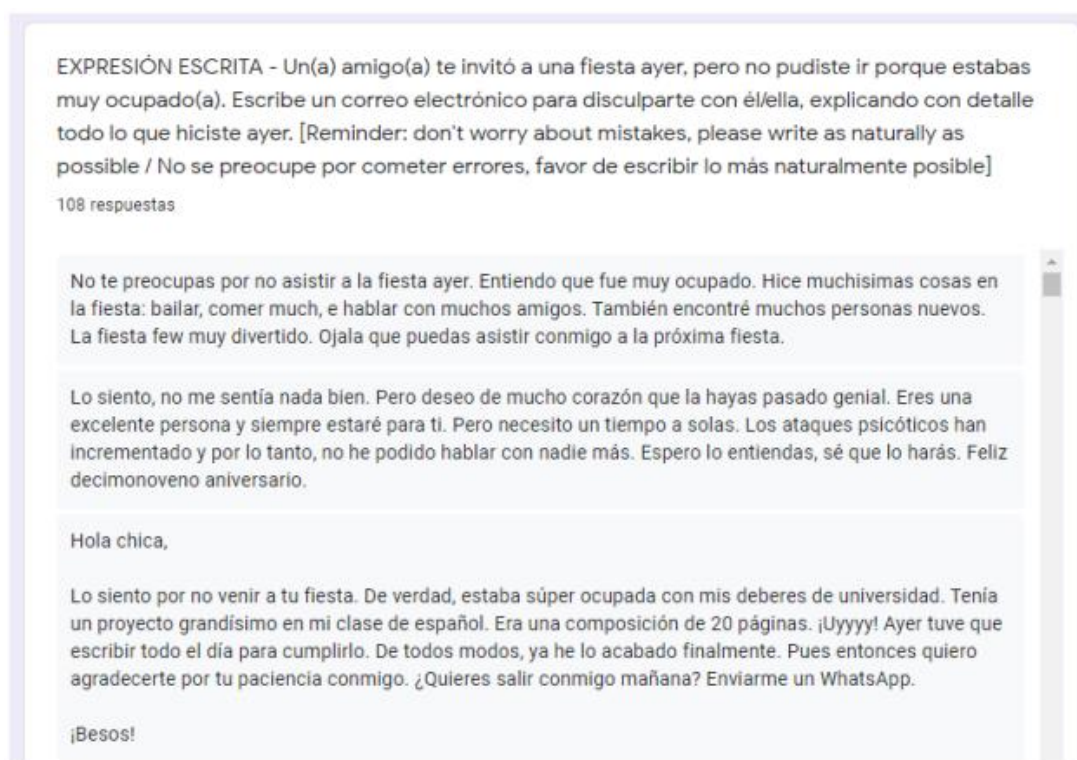
Los resultados preliminares a partir del Estudio de colocaciones ELE mostró que los datos recopilados eran relevantes para la presente investigación. En la primera sección de reactivos de opción múltiple, los resultados fueron similares a los de la prueba piloto 2, lo que

permitió identificar reactivos con mayor índice de error y con mayor grado de reconocimiento, según la taxonomía de las colocaciones del apartado 2.1.3. y la tabla 1.

En cuanto a la segunda sección de producción escrita, la figura 3 proporciona algunos ejemplos de las muestras de escritura obtenidas en el Estudio de colocaciones ELE.

Figura 3

Ejemplos de producción escrita



Fuente: Elaboración propia

En la figura 3 se observa que las muestras de producción escrita en la versión final son más extensas que en las pruebas piloto anteriores. Así mismo, si bien hay errores en la lengua, se logró identificar el uso de colocaciones en estos escritos de los participantes. Entre las

respuestas, se encontraron combinaciones léxicas comunes, así como colocaciones producidas con errores.

A partir de estos resultados, se concluyó que el Estudio de colocaciones ELE representa la versión final del instrumento de recopilación de datos. Con este diseño definitivo, el cuestionario se difundió en diversos grupos de redes sociales dirigidos a aprendices del español. Así mismo, se compartió con estudiantes extranjeros, no nativo-hablantes del español, que han cursado estudios de ELE en UABC.

3.3.2.1 Compilación, preparación y arquitectura del corpus

Una vez recopilados los datos de la sección de producción escrita del Estudio de colocaciones ELE, se procedió a conformar el corpus de aprendices. Para facilitar el análisis de los textos, se prepararon para su procesamiento mediante un gestor de corpus. Con este fin, cada muestra de producción escrita se trató como un texto independiente y se convirtieron en texto plano para su procesamiento.

Cada muestra de producción escrita se identificó con una nomenclatura única basada en los datos estadísticos proporcionados por los participantes, como se detalla en la tabla 5. Esta nomenclatura clasifica cada muestra según el tipo de sujeto (redes sociales o de la institución), lengua nativa, género, nivel de dominio del español, y la posesión de alguna certificación oficial de dominio de la lengua española. Para más detalles sobre la clasificación de las muestras, ver la tabla 5.

El proceso de asignación de nomenclatura individual permite organizar y categorizar las muestras de manera que facilita su análisis y comparación en el estudio. Esta estructura garantiza

la identificación precisa de cada muestra dentro del Corpus de Aprendices de Colocaciones de ELE (CACE), lo que resulta fundamental para realizar un análisis detallado y exhaustivo de los datos recopilados.

Tabla 5
Nomenclatura del CACE

Tipo de sujeto	L1	Código L1	Género	Nivel de español	Certificado
FB	ALEMAN	AL	H	BAS	CERT
UNI	ARABE	AR	M	INT	NOCERT
	AZERI	AZ		AV	
	BULGARIAN	BL		SOB	
	BURMESE	BR			
	CEBUANO	CB			
	CHINO	CH			
	DANISH	DN			
	ESLOVACO	SL			
	FILIPINO	FL			
	FRANCES	FR			
	GRIEGO	GR			
	HEBREW	HB			
	HINDI	HI			
	HÚNGARO	HUN			
	INDONESIO	IN			
	INGLES	EN			
	ITALIANO	IT			
	KIRUNDI	KI			
	LITHUANIAN	LI			
	NEERLANDES	NE			
	NEPALÍ	NP			
	NORUEGO	NO			
	POLACO	PL			
	PORTUGUES	PR			
	RUMANO	RM			
	RUSO	RS			
	SERBOCROATA	SC			
	SWEDISH	SW			
	TAILANDES	TI			

TURCO	TR
URDU	UR
VIETNAMITA	VT

Nota: Nomenclatura para la clasificación de muestras de producción escrita

La estructura de metadatos utilizada en el CACE se diseñó con el propósito de garantizar una organización sistemática de la información y facilitar la búsqueda y clasificación de los textos. En este sentido, las variables incluidas en la tabla 5 guardan correspondencia con los lineamientos recientes propuestos en la investigación de corpus de aprendices. Particularmente, el *Core Metadata Schema for Learner Corpora* (LC-meta, v.2) desarrollado por Paquot et al. (2024) sugiere un conjunto básico de categorías lingüísticas, sociolingüísticas y pedagógicas, destinadas a estandarizar la codificación de la información de los participantes y de las tareas de recopilación. El esquema empleado para el CACE incorpora varios de estos elementos, como la lengua materna, el nivel de dominio del español, la edad, el género, los años de estudio y el contexto de aprendizaje, lo cual favorece la compatibilidad del corpus con otros proyectos de investigación y su posible integración en estudios comparativos futuros.

De acuerdo con los parámetros de la tabla 5, una muestra de producción escrita podría tener la siguiente nomenclatura: FB-AL-M-INT-CERT. Esto permite identificar que el texto fue producido por un individuo del grupo de redes sociales, cuya lengua nativa es el alemán, de género masculino, con un nivel intermedio de dominio de español y que cuenta con un certificado oficial de dominio de la lengua española. Esta nomenclatura es de utilidad para realizar comparaciones entre los fenómenos colocacionales observados y los individuos que los generaron, utilizando el Corpus de Aprendices de Colocaciones de ELE (CACE)

Como se mencionó en el párrafo anterior, se contrastó el corpus conformado por las muestras de producción escrita proporcionadas por los sujetos de redes sociales con el corpus conformado por las muestras de producción escrita producidas por los estudiantes de la institución.

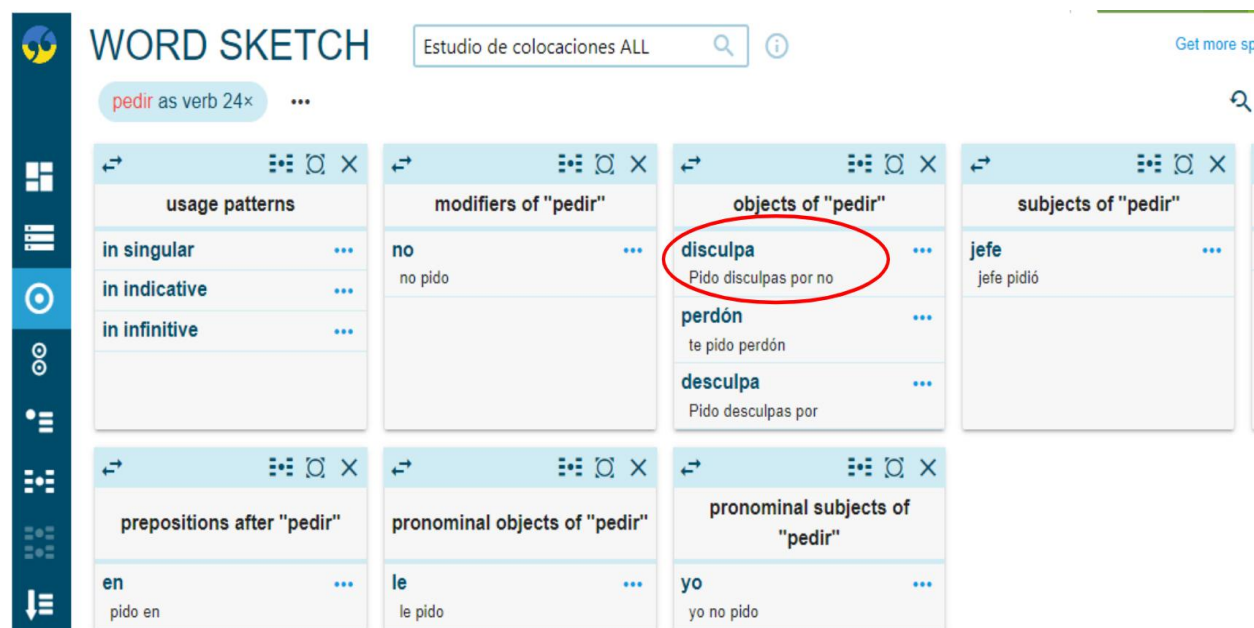
3.3.2.2 Procesamiento a través de software de gestión de corpus

Para la preparación del corpus, se utilizó *Sketch Engine* como herramienta de procesamiento de datos de corpus. Se emplearon diversas funciones de esta herramienta para realizar el análisis del corpus de aprendices. A continuación, se presenta cómo se utilizaron algunas de las funciones del software de gestión de corpus para el análisis de colocaciones.

Se generaron listas de palabras haciendo uso de la función *Wordlist*. Basándose en las taxonomías de colocaciones ya presentadas, se generaron listas de todos los sustantivos, verbos, adjetivos y adverbios dentro del CACE. Como se menciona anteriormente, 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 aparecieran al menos cinco veces en cada una de las listas. Así, cada una de estas unidades léxicas se analizaron en la función *Word sketch*, según la figura 4, permitiendo observar la palabra en contexto y con cuáles otras palabras se combina dentro del CACE.

Figura 4

Ejemplo del CACE en software de gestión de corpus



Nota: Función *Word sketch*, la cual muestra la palabra en contexto y las palabras que se relacionan con esta
Fuente: Elaboración propia, a partir de Sketch Engine

Como se puede observar en la figura 4, la función denominada *Word sketch*, de *Sketch Engine* (Kilgarriff et al., 2014), fue utilizada para observar los elementos con los que coaparece una palabra. Por ejemplo, al analizar el verbo *pedir* mediante esta función, se pudo observar los modificadores con los que comúnmente se combina, así como los verbos con los que está colocando. Es decir, la función *Word Sketch* nos muestra los elementos colocacionales que, de manera más frecuente, se relacionan con la palabra a analizar dentro del corpus.

Se decidió considerar cualquier par de palabras que coaparecen al menos dos veces dentro del CACE como una colocación potencial. La función de concordancia de *Sketch Engine* se utilizó para este fin. La figura 5 muestra un ejemplo de cómo utilizó esta función.

Figura 5

Ejemplo de análisis de colocaciones dentro del software de gestión de corpus

CONCORDANCE Estudio de colocaciones ALL

CQL **pedir + disculpa** • 13
896.37 per million tokens • 0.09%

Left context KWIC Right context

1	doc#0 a no pudo gracias a su nuevo empleo. De todos modos, quiero pedirte las disculpas por no poder llegar a lo que seguramente fue la fiesta de
2	doc#0 or tu comprensión. Hola. Pido disculpas por no poder asistir Tu fiesta. Estaba a punto de irme
3	doc#0 rte abrazo, Julia. Hola, le pido disculpas que no he podido ir a su fiesta. Ayer fue un día llena
4	doc#0 le las 5PM..increíble. También se me olvidó avisarte. Te pido disculpa y espero que haya otra ocasión de vernos muy pronto.
5	doc#0 fue la fiesta? Creo que fue increíblemente asombroso! Pido disculpas por no ir, ayer estuve muy ocupado. Estudié mucho
6	doc#0 lo amigo, no pude asistir a tu fiesta por problemas personales. Pido disculpas Querido amigo, ti ruego
7	doc#0 ola, Juan ¿Como estás? Espero que estés bien. Quiero pedirte disculpa por no comparecer en tu fiesta ayer. Estuve en el cen
8	doc#0 ro que todos hayan disfrutado y disfrutado de su hospitalidad. Pido disculpas por no poder asistir. Estoy haciendo un curso de esp
9	doc#0 io fui a la fiesta. Querido amigo, te pido disculpas por no haber ido contigo a la fiesta. Desafortunadam
10	doc#0 ien por su fiesta y malo por quedarse más vieja. (jejeje) Bueno, quiero pedirte disculpa por mi ausencia en su fiesta. Estoy num período de rr
11	doc#0 ia demasiado agotado por las tareas que estaba realizando. Te pido disculpas , te visitaré esta noche. besos!
12	doc#0 Hola buenos días amigo, Pido disculpas por no asistir a tu fiesta ayer. Estaba enseñando aul

Nota: Función *Concordance*, la cual muestra la concordancia lingüística de la colocación *pedir disculpas* dentro del corpus

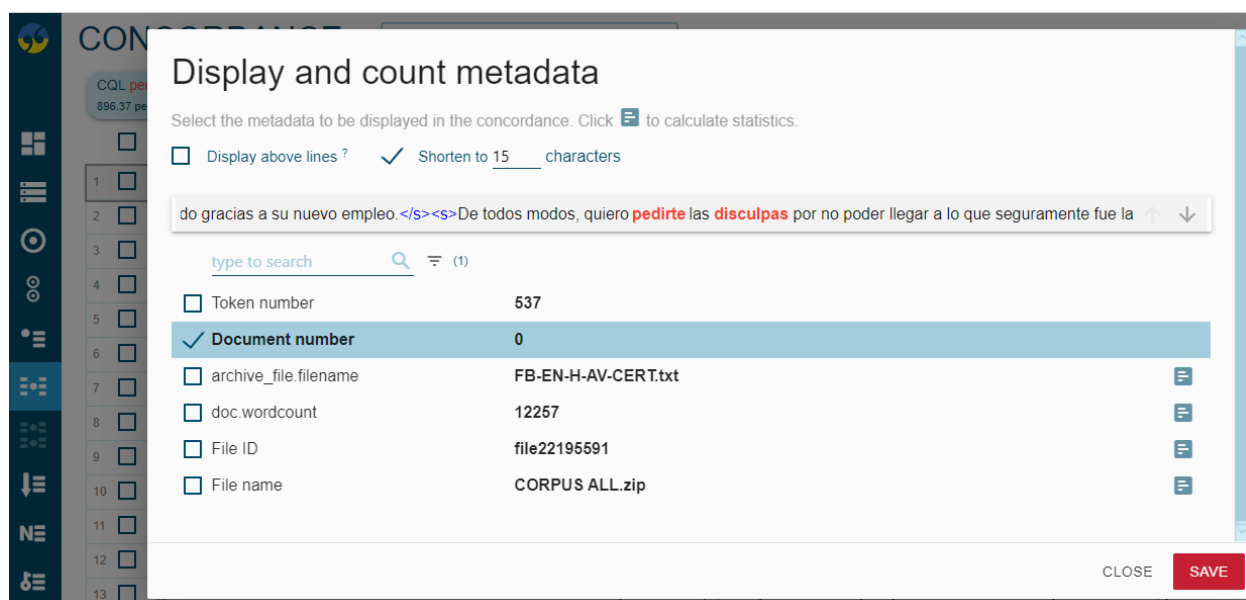
Fuente: Elaboración propia, a partir de Sketch Engine

En la figura 5, se puede observar la función *Concordance* de *Sketch Engine*. Esta herramienta fue utilizada para examinar el contexto en el que una colocación en particular es producida dentro del corpus. Permite observar las frases producidas por aprendices, poniendo énfasis en la colocación buscada. En este ejemplo, se utilizó la colocación léxica simple *pedir + disculpas* (adjetivo + sustantivo).

Tanto la función de *Word sketch*, que permite analizar colocaciones, como la de *Concordance*, que permite ver ejemplos de uso de una palabra en contexto, proporcionan, además, metadatos sobre cada uno de los textos que conforman el corpus. Esto se ilustra en la figura 6.

Figura 6

Metadatos del CACE



Fuente: Elaboración propia, a partir de Sketch Engine

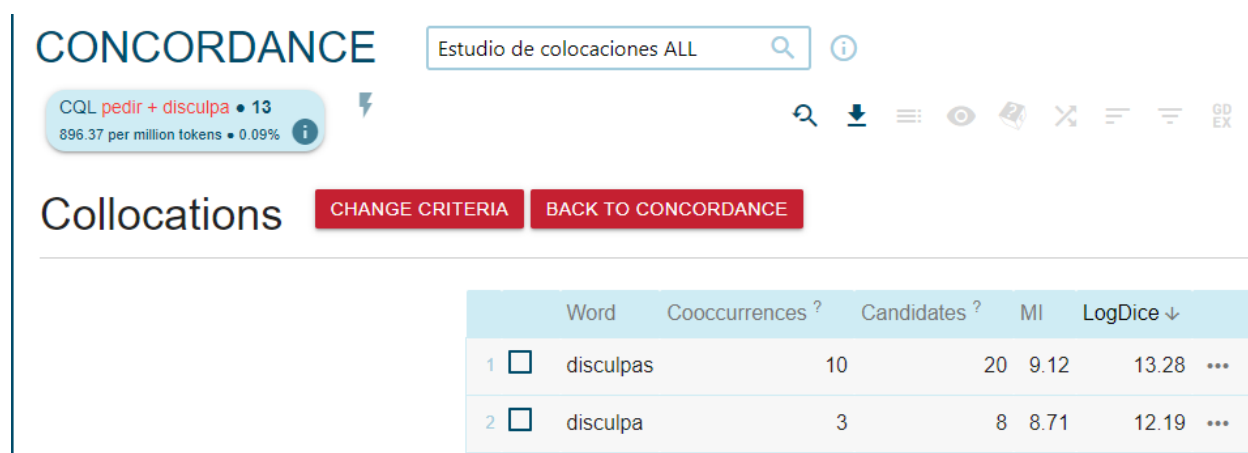
La figura 6 ilustra la función de *Sketch Engine* que proporciona información sobre cada entrada del CACE. Se puede apreciar que en esta función se visualiza la nomenclatura que se asignó a cada muestra de producción escrita del CACE, de acuerdo al perfil de los sujetos participantes. Los detalles sobre la nomenclatura del CACE se pueden consultar en tabla 5.

Adicionalmente, se consultó el índice de información mutua entre las combinaciones de palabras encontradas en el corpus de aprendices. El índice de información mutua (IM), 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. En este estudio, el índice de información

mutua se empleó como una medida de la fuerza de asociación entre los elementos de una colocación, es decir, del grado de atracción estadística que presentan dos palabras cuando aparecen juntas con una frecuencia mayor a la esperada por azar. Las autoras afirman que un número mayor a tres debería considerarse importante. En la figura 7 se muestra el IM de la colocación *pedir* + *disculpas*.

Figura 7

Índice de información mutua



Fuente: Elaboración propia, Sketch Engine

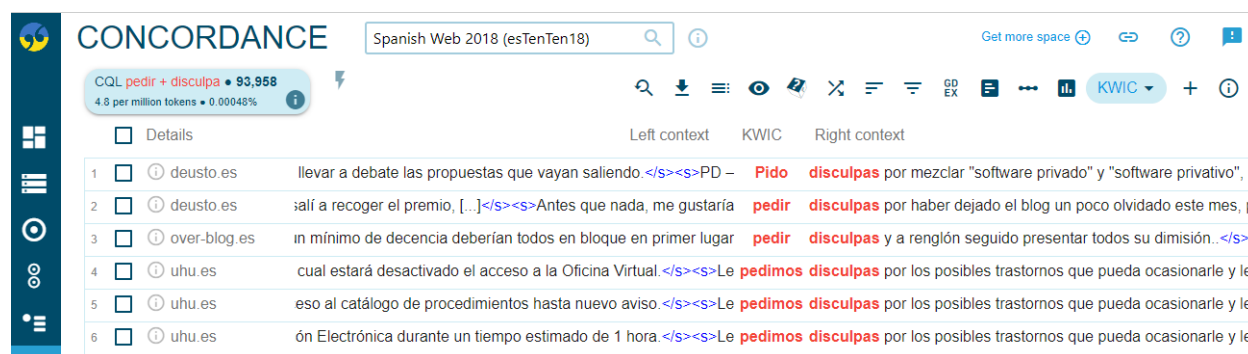
Al revisar el IM, se puede apreciar que aparece la palabra *disculpa* tanto en singular como en plural, de acuerdo con la figura 7. Ambos renglones cuentan con un $IM \geq 3$. Esto permite visualizar que la combinación *pedir* + *disculpa* es, en efecto, una colocación válida.

Alonso Ramos (2016) señala que, para analizar datos, es común emplear algún método contrastivo. En este caso, para contrastar los datos obtenidos en el CACE se decidió hacer uso del corpus titulado *Spanish Web 2018* (Kilgarriff et al., 2014). Este corpus, disponible en la plataforma *Sketch Engine*, está conformado por textos producidos por hablantes nativos y constituye un corpus de español general, no de aprendices. Su función dentro de este estudio es

servir como corpus de referencia para contrastar las colocaciones producidas por los aprendices con aquellas empleadas en el uso natural de la lengua. Dentro del *Spanish Web 2018* se consultaron todas las colocaciones encontradas en el CACE con la finalidad de verificar su validez a partir del *IM Score*, como se muestra en la figura 8.

Figura 8

CACE vs. Spanish Web 2018



Fuente: Sketch Engine

En esta figura 8, se puede apreciar que la colocación de *pedir + disculpas* aparece más 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 Capítulo IV se presenta un detalle de las colocaciones encontradas en el CACE.

3.4. Población y muestra

La presente investigación se llevó a cabo durante el confinamiento a causa de Covid-19. Debido a esto, no fue posible tener contacto directo con aprendices de ELE en la primera etapa de la investigación, pero sí en una segunda. Por esta razón, se tomó la decisión de trabajar con dos grupos distintos: por un lado, se recurrió a redes sociales para buscar estudiantes de ELE; por otro, se trabajó con un grupo de estudiantes extranjeros que han participado en cursos virtuales

de ELE con la UABC. Así, fue posible recopilar datos de participantes con antecedentes y perfiles variados, con base en los criterios de la tabla 2.

El primer grupo con el que se trabajó, se denomina *redes sociales*. De acuerdo con Farzindar e Inkpen (2018), en los últimos años las redes sociales en línea han revolucionado la comunicación interpersonal. El procesamiento del lenguaje natural (PNL) es una de las vías más utilizada para el tratamiento de datos de las redes sociales. Otros autores como Hirschberg y Manning (2015), también establecen que los investigadores en materia de procesamiento de lenguaje natural se sirven cada vez más de las redes sociales para obtener información real y actualizada sobre el uso que se les da a las lenguas.

El segundo grupo de sujetos participantes, conformado por estudiantes extranjeros de la universidad, se denomina *UABC*. Con relación a este grupo que ha participado en cursos de ELE con la UABC, se puede decir que este sector de la población representa una muestra con parámetros estadísticos más uniformes que aquellos participantes de redes sociales. Dichos estudiantes extranjeros de UABC coinciden en rangos similares de edad, grados similares de estudios, e incluso tienen experiencia con la misma variedad dialectal de español. Así mismo, el proceso de aprendizaje de ELE que han llevado a cabo, proviene de cursos con una estructura más estandarizada, debido a la estructura de los diseños de cursos de lenguas extranjeras de la universidad.

La muestra está conformada por sujetos del grupo de redes sociales y sujetos del grupo de UABC, ambos grupos contando con la participación de 144 aprendices respectivamente. Así, se logró obtener datos de un total de 288 participantes. De acuerdo con las preguntas sobre el perfil

de los participantes del cuestionario, detalladas en la tabla 2, se cuantificó el perfil de los participantes. A continuación, en la tabla 6, se muestran las edades señaladas por ambos grupos.

Tabla 6

Edad de los participantes

Redes Sociales		UABC		Total
Rango de Edad	Porcentaje	Rango de Edad	Porcentaje	Porcentaje
15-20	9%	15-20	15%	12%
21-25	23%	21-25	33%	28%
26-30	20%	26-30	8%	14%
31-35	15%	31-35	8%	12%
36-40	12%	36-40	7%	10%
41-45	4%	41-45	8%	6%
46-50	3%	46-50	6%	5%
50-55	5%	50-55	7%	6%
56-60	5%	56-60	6%	5%
61 en adelante	4%	61 en adelante	2%	3%

Fuente: Elaboración propia

De acuerdo con la tabla 6, las edades de los sujetos de redes sociales se ubican principalmente entre 21 a 40 años de edad. Por su parte, las edades de los estudiantes de UABC presentan un rango de edad más variado, desde los 20 hasta los 60 años de edad. En cuanto al género de los participantes, el 65% del grupo de redes sociales declaró ser mujer, mientras que el 74.3% del grupo de UABC se identifica como género femenino.

Por otro lado, se consideró relevante obtener información sobre el nivel de dominio de la lengua española de los participantes. Sin embargo, debido a la pandemia y a los tiempos de esta investigación, no fue posible aplicar un examen de dominio de ELE a los sujetos. Por este

motivo, se les solicitó que nos proporcionaran el nivel de dominio de español que cada uno de ellos considera tener. De acuerdo con los parámetros señalados en la tabla 2, la tabla 7 muestra la información compartida por los participantes de ambos grupos.

Tabla 7

Nivel de dominio de español

Redes Sociales		UABC		Todos
Nivel de Español	Porcentaje	Nivel de español	Porcentaje	Porcentaje
Básico	24%	Básico	52%	38%
Intermedio	37%	Intermedio	36%	37%
Avanzado	26%	Avanzado	10%	18%
Sobresaliente	13%	Sobresaliente	1%	7%

Nota: Los niveles de español representan la opinión personal de cada sujeto participante, en relación al nivel de dominio de la lengua que cada uno de ellos considera tener

A partir de la tabla 7, se puede observar que los participantes de redes sociales ubican el nivel de dominio del español mayormente entre intermedio y avanzado. En cambio, los participantes de UABC, entre los niveles básico e intermedio. Los sujetos de redes sociales mencionaron que el 69.7% no cuenta con alguna certificación oficial de dominio de la lengua. En cuanto a los sujetos de UABC, el 86.8% no cuenta con alguna certificación oficial de dominio de la lengua. Únicamente el 23% de la población total afirmó contar con algún certificado oficial.

Sobre el país en donde estudiaron el español, el 81.7% de los participantes del grupo de redes sociales indicó que estudiaron español dentro de su país de origen, mientras que el 82.5% de los participantes del grupo de UABC afirmó que estudió en su país.

Las lenguas nativas de los participantes se pueden apreciar en la tabla 8. El grupo de redes sociales cuenta con la mayor diversidad de L1, con 33 lenguas nativas distintas. El inglés es la lengua con mayor número de participantes, representando un 46% de la población, seguido por el italiano con un 4.8%. En el grupo de UABC, se identificaron nueve lenguas nativas distintas, destacando el portugués como la más común entre los participantes con un 92.3%, seguido por el turco con un 3.4%.

Tabla 8

L1 de los participantes

Redes Sociales		UABC		Total
Lengua nativa	Porcentaje	Lengua nativa	Porcentaje	Porcentaje
Alemán	2%	Alemán	2%	1.92%
Árabe	2%			1.15%
Azerí	1%			0.38%
Birmano	2%			0.76%
Búlgaro	2%			0.76%
Cebuano	1%			0.38%
Chino		Chino	2%	0.77%
Danés	1%			0.38%
Eslovaco	1%			0.38%
Filipino	3%			1.53%
Francés	6%	Francés	2%	3.82%
Griego	2%			1.15%
Hebreo	2%			0.76%
Hindi	2%			0.76%
Húngaro		Húngaro	1%	0.38%
Indonesio	2%			0.76%
Inglés	51%	Inglés	1%	25.96%
Italiano	5%	Italiano	2%	3.44%
Kirundí	1%			0.38%
Lituano	1%			0.38%
Neerlandés	1%			0.38%
Nepalí	1%			0.38%
Noruego	2%			0.76%
Polaco	2%			1.15%

Portugués	3%	Portugués	87%	44.99%
Rumano	2%			0.76%
Ruso	2%			0.76%
Serbocroata	1%			0.38%
Sueco	1%			0.38%
Tailandés	1%	Tailandés	1%	0.77%
Turco	1%	Turco	4%	2.30%
Urdu	1%			0.38%
Vietnamita	1%			0.38%

Nota: Los espacios vacíos son equivalentes a 0 (cero) sujetos participantes.

La tabla 8 muestra que las lenguas nativas de los participantes de ambos grupos son principalmente lenguas indo-europeas, destacando el inglés y portugués como las más representadas en la muestra.

Por último, en cuanto a las muestras de producción escritas, si bien se contó con 144 participantes del grupo de redes sociales, únicamente 94 sujetos proporcionaron muestras, mientras que de los 144 participantes del grupo de estudiantes de UABC, 128 sujetos proporcionaron muestras de producción escrita.

En este capítulo se presentó el procedimiento que se llevó a cabo para diseñar un instrumento de recopilación de datos que permitiera observar la producción de colocaciones de aprendices de ELE. Así mismo, se ofrece una descripción de la población participante, así como algunos ejemplos de los pasos que siguieron para la identificación de colocaciones. El siguiente capítulo describe el análisis de datos realizado a partir de los datos recolectados y la discusión en torno a los resultados obtenidos.

Capítulo IV. Análisis y discusión de resultados

En este capítulo se presentan los resultados obtenidos a partir del Estudio de colocaciones ELE y sobre la compilación del corpus de aprendices, denominado CACE. Posteriormente, se ofrece un análisis y una categorización de los fenómenos observados.

4.1. Resultados sobre el conocimiento colocacional de los aprendices de ELE a partir del cuestionario

En este apartado se presentan los resultados sobre el conocimiento colocacional, obtenidos a través de la primera sección del instrumento de recopilación de datos. Primero, se introducen los resultados del grupo de redes sociales; después, los del grupo de UABC. Finalmente, se presenta un concentrado del resultado de ambos grupos, para después ofrecer un análisis y discusión de los mismos.

4.1.1. El conocimiento colocacional del grupo de redes

A continuación, en la tabla 9, se resumen los resultados sobre conocimiento colocacional del grupo de participantes de redes sociales. En ella se describen los resultados según la taxonomía de colocaciones propuesta por Corpas Pastor (1996), K. Koike (2001) y Travalia (2006).

Tabla 9

Resultados del cuestionario de redes sociales

Redes Sociales				
Categoría	Tipo de colocación	% de respuestas correctas	% de respuestas incorrectas	% sin respuesta
Léxica simple	Sustantivo (Sujeto) + Verbo	50.3%	45.5%	4.2%

Léxica simple	Verbo + Adverbio (mente)	40.6%	*53.1%	6.3%
Léxica simple	Verbo + Sustantivo (Objeto)	29.9%	*65.6%	4.5%
Léxica simple	Adjetivo + Sustantivo	27.8%	*68.1%	4.2%
Léxica simple	Sustantivo + Preposición + Sustantivo	44.1%	*52.1%	3.8%
Léxica simple	Verbo participio + Adverbio	44.8%	*51.0%	4.2%
Léxica simple	Verbo + Preposición + Sustantivo	63.5%	33.0%	3.5%
Léxica simple	Adverbio + Adjetivo	44.8%	*51.4%	3.8%
Léxica simple	Sustantivo + Adjetivo	45.5%	50.3%	4.2%
Léxica simple	Verbo + Adjetivo	42.4%	*52.8%	4.9%
Léxica compleja	Verbo + locución nominal	53.8%	41.3%	4.9%
Léxica compleja	Locución verbal + sustantivo	42.7%	*52.8%	4.5%
Léxica compleja	Verbo + locución adverbial	29.9%	*63.5%	6.6%
Léxica compleja	Adjetivo + locución adverbial	58.3%	35.8%	5.9%
Gramatical	verbo + preposición	38.2%	*57.6%	4.2%
Gramatical	sustantivo + preposición (+ sustantivo / oración subordinada)	61.1%	34.7%	4.2%
Gramatical	(artículo +) sustantivo + preposición (+ infinitivo)	58.7%	36.1%	5.2%
Gramatical	preposición (+ artículo) + sustantivo (+ preposición)	31.3%	*64.2%	4.5%
Gramatical	(verbo) + adjetivo + preposición	62.2%	33.7%	4.2%
Gramatical	(verbo + sustantivo) + preposición + sustantivo	59.0%	34.7%	6.3%

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.
Fuente: (Corpas Pastor, 1996; K. Koike, 2001; Travalía, 2006)(Corpas Pastor, 1996; K. Koike, 2001; Travalía, 2006)

Así, en la tabla 9 se observa que de los diez tipos de colocaciones léxicas simples (Corpas Pastor, 1996; K. Koike, 2001) sólo tres obtuvieron mayoría de respuestas correctas. Estas fueron

las colocaciones de tipo *Sustantivo (Sujeto) + Verbo*, *Verbo + Preposición + Sustantivo*, y *Sustantivo + Adjetivo*.

En relación con las colocaciones léxicas complejas (K. Koike, 2001), que comprenden cuatro tipos, sólo dos obtuvieron mayoría de respuestas correctas. las de tipo *Verbo + locución nominal* y *Adjetivo + locución adverbial*.

Por último, en relación con las colocaciones gramaticales (Travalia, 2006), esta categoría mostró el mayor grado de reconocimiento positivo entre sus tipos de colocaciones. De los seis tipos propuestos por Travalia (2006), cuatro recibieron mayoría de respuestas correctas. Las únicas colocaciones gramaticales que obtuvieron una mayoría de respuestas incorrectas fueron las de *verbo + preposición* y la de *preposición (+ artículo) + sustantivo (+ preposición)*. Una discusión sobre las posibles implicaciones de estos resultados se presenta en el apartado 4.3.

4.1.2. El conocimiento colocacional del grupo de UABC

En lo que concierne a los resultados sobre el conocimiento colocacional del grupo de participantes de UABC, estos se resumen en la tabla 10, siguiendo la misma taxonomía de colocaciones mencionada en el apartado anterior.

Tabla 10
Resultados del cuestionario de UABC

UABC				
Categoría	Tipo de colocación	% de respuestas correctas	% de respuestas incorrectas	% sin respuesta
Léxica simple	Sustantivo (Sujeto) + Verbo	61.5%	38.2%	0.3%
Léxica simple	Verbo + Adverbio (mente)	37.8%	*61.8%	0.3%
Léxica simple	Verbo + Sustantivo (Objeto)	44.1%	*55.9%	0.0%

Léxica simple	Sustantivo + Preposición + Sustantivo	44.4%	*54.2%	1.4%
Léxica simple	Verbo participio + Adverbio	60.8%	38.9%	0.3%
Léxica simple	Verbo + Preposición + Sustantivo	87.8%	12.2%	0.0%
Léxica simple	Adverbio + Adjetivo	71.2%	28.1%	0.7%
Léxica simple	Sustantivo + Adjetivo	62.5%	37.2%	0.3%
Léxica simple	Verbo + Adjetivo	41.3%	*58.3%	0.3%
Léxica compleja	Verbo + locución nominal	65.3%	34.7%	0.0%
Léxica compleja	Locución verbal + sustantivo	37.5%	*62.2%	0.3%
Léxica compleja	Verbo + locución adverbial	41.3%	*58.3%	0.3%
Léxica compleja	Adjetivo + locución adverbial	58.0%	41.3%	0.7%
Gramatical	verbo + preposición	59.0%	40.6%	0.3%
Gramatical	sustantivo + preposición (+ sustantivo / oración subordinada)	82.3%	17.7%	0.0%
Gramatical	(artículo +) sustantivo + preposición (+ infinitivo)	51.4%	48.3%	0.3%
Gramatical	preposición (+ artículo) + sustantivo (+ preposición)	29.5%	*69.8%	0.7%
Gramatical	(verbo) + adjetivo + preposición	51.4%	47.9%	0.7%
Gramatical	(verbo + sustantivo) + preposición + sustantivo	59.7%	40.3%	0.0%

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.
Fuente: (Corpas Pastor, 1996; K. Koike, 2001; Travalía, 2006)

Al igual que con el grupo participante anterior, la tabla 10 permite observar cuáles son los tipos de colocaciones que tuvieron mayor y menor reconocimiento positivo. De los diez tipos de colocaciones léxicas simples (Corpas Pastor, 1996; K. Koike, 2001) seis obtuvieron mayoría de respuestas correctas. Estas fueron las colocaciones de tipo *Sustantivo (Sujeto) + Verbo, Adjetivo + Sustantivo, Verbo participio + Adverbio, Verbo + Preposición + Sustantivo, Adverbio + Adjetivo*, y *Sustantivo + Adjetivo*.

En relación con la categoría de colocaciones léxicas complejas (K. Koike, 2001), según la misma tabla 10, los resultados fueron similares a los del grupo redes sociales. De los cuatros tipos existentes, dos tuvieron mayoría de respuestas correctas y dos mayoría de respuestas incorrectas.

Por último, en cuanto a la categoría de colocaciones gramaticales (Travalia, 2006), esta también fue la categoría que obtuvo mayor grado de reconocimiento positivo entre sus tipos de colocaciones. De los seis tipos propuestos por Travalia (2006), cinco recibieron mayoría de respuestas correctas. La única colocación gramatical que obtuvo mayoría de respuestas incorrectas fue el de *preposición (+ artículo) + sustantivo (+ preposición)*.

4.1.3. Síntesis de los hallazgos del conocimiento colocacional a partir del cuestionario

En este apartado se ofrece un análisis de los resultados obtenidos de los participantes de redes sociales y UABC, concentrando los datos de ambos grupos. La tabla 11 resume el porcentaje de respuestas correctas e incorrectas de manera global.

Tabla 11

Concentrado de resultados del cuestionario

Concentrado de respuestas de todos los participantes				
Categoría	Tipo de colocación	% de respuestas correctas	% de respuestas incorrectas	% sin respuesta
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%

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.

Fuente: (Corpas Pastor, 1996; K. Koike, 2001; Travalía, 2006)

En la tabla 11 se observa que la categoría de colocaciones gramaticales, de acuerdo con la clasificación propuesta por Travalía (2006), fue la que presentó el mayor grado de reconocimiento entre los participantes, tanto del grupo de Redes Sociales como del grupo de UABC. Este hallazgo coincide con los resultados mostrados en la tabla 9 y la tabla 10, lo cual sugiere una tendencia consistente en el reconocimiento de este tipo de colocaciones.

Un análisis más detallado revela que el grupo de UABC obtuvo un reconocimiento ligeramente superior, con un promedio del 55.55%, mientras que el grupo de Redes Sociales

alcanzó un promedio del 51.75%. En el grupo de Redes Sociales, de la tabla 9 solo cuatro de los seis tipos de colocaciones gramaticales fueron reconocidos correctamente por la mayoría. En contraste, el grupo de UABC, de la tabla 10 mostró un desempeño más sólido, con cinco de los seis tipos reconocidos adecuadamente. Este patrón sugiere que, si bien ambos grupos presentan niveles similares de familiaridad con las colocaciones gramaticales, el grupo de UABC podría contar con un conocimiento más sistemático, posiblemente favorecido por un contexto educativo más estructurado y una instrucción explícita en estructuras gramaticales complejas (Koike, 2001; Travalía, 2006).

Ahora bien, al comparar el reconocimiento de colocaciones gramaticales con el de las colocaciones léxicas, que en general obtuvieron porcentajes más bajos, se observa una diferencia relevante en términos de adquisición lingüística. Las colocaciones gramaticales incluyen palabras como preposiciones, artículos o auxiliares, que se usan con mucha frecuencia y de forma bastante fija en el idioma. Por eso, los hablantes suelen reconocerlas con más facilidad, ya que forman parte de estructuras que se repiten con frecuencia. Como explica Travalía (2006), estas combinaciones funcionan como bloques que se usan de manera muy predecible en la gramática, lo cual puede explicar por qué obtuvieron mejores resultados. En este sentido, se refuerza la importancia de considerar el contexto de aprendizaje como variable explicativa en el desarrollo del conocimiento colocacional (Cheikh-Khamis Cases, 2013b; Ibarretxe Antuñano, 2010).

En cuanto a las colocaciones léxicas simples, en la tabla 11 se observa que, de los diez tipos identificados (Corpas Pastor, 1996; Koike, 2001), cinco presentan un mayor número de respuestas correctas: *Sustantivo (Sujeto) + Verbo*, *Verbo participio + Adverbio*, *Verbo + Preposición + Sustantivo*, *Adverbio + Adjetivo*, y *Sustantivo + Adjetivo*. Este patrón puede

explicarse por la frecuencia de uso y la menor opacidad semántica de estas combinaciones (Alonso Ramos, 1994; Leech, 2011), lo cual facilita su reconocimiento en contraste con otras estructuras menos comunes o más idiosincráticas (Cruse, 1986).

Sin embargo, observar el desempeño general en esta categoría, se constata que las colocaciones léxicas obtuvieron porcentajes de reconocimiento más bajos que las gramaticales. Esto se debe en parte a que las colocaciones léxicas están formadas por palabras de contenido abierto, tal como verbos, sustantivos y adjetivos, los cuales pueden variar de una lengua a otra en cuanto a su combinación habitual. Un ejemplo de ello es la diferencia entre el español *tomar una decisión* y el inglés *make a decision*. Aunque ambas expresiones comunican la misma idea, los verbos involucrados (*tomar* vs. *make*) no son equivalentes ni intercambiables, lo que da lugar a errores por interferencia interlingüística (Hausmann, 1989; Granger, 1998).

Tal como lo señalan Cheikh-Khamis Cases (2013b) e Ibarretxe Antuñano (2010), este tipo de colocaciones depende en gran medida de la competencia pragmática y cultural del hablante, ya que involucra no solo el significado de las palabras individuales, sino también las convenciones de uso propias de cada lengua. A diferencia de las colocaciones gramaticales, que se apoyan en elementos estructurales del sistema lingüístico (como las preposiciones o artículos que tienden a ser fijos y predecibles), las colocaciones léxicas no siguen reglas gramaticales explícitas, sino patrones de uso más difíciles de generalizar. Esta diferencia explica por qué los aprendientes tienden a reconocer con mayor facilidad estructuras gramaticales como *en la mesa* o *por la mañana*, donde la preposición suele ser estable dentro del sistema (Travalia, 2006), mientras que enfrentan más dificultades al intentar identificar combinaciones como *desempeñar un cargo* o *entablar una amistad*, que no siempre tienen equivalentes directos o lógicos en sus lenguas maternas (Koike, 2001; Coseriu, 1981).

Al comparar las tablas 9 y 10, se confirma que el grupo de UABC muestra un porcentaje más alto de reconocimiento tanto en colocaciones léxicas simples como en gramaticales. Resulta especialmente significativo que, a pesar de haber reportado un nivel de dominio del español inferior, este grupo obtuvo mejores resultados que el grupo de Redes Sociales. Esta contradicción refuerza la idea de que el conocimiento colocacional no depende exclusivamente del nivel lingüístico declarado, sino también de factores como la instrucción explícita, el tipo de exposición lingüística, y la metacognición sobre el uso del idioma (Read, 2000; Cheikh-Khamis Cases, 2013b).

Finalmente, ambos grupos muestran desempeños similares en el reconocimiento de colocaciones léxicas complejas. Este dato resulta interesante si se considera que este tipo de colocaciones, a pesar de presentar un mayor grado de fijación y menor variabilidad sintáctica (Koike, 2001), suelen tener un carácter idiomático más marcado, lo que las convierte en unidades más estables, pero menos transparentes desde el punto de vista semántico. Su adquisición requiere una exposición considerable al lenguaje auténtico y al reconocimiento de patrones formulaicos, lo que resalta la importancia del enfoque basado en corpus y la enseñanza explícita de fórmulas lexicalizadas (Jiménez Calderón y Rufat, 2019; Conzett, 2000).

A diferencia de las colocaciones simples, donde los elementos léxicos pueden combinarse con mayor flexibilidad y generar más interferencia entre lenguas, las colocaciones complejas tienden a funcionar como unidades fijas de significado, menos transparentes y más difíciles de interpretar palabra por palabra. Su comprensión depende menos de traducir literalmente y más de haberlas aprendido dentro de un contexto. Como señalan Granger (1998) y Nesselhauf (2005), estas expresiones prefabricadas pueden representar un reto similar para estudiantes de distintos entornos, ya que su aprendizaje no está tan relacionado con el conocimiento gramatical, sino con

el reconocimiento repetido en situaciones reales de comunicación. Esto podría explicar por qué los participantes, sin importar su contexto de aprendizaje, obtuvieron resultados parecidos en esta categoría.

4.1.3.1 Síntesis de los hallazgos del conocimiento de las colocaciones léxicas simples.

En esta sección se analizan las observaciones sobre el conocimiento colocacional de las colocaciones léxicas simples (Corpas Pastor, 2001; K. Koike, 2001) a partir de los datos obtenidos en el cuestionario. Todos los datos provienen de la tabla 9 y la tabla 10.

En el caso de la colocación léxica simple de *Sustantivo (Sujeto) + Verbo*, se observó que se obtuvo una mayoría de respuestas correctas tanto en el grupo de redes sociales como en el grupo de UABC. Algunos ejemplos de este tipo de colocación que se identificaron correctamente incluyen *correr un rumor* y *confirmar una sospecha*. Se constató que estas expresiones son muy similares tanto en inglés como en portugués, tomando en cuenta que son las L1 predominantes en los grupos de redes sociales y de UABC, respectivamente. Por tanto, se infiere que el reconocimiento positivo de este tipo de colocación se debe a la transferencia interlingüística.

A continuación, se presentan otras colocaciones léxicas simples que también mostraron un alto grado de reconocimiento por parte de los participantes (Corpas Pastor, 1996; Koike, 2001). En el cuestionario, los sujetos debían seleccionar la colocación entera que completaba correctamente la oración. Una de estas fue la colocación del tipo *Verbo + Adjetivo*, representada por la expresión *quedó claro*. Esta estructura es semánticamente y formalmente cercana al portugués *deixar claro* y al inglés *to make clear*, aunque los verbos no son equivalentes funcionalmente. La similitud radica principalmente en el adjetivo *claro*, que en las tres lenguas mantiene su forma. Por ello, se considera que el reconocimiento positivo puede estar influido por

un fenómeno de transferencia léxica parcial, donde el elemento compartido (*claro*) facilita la identificación de la colocación correcta, a pesar de que el verbo que la acompaña varíe entre lenguas (Coseriu, 1981; Cheikh-Khamis Cases, 2013b).

De manera similar, la colocación del tipo *Adverbio + Adjetivo*, *visiblemente afectado*, mostró un alto nivel de reconocimiento. En este caso, la transferencia parece centrarse en el adverbio *visiblemente*, cuya forma y función son similares a sus equivalentes en portugués (*visivelmente*) e inglés (*visibly*). Esta coincidencia interlingüística refuerza la hipótesis de que los aprendices pueden apoyarse en el conocimiento previo de otras lenguas para identificar colocaciones en español. Este tipo de apoyo se ha documentado en la literatura sobre interlengua y transferencia lingüística en el aprendizaje de L2 (Cheikh-Khamis Cases, 2013b; Read, 2000). En cuanto a las colocaciones del tipo *Sustantivo + Adjetivo*, como *amor ciego y conducta intachable*, se observaron similitudes tanto en el significado como en la forma de la frases *amor é cego / love is blind* y *conduta irrepreensível / impeccable behavior*. Si bien no son traducciones exactas, la cercanía en el contenido conceptual y el uso de adjetivos con funciones equivalentes apunta a una posible influencia de la transferencia entre lenguas, particularmente en lo que respecta a construcciones adjetivales comunes en idiomas romances y germánicos (Koike, 2001; Alonso Ramos, 1994).

Fenómenos similares se observaron en colocaciones del tipo *Verbo participio + Adverbio* con las colocaciones *estrechamente ligado y profundamente dormido*, y *Adjetivo + Sustantivo* con las colocaciones *fuerza fidedigna* y *relación estrecha*, también reconocidas correctamente por los participantes. En estos casos, tanto la forma en que están construidas las expresiones, como algunas de las palabras utilizadas, son bastante parecidas a sus equivalentes en portugués e

inglés. Esto apoya la idea de que los aprendices pueden reconocer más fácilmente ciertas colocaciones cuando estas comparten características con otras lenguas que conocen, como señala Alonso Ramos (1994) al hablar de la probabilidad y aceptación de ciertas combinaciones de palabras.

Cabe reiterar que todas las colocaciones utilizadas fueron tomadas de las clasificaciones propuestas por Corpas Pastor (1996) y Koike (2001), seleccionadas para representar distintos tipos de combinaciones léxicas, lo cual permite analizar la posible transferencia desde la L1 y otras lenguas previamente adquiridas por los aprendices (Cheikh-Khamis Cases, 2013b; Schmitt y Schmitt, 2020).

Adicionalmente, la colocación de tipo *Verbo + Preposición + Sustantivo* fue reconocida correctamente por los participantes con las combinaciones *andar con bromas* y *poner (algo) en práctica*. Como ejemplo, se puede mencionar que la última combinación se dice *colocar (algo) em prática* en portugués y *put (something) into practice* en inglés. Aunque en las L1 de los participantes estas frases no usan el mismo verbo, se teoriza que los participantes cuentan con conocimiento preposicional que facilita su reconocimiento.

Por otro lado, hubo colocaciones léxicas simples que fueron identificadas de manera incorrecta. El tipo de colocación *Verbo + Adverbio (mente)* fue uno de los tipos de colocación que ambos grupos de participantes respondieron incorrectamente. Por ejemplo, los participantes debían de reconocer la colocación *llorar amargamente*. Sin embargo, ambos grupos escogieron la combinación incorrecta *llorar tristemente*.

Esta tendencia sugiere una posible interferencia lingüística, ya que combinaciones similares existen en las lenguas maternas de los participantes. Sin embargo, un análisis contrastivo de las colocaciones en español, inglés y portugués revela diferencias sutiles en la frecuencia de uso y en la connotación semántica de términos como *amargamente* y *bitterly*. Por ejemplo, aunque *bitterly* es una traducción común de *amargamente*, un análisis en *Sketch Engine* sugiere que en inglés son más frecuentes expresiones como *uncontrollably* y *profusely* en ciertos contextos. En portugués, *Sketch Engine* indica que la colocación *chorar copiosamente* es utilizada con mayor frecuencia que la traducción literal, *chorar amargamente*.

Además de la interferencia lingüística, otros factores podrían haber influido en los resultados como: el enfoque pedagógico utilizado en el aula y las características cognitivas individuales de los participantes. Futuras investigaciones podrían explorar en mayor profundidad la influencia de la lengua materna, así como el papel de otros factores en la adquisición de colocaciones léxicas.

Continuando con el análisis de los tipos de colocaciones léxicas simples que fueron identificadas de manera incorrecta, se observó que algunas de estas combinaciones no tienen una traducción directa a la L1 de los participantes. Entre ellos está el tipo de colocación de *Verbo + Sustantivo (Objeto)* con las combinaciones *desempeñar un cargo* y *entablar una amistad*, y el tipo *Sustantivo + Preposición + Sustantivo* con las combinaciones *rebanada de pan* y *ciclo de conferencias*.

En estos casos, se teoriza que también se trata de un problema de interferencia lingüística. Por ejemplo, los hablantes de español podrían intentar traducir literalmente *desempeñar un cargo* al inglés como **to play a charge*, lo cual es incorrecto. Esta tendencia a aplicar las reglas

gramaticales y el vocabulario de la lengua materna a la lengua meta, conocida como interferencia lingüística, es un fenómeno bien documentado en la adquisición de segundas lenguas. En el caso del inglés y el portugués, esta interferencia puede manifestarse en la formación de colocaciones léxicas incorrectas, como hemos visto en los ejemplos anteriores. Por ejemplo, en español decimos *tomar una decisión*, mientras que en inglés se dice *make a decision*. Esta diferencia se debe a que las colocaciones léxicas son idiomáticas y varían de una lengua a otra.

En este apartado se analizó el conocimiento de colocaciones léxicas simples en ambos grupos de participantes, evaluando su capacidad para identificar combinaciones de palabras correctas. Los resultados mostraron que la influencia de la L1 (inglés y portugués) fue un factor determinante en el desempeño de los participantes.

Se observó que las colocaciones con equivalentes similares en la L1, como *correr un rumor* o *quedar claro*, fueron identificadas correctamente con mayor frecuencia. Esto se atribuye a la transferencia interlingüística, es decir, la aplicación de conocimientos de una lengua a otra.

Sin embargo, también se identificaron casos donde la interferencia lingüística dificultó la identificación de colocaciones. Por ejemplo, la combinación *llorar amargamente* fue frecuentemente confundida con *llorar tristemente*, debido a la existencia de expresiones similares en las L1 de los participantes.

4.1.3.2 Síntesis de los hallazgos del conocimiento de las colocaciones léxicas complejas.

La categoría de colocaciones léxicas complejas incluye cuatro tipos distintos de combinaciones colocacionales (K. Koike, 2001). Los resultados fueron analizados de la misma manera que las colocaciones descritas en el apartado anterior. En términos generales, los

resultados similares obtenidos en ambos grupos participantes fueron similares. Tanto el grupo de redes sociales como el de UABC identificaron correctamente los mismos dos tipos de colocaciones y de manera incorrecta los otros dos tipos.

El primer tipo de colocación léxica compleja que fue identificado correctamente por ambos grupos es el de *Verbo + locución nominal*. Ejemplos de este tipo incluyen *golpe de estado* y *estilo de vida*. Se constató que estas expresiones comparten una estructura similar y un significado conceptual cercano en las tres lenguas analizadas, español, inglés y portugués. Por ejemplo, en inglés, *coup d'état* y *lifestyle* son equivalentes directos a las expresiones en español. En portugués, la equivalencia es aún más directa pues su traducción es *golpe de Estado* y *estilo de vida*.

Esta similitud estructural y semántica sugiere que la transferencia lingüística ha facilitado el reconocimiento de estas colocaciones por parte de los participantes. Estos resultados respaldan la hipótesis de Cheikh-Khamis Cases (2013), quien plantea que el conocimiento previo del aprendiz influye en la comprensión de conceptos lingüísticos. Sin embargo, es importante considerar que otros factores, como la frecuencia de exposición a estas expresiones en los materiales de aprendizaje, también podrían haber influido en los resultados obtenidos.

El segundo tipo de colocación léxica compleja que fue identificado con mayor precisión por los participantes es *Adjetivo + locución adverbial*. Colocaciones como *loco de remate* y *claro como el agua* fueron reconocidas correctamente. Este reconocimiento puede deberse a varios factores: la experiencia lingüística, es decir que los participantes han adquirido estas colocaciones a través de su experiencia con la lengua española, ya sea a través de la lectura, la escucha o la interacción comunicativa (Higueras García, 2006)., o bien, las asociaciones

semánticas entre el español y la L1 de los participantes. Por ejemplo, en *loco de remate*, la palabra *remate* evoca una imagen de algo llevado al extremo, reforzando el significado de *loco*.

Es interesante observar que expresiones similares existen en otras lenguas. En inglés, por ejemplo, *crystal clear* (cristalino) es equivalente a *claro como el agua* y connota una gran claridad o transparencia. En portugués, se pueden encontrar expresiones similares, como *claro como a água* (claro como el agua). Esta similitud entre lenguas sugiere que las asociaciones semánticas entre estas colocaciones son universales en cierta medida, lo que facilita su adquisición y reconocimiento.

Con relación a los tipos de colocaciones que no fueron identificados de manera correcta, se encuentra el tipo *Locución verbal + sustantivo*. La primera colocación que los participantes debían identificar es *llevar a cabo un proyecto*. Tanto los participantes de redes sociales como los de UABC escogieron incorrectamente el verbo *ejecutar*, en lugar de *llevar a cabo*. Se teoriza que se puede tratar de un caso de selección semántica o de interferencia con el inglés *execute*. La segunda colocación que no lograron identificar correctamente es la de *dejar plantado a alguien*. En este caso se observa que hubo falta de conocimiento metafórico por parte de los participantes.

El segundo tipo de colocación léxica compleja que obtuvo mayoría de respuestas incorrectas es el de *Verbo + locución adverbial*. Para este tipo, los participantes no lograron reconocer correctamente las combinaciones *llorar a moco tendido* y *estar en ayunas*. En el primer caso, ambos grupos seleccionaron la combinación incorrecta *llorar a moco echado*. En el segundo caso ambos grupos seleccionaron la combinación incorrecta *quedarme en ayunas*. Para ambos casos, se teoriza que se trata de un caso de selección semántica errónea. Cabe mencionar

que todos los ejemplos de combinaciones fueron tomados a partir de los ejemplos propuestos por Koike (2001).

Los datos analizados en este apartado mostraron que los participantes lograron identificar con mayor precisión colocaciones léxicas complejas como *golpe de estado* y *loco de remate*, las cuales presentan estructuras y significados similares en español, inglés y portugués. Esto sugiere que la transferencia lingüística y la experiencia previa con la lengua española facilitan el reconocimiento de estas expresiones idiomáticas. Sin embargo, los participantes tuvieron dificultades para identificar correctamente colocaciones como *llevar a cabo un proyecto* y *llorar a moco tendido*. En estos casos, se observó una tendencia a seleccionar opciones semánticamente cercanas pero incorrectas, lo que indica una falta de dominio de las expresiones idiomáticas y sus connotaciones específicas. En resumen, los datos muestran que la comprensión de colocaciones léxicas complejas está influenciada por diversos factores, como la transferencia lingüística, la experiencia previa con la lengua, y la familiaridad con las connotaciones culturales y metafóricas de las expresiones. La identificación correcta de estas colocaciones requiere un conocimiento profundo del idioma y de las convenciones idiomáticas propias de cada lengua.

4.1.3.3 Síntesis de los hallazgos del conocimiento de las colocaciones gramaticales.

La categoría de colocaciones gramaticales cuenta con seis tipos distintos de combinaciones colocacionales (Travalia, 2006). Esta fue la categoría con el mayor grado de reconocimiento por parte de los sujetos, de acuerdo con la tabla 11. Tanto el grupo de redes sociales como el de UABC identificaron de manera correcta cuatro de los seis tipos de colocaciones gramaticales.

A continuación, se enlistan las colocaciones gramaticales que obtuvieron un mayor porcentaje de respuestas correctas en ambos grupos participantes. El primer tipo corresponde a *sustantivo + preposición (+ sustantivo / oración subordinada)*. Ejemplos de este tipo son *miedo de* y *amor por*; ambos fueron identificados de manera correcta por los dos grupos. En reconocimiento correcto podría atribuirse al conocimiento lingüístico y preposicional de los participantes. Lo mismo sucede con el resto de las colocaciones gramaticales que también fueron reconocidas correctamente. Los tipos son: (*artículo +*) *sustantivo + preposición (+ infinitivo)* con las combinaciones *el derecho a* y *el placer de*, (*verbo + adjetivo + preposición* con las combinaciones *estar seguro de* y *estar convencido de*, y (*verbo + sustantivo + preposición + sustantivo* con las combinaciones *anunciar algo por los altavoces* y *ver algo en la televisión*.

Por otro lado, algunas colocaciones gramaticales registraron un mayor número de respuestas incorrectas. Una de ellas fue la colocación *preposición (+ artículo) + sustantivo (+ preposición)*. Un ejemplo de este tipo es la colocación *de por vida*. En este caso, ambos grupos seleccionaron erróneamente la combinación *por la vida*. De manera similar, al solicitarles identificar la colocación *en comparación con*, los participantes eligieron incorrectamente la combinación *en comparación de*. Estos casos sugieren confusión en el uso de preposiciones u omisión, lo que podría atribuirse a interferencia lingüística entre la L1 y la lengua meta.

La interferencia lingüística puede manifestarse de diversas maneras. En particular, el empleo de preposiciones puede variar considerablemente entre el español y otras lenguas. Por ejemplo, en inglés, la expresión equivalente a *de por vida* es *for life*, donde la preposición *for* cumple una función similar a *por* en español. Sin embargo, en otras lenguas, la expresión correspondiente podría requerir una estructura completamente diferente.

El segundo tipo de colocación gramatical con reconocimiento incorrecto fue la de *verbo + preposición*. Si bien tuvo un grado de reconocimiento positivo en el grupo de UABC, en el grupo de redes sociales, la mayoría de los participantes respondió incorrectamente, de acuerdo con los datos presentados en la tabla 9, tabla 10 y tabla 11. Entre los ejemplos que no fueron identificados correctamente se encuentran *hablar de* y *pensar en*. Dado el uso de preposiciones puede variar entre las distintas lenguas, es probable que estos errores sean producto de la interferencia lingüística entre la L1 y la lengua meta. Lo que en español se expresa con una determinada preposición, en inglés o portugués puede requerir otra diferente o incluso una construcción completamente distinta. En inglés, por ejemplo, hablar *de* se traduce como *to talk about* y pensar *en* como *to think about*, mientras que en portugués se utilizan *falar sobre* y *pensar em*, respectivamente.

A partir de estos resultados, es posible observar que el comportamiento colocacional de los aprendices responde a patrones diferenciados según su grado de dominio y exposición al español. En síntesis, los resultados obtenidos muestran que el conocimiento colocacional de los aprendices se ve influido por distintos factores lingüísticos y de exposición. Entre ellos destacan la interferencia de la lengua materna, la frecuencia con la que se presentan las expresiones, la complejidad semántica y las características estructurales de las colocaciones. En general, las colocaciones más transparentes o con equivalentes directos en la L1 tienden a reconocerse con mayor facilidad, mientras que aquellas de tipo idiomático o con estructuras menos convencionales presentan mayores dificultades para los aprendientes. Estos resultados refuerzan la idea planteada por Cheikh-Khamis Cases (2013) respecto a la necesidad de que el aprendiz

identifique las asociaciones léxicas que suelen acompañar a una unidad léxica, ya que ello favorece un uso más adecuado y natural de las colocaciones en la lengua meta.

Tras sintetizar los hallazgos generales del cuestionario, se procede a un análisis estadístico complementario que permite contrastar las diferencias entre grupos y tipos de colocaciones, acompañado de ejemplos que ilustran los patrones más frecuentes.

4.1.4. Análisis estadístico y ejemplos ilustrativos

Con el fin de respaldar cuantitativamente las diferencias observadas entre grupos y tipos de colocaciones, se aplicó la prueba de chi-cuadrado (χ^2) para comparar la distribución de respuestas correctas e incorrectas entre el grupo de UABC y el grupo de redes sociales. Los resultados se resumen en la tabla 12.

Tabla 12

Resultados del análisis de chi-cuadrado sobre la producción de colocaciones por grupo de aprendices

Tipo de colocación	χ^2	p-valor
Sustantivo (Sujeto) + Verbo	0	1
Verbo + Adverbio (mente)	0	1
Verbo + Sustantivo (Objeto)	0	1
Adjetivo + Sustantivo	0	1
Sustantivo + Preposición + Sustantivo	0	1
Verbo participio + Adverbio	0	1
Verbo + Preposición + Sustantivo	0	1
Adverbio + Adjetivo	0	1
Sustantivo + Adjetivo	0	1

Verbo + Adjetivo	0	1
Verbo + locución nominal	0	1
Locución verbal + sustantivo	0	1
Verbo + locución adverbial	0	1
Adjetivo + locución adverbial	0	1
Verbo + preposición	0	1
Sustantivo + preposición (+ sustantivo / oración)	0	1
(artículo +) Sustantivo + preposición (+ infinitivo)	0	1

Nota: Se empleó la prueba de chi-cuadrado para determinar si existían diferencias significativas entre los grupos de aprendices (UABC y redes sociales) en la producción de colocaciones correctas e incorrectas. $p > .05$ indica que no se observaron diferencias estadísticamente significativas

Como puede observarse, en ninguno de los casos se alcanzaron diferencias estadísticamente significativas ($p > .05$). Esto indica que, si bien existen variaciones descriptivas en los porcentajes de aciertos y errores, el patrón de desempeño entre los dos grupos fue similar en términos estadísticos.

Como complemento a los resultados cuantitativos, algunos ejemplos representativos de colocaciones producidas correctamente por los participantes son: *correr un rumor* (UABC), *confirmar una sospecha* (UABC), *rechazar categóricamente* (Redes Sociales), *estar lleno de trabajo* (producción en corpus CACE). En contraste, algunos ejemplos incorrectos, ya sea por influencia de la L1 o que suenen poco naturales son: *hacer una decisión* (interferencia del inglés *make a decision*), *ejecutar el cargo* (poco natural, en lugar de *asumir el cargo*), *llorar tristemente* (calco de *cry sadly*, cuando en español se prefiere *llorar amargamente*), *negar sistemáticamente* (influencia del portugués, en español más común: *rechazar categóricamente*)

Estos ejemplos permiten visualizar con claridad los patrones observados en los datos, mostrando tanto la interferencia de las lenguas de origen como los aciertos en la producción de combinaciones propias del español nativo.

Los resultados del análisis de chi-cuadrado confirman que las diferencias observadas entre los grupos no alcanzan significación estadística ($p > .05$). No obstante, los ejemplos de colocaciones correctas e incorrectas permiten identificar patrones vinculados con el nivel de competencia y con posibles efectos de la lengua materna. En particular, se observa que las combinaciones más transparentes y con equivalentes directos en la L1 tienden a producirse con mayor frecuencia y corrección, mientras que las colocaciones menos literales o con restricciones preposicionales muestran mayor variabilidad. Este contraste entre la evidencia estadística y los ejemplos cualitativos complementa la interpretación general de los datos y refuerza la necesidad de abordar el componente colocacional mediante instrucción explícita.

Estos resultados concuerdan con lo señalado por Nesselhauf (2005) y Laufer y Waldman (2011), quienes destacan la interacción entre la frecuencia de exposición, la competencia lingüística y la interferencia de la L1 en la producción colocacional de aprendices de segundas lenguas..

4.2. El Corpus de Aprendices de Colocaciones de ELE (CACE)

Uno de los objetivos investigación fue el diseño y construcción de un corpus de aprendices de ELE, con el propósito de analizar la producción de colocaciones de dichos aprendices. A continuación, se presenta una descripción del Corpus de Aprendices de Colocaciones de ELE (CACE) y un análisis a partir de los datos obtenidos en el mismo.

El CACE está compuesto por 222 muestras de producción escrita realizadas por parte de aprendices de español de 33 lenguas nativas diferentes, cuya lista se detalla en la tabla 8. Así mismo, el corpus cuenta con un total de 12,257 tokens y 2,113 tipos de lemas. Los datos recopilados no fueron estructurados, ni procesados inicialmente; es decir, se incorporaron al corpus sin alterar la producción del participante.

El CACE se considera un corpus oportunista, ya que recoge únicamente los datos disponibles para efectos de esta investigación (McEnery y Hardie, 2012). Además, se trata de un corpus de aprendices, conformado por muestras reales de la lengua meta. Aunque se trata de un corpus relativamente pequeño (Ragan, 2001), las muestras que lo conforman permiten realizar tanto análisis cuantitativos sobre las colocaciones producidas como análisis cualitativos sobre los fenómenos lingüísticos observados. En este sentido, el CACE puede clasificarse como un corpus especializado (Beaugrande, 2001; Sierra, 2015).

Si bien, es posible utilizar el CACE para observar diversos fenómenos lingüísticos de los aprendices, su objetivo principal es ser utilizado para observar el fenómeno colocacional. De acuerdo con la tipología de corpus de Sierra (2015), descrita en el apartado 2.3., el CACE se clasifica como un corpus textual, especializado, de autoría variada, sincrónico, monolingüe, pequeño, oportunista y diseñado para propósitos específicos.

Es importante destacar que el CACE no representa una muestra exhaustiva del uso que los aprendices de ELE hacen del español, ni de todas las colocaciones existentes en la lengua. No obstante, proporciona una muestra del uso real que estos aprendices de ELE, con diversos antecedentes lingüísticos, hacen del español. A pesar de ser un corpus pequeño, el CACE

permite analizar cómo los aprendices de ELE verbalizan el fenómeno colocacional a través de muestras de lenguaje natural.

A continuación, se proporciona una descripción de los resultados sobre la producción de colocaciones obtenidas a través del CACE.

4.2.1. Resultados de la producción de colocaciones a partir del CACE

Las listas de palabras del CACE, generadas a partir de *Sketch Engine*, proporcionaron un total de 764 sustantivos, 361 verbos, 175 adjetivos, y 111 adverbios. A partir del análisis de estas unidades léxicas, se identificaron 277 coapariciones dentro del CACE. De éstas, 223 corresponden a colocaciones producidas por los participantes, que fueron empleadas al menos dos veces. Estas colocaciones representan 48 tipos distintos.

Por ejemplo, la colocación *querido amigo* que es una colocación del tipo *Sustantivo + Adjetivo*, aparece 19 veces dentro del CACE. De manera similar, la colocación *pedir disculpas*, del tipo *Verbo + Sustantivo (Objeto)*, se registra 13 veces. Si bien, se encontraron otras combinaciones que podrían clasificarse dentro de la taxonomía de colocaciones, aquellas que sólo tuvieron una aparición dentro del CACE no fueron consideradas para el análisis.

Entre las 48 colocaciones encontradas en el CACE, se evidenció la producción de colocaciones léxicas simples (Corpas Pastor, 1996; K. Koike, 2001) y de colocaciones gramaticales (Travalia, 2006). No se encontró evidencia de producción de colocaciones léxicas complejas (K. Koike, 2001). A continuación, se presentan los resultados tanto de las colocaciones léxicas como gramaticales encontradas, así como un análisis y discusión al respecto.

4.2.1.1. La producción de colocaciones léxicas en el CACE

En este apartado se abordan las colocaciones léxicas identificadas en el CACE. En cuanto a las colocaciones léxicas simples (Corpas Pastor, 1996; K. Koike, 2001), se encontraron cinco tipos distintos de colocaciones dentro del corpus. El tipo más frecuente fue *Sustantivo + Adjetivo*, mientras que el menos frecuente fue *Verbo + Adverbio (mente)*.

Tabla 13

Colocaciones léxicas simples en el CACE

Tipo de colocación	Frecuencia absoluta de aparición en CACE	Ejemplo	
		Base	Colocativo
Sustantivo + Adjetivo	89	amigo	querido
Verbo + Sustantivo (Objeto)	17	disculpa	pedir
Verbo + Adjetivo	15	estar	enfermo
Sustantivo + Preposición + Sustantivo	13	semana	fin
Verbo + Adverbio (mente)	2	llegar	finalmente

Nota: Tipos de colocaciones léxicas simples en el CACE de acuerdo a la taxonomía de Corpas Pastor Corpas Pastor (1996) y Koike (2001)

En la tabla 13 se incluyen ejemplos representativos de colocaciones léxicas simples identificadas en el CACE. Aunque la taxonomía general contempla diez tipos de colocaciones léxicas simples, como se indica en la tabla 1, solo seis de estos tipos estuvieron presentes en los datos analizados. Entre ellos, destaca el tipo *Sustantivo + Adjetivo*, que fue el más frecuente dentro del corpus, en concordancia con lo señalado por Koike (2001). El resto de las colocaciones léxicas simples que no figuran en la tabla 13, no fueron empleadas dentro del CACE, o bien, que no figuraron al menos dos veces dentro del CACE, criterio utilizado para su inclusión en el análisis cuantitativo.

Como se mencionó, la colocación con mayor frecuencia fue *Sustantivo + Adjetivo*. Con combinaciones como *año pasado* y *primer lugar*, con un índice de Información Mutua (IM) de ≥ 11 . Algunos otros ejemplos de la producción de este tipo de colocación *Sustantivo + adjetivo* son *querido amigo* con un IM de 9.11, *correo electrónico* con IM de 11.24, y *mejor amigo* con un IM de 10.02. Se revisó el perfil de cada uno de los participantes que produjeron estas combinaciones, de acuerdo a la nomenclatura del CACE presentado en la tabla 5 y tabla 2, y se observó que fue producida por los participantes de todos los grupos, géneros, con variados niveles de dominio del español, y diversas L1. Es decir, esta colocación fue producida por participantes de todos los perfiles, los cuales se pueden consultar en el apartado 3.4. y en la tabla 2.

El segundo tipo de colocación léxica simple identificada en el CACE es la de *Verbo + Sustantivo (Objeto)*. Entre los ejemplos empleados en el CACE se encontraron las colocaciones *pedir disculpas*, y *pedir perdón*. La colocación *pedir disculpas* fue la más frecuente, ya que la pauta de escritura instruía a los participantes que solicitaran una disculpa.

La siguiente categoría de colocaciones léxicas simples dentro del CACE es aquella de *Verbo + Adjetivo*. El ejemplo más frecuentemente utilizado fue *estar enfermo*, misma que se relaciona con la pauta proporcionada a los participantes. Otros ejemplos de esta colocación fueron *estar presente* y *estar mejor*. Las combinaciones como *estar presente* y *estar mejor* se consideran colocaciones debido a sus características semánticas, ya que pueden emplearse como definidores de otros términos, en particular *presentarse* y *mejorar* (K. Koike, 2001). De acuerdo con el perfil de los participantes que produjeron este tipo de colocación, se trata de hablantes de inglés y portugués como L1, y pertenecen a todos los grupos, niveles y géneros.

El tipo de colocación *Sustantivo + Preposición + Sustantivo* también se encontró en el CACE con las combinaciones *dolor de cabeza*, *periodo de exámenes*, y *fin de semana*. Estas combinaciones obtuvieron un IM de 11.5 para el primer ejemplo y 10.02 para los otros dos. Este caso tipo de colocación fue producido por participantes cuya L1 es inglés y portugués, con variaciones en niveles, géneros, y grupos.

En contraste, la colocación *Verbo + Adverbio (mente)* fue producida únicamente por participantes con nivel sobresaliente de dominio del español, de acuerdo con la tabla 7, esta colocación tuvo una frecuencia absoluta de 2 (dos) dentro del CACE, con la combinación *llegar finalmente*. Esta colocación obtuvo un IM de 11.24. Los dos participantes que produjeron esta colocación pertenecen al grupo de redes sociales y declararon que su L1 es inglés.

En lo que respecta a las colocaciones léxicas complejas (Koike, 2001), no se encontró suficiente evidencia de su empleo dentro del CACE. La ausencia de producción de estas colocaciones dentro del CACE se puede deber a varias causas. En primer lugar, la pauta de escritura, que limita la producción escrita de los aprendices a la temática de una pauta proporcionada. En segundo lugar, la naturaleza de metafórica de las locuciones que forman parte de las colocaciones complejas (K. Koike, 2001). De acuerdo con Cheikh-Khamis Cases (2013), los aprendices requieren apoyo para desarrollar su competencia metafórica. Si no existen imágenes conceptuales compartidas entre la L1 y la L2, es posible que los aprendices tengan dificultades para producir este tipo de combinaciones, especialmente si no coinciden con aquellas de su L1.

Por ejemplo, la colocación *dejar plantado a alguien* (*locución verbal + sustantivo*) no tiene un significado transparente, ya que la locución *dejar plantado* no se deduce literalmente de

sus componentes. Esto sugiere que los aprendices podrían enfrentar problemas relacionados con el conocimiento metafórico de la lengua meta, el cual no es objeto de estudio de la presente investigación. No obstante, Cheikh-Khamis Cases propone que una instrucción específica puede contribuir al desarrollo de la competencia metafórica de los aprendices.

Sin embargo, dentro del CACE se identificó evidencia de conocimiento y uso de ciertas locuciones por parte de los sujetos participantes. Entre las locuciones identificadas, aunque no se consideran en este estudio por no formar parte de su objetivo, se encuentran las expresiones como *lo siento*, *por favor*, y *hasta luego*.

En resumen, en cuanto a las colocaciones registradas en el CACE, se observó que algunas combinaciones aparecieron con notable frecuencia, como es el caso de las colocaciones léxicas simples del tipo *Sustantivo + Adjetivo* y las colocaciones gramaticales *Verbo + Preposición*. No obstante, ciertos tipos mostraron una presencia escasa o nula, como ocurrió con las colocaciones léxicas complejas, que no fueron producidas de manera suficiente para ser incluidas en el análisis. Una posible explicación es que los aprendices prefieren utilizar expresiones que ya conocen y dominan, evitando aquellas cuya estructura les resulta menos familiar. Esta tendencia a priorizar lo conocido sobre lo nuevo ha sido documentada en investigaciones sobre adquisición de vocabulario, donde se señala que los estudiantes suelen apoyarse en fórmulas prefabricadas como estrategia para reducir la carga cognitiva durante la producción lingüística (Rica Peromingo, 2010; Schmitt y Schmitt, 2020).

4.2.1.2. La producción colocaciones gramaticales en el CACE

La categoría de colocaciones gramaticales establecida por Travalia (2006) también se encuentra representada dentro del CACE. Se identificaron dos tipos de colocaciones gramaticales

dentro del corpus de aprendices. El tipo más frecuente es aquel de *verbo + preposición*, con 82 ocurrencias. Algunos ejemplos de esta colocación en el CACE son *asistir a* y *perdonar por*.

El segundo tipo corresponde a *verbo + adjetivo + preposición*, el cual aparece en cinco ocurrencias. La tabla 14 muestra las colocaciones gramaticales identificadas dentro del CACE y algunos ejemplos de estas combinaciones.

Tabla 14

Colocaciones gramaticales en el CACE

Tipo de colocación	Frecuencia de aparición en CACE	Ejemplo	
		Base	Colocativo
verbo + preposición	82	asistir	a
(verbo) + adjetivo + preposición	5	lleno de	estar

Fuente: Travalía (2006).

Como se menciona anteriormente, el tipo de colocación gramatical más frecuentemente identificado en el CACE es el de *verbo + preposición*. Otros ejemplos de producción de esta colocación son *pensar en*, *creer que*, e *invitar a*. El puntaje más alto de IM fue de 6.23 para la colocación *asistir a*. 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 de dominio de la lengua, y grupo.

Adicionalmente, en materia de la colocación gramatical *(verbo) + adjetivo + preposición*, ésta aparece dentro del CACE con los ejemplos *estar lleno de* y *estar presente en*, con un IM de 5.38 y de 9.65 respectivamente. Al analizar el perfil de los participantes que produjeron que estas colocaciones, se identificó que fue producida por sujetos con un nivel de dominio de la lengua básico e intermedio. El resto de los datos estadísticos, tales como edad, género y L1 de los participantes es variado.

Las colocaciones gramaticales que no se produjeron dentro del CACE o que no aparecieron al menos dos veces son: *sustantivo + preposición (+ sustantivo / oración subordinada)*, *(artículo +) sustantivo + preposición (+ infinitivo)*, *preposición (+ artículo) + sustantivo (+ preposición)*, *(verbo + sustantivo) + preposición + sustantivo*. A continuación, se presenta un análisis de los tipos de colocaciones gramaticales que sí fueron producidas en el CACE.

4.2.2. Análisis y discusión de la producción de colocaciones léxicas en el CACE

Los resultados obtenidos a partir del análisis del CACE revelan que los aprendices de ELE produjeron principalmente colocaciones léxicas simples, de las cuales destacan combinaciones del tipo *sustantivo + adjetivo*, como *querido amigo*, y *verbo + sustantivo (objeto)*, como *pedir disculpas*. Estas colocaciones, además de ser frecuentemente utilizadas, son composicionales y semánticamente transparentes, lo que concuerda con la descripción de Koike (2001), quien señala que las colocaciones simples se caracterizan por su flexibilidad morfológica y su relativa transparencia semántica. La frecuencia de uso de este tipo de colocaciones dentro del CACE sugiere que los aprendices se apoyan en estructuras prototípicas y de uso común, posiblemente adquiridas por exposición o a través de transferencias positivas desde su lengua materna.

En este sentido, la alta frecuencia de colocaciones como *querido amigo* podría explicarse por una similitud formal y semántica con expresiones equivalentes en otras lenguas, lo cual, como apunta Coseriu (1981), no constituye en sí mismo una prueba de que se haya interiorizado una colocación en la lengua meta, sino que puede tratarse de una coincidencia interlingüística o

de una elección basada en el conocimiento extralingüístico del mundo. Esta transferencia puede facilitar la producción de colocaciones aceptables, aunque no necesariamente implique un dominio profundo de las restricciones de combinación propias del español.

Por otra parte, el análisis también mostró una producción nula de colocaciones léxicas complejas dentro del corpus. Estas colocaciones, en tanto unidades lexicalizadas parcialmente, requieren no solo un conocimiento formal, sino también una comprensión de su carga semántica y contextual. Koike (2001) destaca que las colocaciones complejas tienden a tener un mayor grado de fijación y menor variabilidad, por lo que su uso adecuado indica una mayor familiaridad con las normas pragmáticas de la lengua.

Sin embargo, la falta de evidencia de producción de estas colocaciones en el corpus sugiere que los aprendices aún no dominan completamente este tipo de combinaciones. Como señala Alonso Ramos (1994), las colocaciones no deben analizarse desde un punto de vista meramente sistemático, sino según la norma, es decir, de acuerdo con las preferencias de uso establecidas socialmente. Por tanto, la ausencia de colocaciones más idiomáticas podría deberse a una falta de instrucción explícita en torno a este tipo de unidades léxicas, o bien a una exposición limitada a contextos comunicativos auténticos en los que estas expresiones son comunes.

Adicionalmente, Corpas Pastor (1996) advierte que las colocaciones no se distinguen de combinaciones libres por su estructura, sino por las restricciones impuestas por el uso. El hecho de que los aprendices prefieran combinaciones más previsibles, o incluso generen combinaciones no idiomáticas, sugiere una comprensión incompleta de estas restricciones, especialmente en el

caso de colocaciones cuya aceptabilidad depende más del uso establecido que de la lógica composicional.

4.2.3. Análisis y discusión de la producción de colocaciones gramaticales en el CACE

Los hallazgos en torno a las colocaciones gramaticales del CACE permiten reflexionar sobre el grado de familiaridad de los aprendices con estructuras léxicas parcialmente fijas que involucran elementos gramaticales. Tal como lo señala Travalia (2006), este tipo de colocaciones no ha recibido suficiente atención en la investigación del ELE, a pesar de que constituyen un componente esencial del conocimiento fraseológico de la lengua. El hecho de que solo se hayan producido dos de los seis tipos contemplados en la clasificación de dicha autora parece confirmar esta falta de sistematización en su enseñanza y práctica.

La frecuencia relativamente alta de la combinación *verbo + preposición* indica que los aprendices tienden a consolidar primero aquellas colocaciones que les resultan más accesibles tanto por su exposición en el aula como por su uso en contextos formales y cotidianos. En este sentido, cabe recordar que muchas de estas combinaciones también se encuentran en otras lenguas, lo cual puede facilitar su adquisición por analogía o transferencia positiva. No obstante, como advierte Koike (2001), la existencia de una estructura formal similar no garantiza que se produzca un uso correcto, ya que las restricciones léxicas y combinatorias pueden variar significativamente entre lenguas.

En contraste, la presencia limitada del patrón (*verbo*) + *adjetivo + preposición* y la ausencia del resto de las estructuras gramaticales propuestas por Travalia (2006) sugieren que los aprendices aún no han interiorizado plenamente las restricciones que rigen estas combinaciones.

Esta dificultad podría atribuirse, al menos en parte, al hecho de que tales colocaciones suelen pasar desapercibidas tanto en los materiales didácticos como en la instrucción directa. Como lo plantea Cheikh-Khamis Cases (2013b), sin una enseñanza explícita de estas construcciones, los estudiantes no desarrollan la conciencia necesaria para identificar y reproducir patrones propios de la lengua meta.

Además, la escasa producción de colocaciones gramaticales más complejas podría estar relacionada con el nivel de dominio general de los aprendices. A medida que aumenta la competencia lingüística, se espera una mayor capacidad de identificar y producir las restricciones combinatorias impuestas por la norma, especialmente en estructuras donde intervienen elementos gramaticales y léxicos en simultáneo (Corpas Pastor, 1996).

En síntesis, aunque el CACE proporciona evidencia del uso productivo de algunas colocaciones gramaticales, los datos también reflejan limitaciones importantes en la variedad y complejidad de las estructuras utilizadas. Esto refuerza la necesidad de incorporar un enfoque más sistemático sobre este tipo de combinaciones en la enseñanza del ELE, de modo que los aprendices puedan ampliar su repertorio colocacional y alcanzar una mayor naturalidad en el uso del español.

4.3. Análisis y discusión del conocimiento colocacional vs. la producción colocacional

Con el objetivo de obtener una visión más amplia del manejo del fenómeno colocacional de los aprendices de ELE, se realizó un análisis contrastivo entre el conocimiento colocacional de los participantes y su producción escrita. Para ello, se analizaron los 20 tipos de colocaciones identificados por Corpas Pastor (1996), Koike (2001) y Travalía (2006) presentados en la tabla 1.

Este análisis, que se describe en el siguiente apartado, se divide en las tres categorías de colocaciones propuesto por las autoras previamente mencionadas.

4.3.1. El conocimiento colocacional vs. producción de las colocaciones léxicas simples

La categoría de colocaciones léxicas simples, identificadas por Corpas Pastor (1996) y Koike (2001) consta de diez tipos de colocaciones distintos. Por cada uno de ellos, se contrastaron los resultados sobre el conocimiento colocacional con la producción de los participantes en el CACE para, más adelante, hacer observaciones al respecto.

El primer tipo de colocación léxica simple es el de *Sustantivo (Sujeto) + Verbo*, el cual obtuvo una mayoría de respuestas correctas en el cuestionario. Los ejemplos incluidos en dicha evaluación fueron *correr un rumor* y *confirmar una sospecha*, expresiones que, como se discutió anteriormente, presentan una relativa transparencia semántica y una posible transferencia positiva desde las lenguas maternas de los participantes. Sin embargo, no se encontró evidencia de la producción de este tipo de colocación en el corpus CACE. Este hallazgo sugiere una discrepancia entre el reconocimiento receptivo y la producción activa de este patrón colocacional. Una posible explicación para esta ausencia radica en las características discursivas y estructurales de la tarea de escritura propuesta. Tal como señala Koike (2001), el uso de determinadas colocaciones está estrechamente vinculado con el contexto comunicativo y con la función discursiva que estas cumplen dentro del texto. En este sentido, la pauta de redacción semidirigida pudo haber limitado el uso de estructuras oracionales en las que un sustantivo funcione como sujeto explícito seguido de un verbo que active el patrón colocacional. De acuerdo con Corpas Pastor (1996), la frecuencia y distribución de las colocaciones en un texto dependen no solo del conocimiento del hablante, sino también de las restricciones textuales y

pragmáticas que impone el género discursivo. Así, se teoriza que los participantes, aunque capaces de reconocer este tipo de colocaciones en un contexto evaluativo, no encontraron oportunidades discursivas naturales para su producción escrita, o bien no supieron cómo integrarlas adecuadamente en un tipo de texto que posiblemente priorizaba estructuras informativas o argumentativas más impersonales.

En cuanto al tipo de colocación léxica simple *Verbo + Adverbio (-mente)*, los resultados del cuestionario reflejan un menor grado de familiaridad por parte de los participantes, como lo evidencia el alto porcentaje de respuestas incorrectas. No obstante, en el CACE se identificó la colocación *llegar finalmente*, empleada en dos ocasiones por estudiantes que declararon tener un nivel sobresaliente de español. Este dato sugiere que este tipo de colocación, al requerir un mayor dominio de matices semánticos y una mayor competencia discursiva, podría estar más estrechamente vinculado con niveles avanzados de competencia lingüística. Sin embargo, esta observación es de carácter exploratorio y podría confirmarse o refutarse mediante la comparación con otros corpus de aprendices de ELE que integren una muestra más amplia y diversa de producciones escritas. En relación con la ausencia de otros ejemplos de esta categoría en el corpus, es importante subrayar que la no producción de una colocación no puede asumirse de forma directa como indicio de desconocimiento. Como señalan Koike (2001) y Corpas Pastor (1996), la producción de colocaciones está condicionada no solo por el conocimiento lingüístico, sino también por factores contextuales y discursivos, como el género textual, la intencionalidad comunicativa y las restricciones temáticas impuestas por la tarea. En este caso, la pauta de escritura semidirigida pudo haber limitado las oportunidades para que los aprendices integraran

este tipo específico de combinaciones, especialmente si su uso no resultaba funcional o pertinente dentro del desarrollo temático del texto solicitado.

De manera similar, la colocación del tipo *Verbo + Sustantivo (Objeto)* presentó un alto porcentaje de respuestas incorrectas en el cuestionario, pero, paradójicamente, se halló evidencia significativa de su producción dentro del corpus CACE por parte de hablantes de inglés y portugués. En total, este tipo de colocación fue producido en 32 ocasiones, con ejemplos como *pedir disculpas* y *pedir perdón*. No obstante, según los datos resumidos en la tabla 11, fue identificada incorrectamente por el 60.7% de los participantes en el cuestionario. Esta discrepancia entre reconocimiento y producción sugiere que, si bien los aprendientes pueden emplear estas combinaciones de forma intuitiva o basada en fórmulas memorizadas, no siempre logran identificarlas conscientemente como colocaciones en tareas de tipo evaluativo, como lo fue el cuestionario. Como señalan Jiménez Calderón y Rufat (2019), este fenómeno podría deberse a una limitada conciencia intralingüística, es decir, una escasa reflexión metalingüística sobre el comportamiento combinatorio de las palabras. Esta falta de conciencia dificulta el reconocimiento explícito de estructuras que, aunque son utilizadas correctamente en contextos comunicativos reales, no son percibidas como combinaciones comunes dentro de pruebas de tipo formal o académico, lo cual refuerza la importancia de una instrucción explícita del componente colocacional en el aula, como lo proponen Schmitt y Schmitt (2020).

La colocación del tipo *Adjetivo + Sustantivo* fue identificada por Koike (2001), mientras que Corpas Pastor (1996) clasifica la misma estructura como *Sustantivo + Adjetivo*. En el cuestionario se incluyeron ejemplos representativos propuestos por ambas autoras, lo que permitió observar el comportamiento de los aprendices frente a diferentes manifestaciones de

este patrón colocacional. En primer lugar, al trabajar con los ejemplos clasificados por Koike (2001) bajo la categoría *Adjetivo + Sustantivo*, se observó un porcentaje elevado de respuestas incorrectas en el cuestionario. No obstante, se encontró evidencia de la producción de esta estructura en el corpus CACE, específicamente con expresiones como *primer lugar*. Asimismo, se registró la colocación *año pasado*, que corresponde a la estructura *Sustantivo + Adjetivo* descrita por Corpas Pastor (1996). Tras revisar los datos, se confirmó que los participantes que produjeron estas colocaciones también seleccionaron correctamente este tipo de combinación en el cuestionario, lo cual indica una correspondencia positiva entre el reconocimiento explícito y la capacidad de producción escrita en estos casos.

Por otro lado, en cuanto a la estructura *Sustantivo + Adjetivo*, identificada por Corpas Pastor (1996), el cuestionario arrojó un nivel de reconocimiento relativamente favorable, con un 54 % de respuestas correctas, como se muestra en la tabla 11, en ejemplos como *amor ciego* y *conducta intachable*. Además, este tipo de colocación fue la más frecuente dentro del corpus CACE, de acuerdo con los datos presentados en la tabla 13. Se registraron múltiples ejemplos, entre ellos *año pasado*, *próxima vez*, *mejor amigo* y *hermano menor*. Al analizar el perfil de los participantes, se constató que esta estructura fue producida por aprendices de todos los niveles de competencia y contextos educativos, lo que sugiere que se trata de una combinación ampliamente internalizada y común en el uso escrito del español por parte de aprendices de ELE.

La discrepancia observada entre los resultados de ambas subcategorías puede explicarse por varios factores. En primer lugar, la frecuencia de uso en el input recibido por los aprendices juega un papel determinante. Como señalan Alonso Ramos (Alonso Ramos, 1994) y Hoey

(2012), la aceptabilidad de una colocación está estrechamente vinculada a su grado de lexicalización y a la frecuencia con la que los aprendices la encuentran en el discurso. Desde la perspectiva del *lexical priming*, propuesta por Hoey, las palabras tienden a activarse en combinaciones que reflejan los patrones de uso a los que el hablante ha estado expuesto, lo cual favorece la consolidación de ciertas colocaciones sobre otras. Las colocaciones tipo *Sustantivo + Adjetivo* tienden a ser más comunes en contextos descriptivos o narrativos, y son más fácilmente reconocibles como unidades cohesivas. En contraste, las combinaciones *Adjetivo + Sustantivo*, como *primer lugar*, aunque frecuentes en ciertos géneros textuales, pueden percibirse como menos prototípicas o más sujetas a variación estilística, lo cual afecta su reconocimiento en tareas evaluativas (Koike, 2001). Asimismo, la transparencia semántica y la direccionalidad de la colocación, aspectos señalados por Coseriu (Coseriu, 1981) y Cruse (1986), también influyen. En *amor ciego*, el sustantivo es la base semántica y el adjetivo modifica su significado dentro de un patrón conocido, mientras que, en *primer lugar*, el adjetivo tiene una función más clasificatoria o estructural, menos marcada semánticamente, lo cual puede dificultar su categorización como colocación para los aprendices. En conjunto, estos resultados refuerzan la importancia de considerar no solo la estructura formal de las colocaciones (Corpas Pastor, 1996; Koike, 2001), sino también su funcionalidad discursiva (Koike, 2001), su frecuencia en el input (Alonso Ramos, 1994; Hoey, 2005; Schmitt y Schmitt, 2020) y su grado de fijación semántica (Cruse, 1986; Coseriu, 1981) como variables que afectan tanto su reconocimiento como su producción en el aprendizaje de ELE.

En el caso de la colocación de tipo *Sustantivo + Preposición + Sustantivo*, los resultados muestran la misma tendencia: un alto porcentaje de respuestas incorrectas dentro del

cuestionario, pero evidencia de producción en el corpus de aprendices. Los ejemplos identificados en el CACE fueron *dolor de cabeza*, *período de exámenes* y *fin de semana*. En estos casos, los participantes que produjeron estas colocaciones tenían como L1 el inglés o el portugués, lo que sugiere una posible transferencia lingüística, ya que estas construcciones son comunes en ambas lenguas. Por ejemplo, *dolor de cabeza* se dice *dor de cabeça* en portugués. La estructura *Sustantivo + Preposición + Sustantivo* se mantiene entre el español y el portugués. Sin embargo, también se identificaron errores como *dolor de la cabeza* (en lugar de *dolor de cabeza*). De acuerdo con Álvarez y Gil Martínez (2018), estos errores podrían atribuirse a una sobre generalización de las reglas del español o interferencias de la L1.

Las colocaciones del tipo *Verbo participio + Adverbio*, *Verbo + Preposición + Sustantivo* y *Adverbio + Adjetivo* mostraron un alto porcentaje de respuestas correctas en el cuestionario, lo cual refleja un nivel de reconocimiento léxico relativamente sólido entre los participantes. No obstante, no se registró evidencia de su uso en los textos producidos para el corpus CACE. A diferencia de otras colocaciones léxicas simples, como las de tipo *Sustantivo + Preposición + Sustantivo*, que sí fueron producidas por los aprendices, estas estructuras estuvieron ausentes en la producción escrita. Una posible explicación para este fenómeno se encuentra en la naturaleza de la tarea solicitada. Dado que la pauta de escritura consistía en redactar una carta de disculpa, es probable que el tipo de discurso exigido no favoreciera el uso frecuente de este tipo de colocaciones, las cuales tienden a aparecer en registros más descriptivos o narrativos. Como señalan Koike (2001) y Corpas Pastor (1996), la aparición de una colocación depende no solo del conocimiento lingüístico del hablante, sino también del contexto discursivo en el que se utiliza, ya que ciertas estructuras se activan o inhiben en función de la intención

comunicativa. Desde esta perspectiva, las colocaciones analizadas pueden no haber sido percibidas como necesarias o pertinentes dentro del desarrollo de la carta, lo cual concuerda con lo que plantea Alonso Ramos (1994) sobre la selección léxica guiada por la aceptabilidad pragmática y el grado de lexicalización. Además, aunque los participantes lograron identificar estas colocaciones en el cuestionario, su ausencia en los textos puede sugerir un conocimiento aún no plenamente integrado o automatizado. Schmitt y Schmitt (2020) diferencian claramente entre el conocimiento receptivo y el productivo de las colocaciones, señalando que este último requiere un mayor grado de consolidación y práctica. En esta línea, también Coseriu (1981) y Firth (1957) sostienen que la competencia colocacional no se limita a reconocer combinaciones plausibles, sino que implica saber activarlas en contextos reales de uso, en función de convenciones idiomáticas y restricciones de coocurrencia que muchas veces escapan a la conciencia explícita del aprendiz.

Por último, dentro de la categoría de colocaciones léxicas simples, tenemos la colocación de *Verbo + Adjetivo*. Esta obtuvo un porcentaje alto de respuestas incorrectas en el cuestionario y no se encontró evidencia de su producción dentro del CACE. Según Koike (2001), los verbos de este tipo de combinaciones carecen de contenido semántico, como en las combinaciones *resultar ileso* y *quedar claro*, lo que podría dificultar la identificación del colocativo correspondiente.

4.3.2. El conocimiento colocacional vs. producción de las colocaciones léxicas complejas

En cuanto a las colocaciones léxicas complejas, los resultados del cuestionario reflejaron un nivel intermedio de reconocimiento, con un promedio del 50% de respuestas correctas. No obstante, al analizar la producción escrita en el corpus CACE, no se encontró evidencia de uso de

ninguno de los tipos pertenecientes a esta categoría, lo que contrasta con el desempeño observado en otras categorías colocacionales. Esta discrepancia entre el conocimiento receptivo y la producción activa puede interpretarse a partir de dos niveles: el primero, las condiciones específicas de la tarea de escritura, y el segundo, factores más generales vinculados del conocimiento colocacional.

En cuanto al primer nivel, se reconoce que la pauta de redactar una carta de disculpa limitaba las oportunidades naturales para emplear ciertas colocaciones léxicas complejas, especialmente aquellas con estructuras más idiomáticas, como *llevar a cabo un proyecto*, *estar en ayunas*, o *claro como el agua*. Esta ausencia puede atribuirse a que los participantes no encontraron contextos propicios para utilizar dichas combinaciones dentro del género solicitado, caracterizado por un registro directo, claro y socialmente convencional, y que tiende apoyarse en fórmulas lexicalizadas de carácter más sencillo y explícito (Koike, 2001; Corpas Pastor, 1996).

Por otro lado, incluso en contextos comunicativos más amplios, este tipo de colocaciones representan un reto para los aprendices, debido a su opacidad semántica, baja frecuencia relativa en el input, y carácter formulaico o parcialmente idiomático (Schmitt y Schmitt, 2020; Coseriu, 1981). Como señala Alonso Ramos (1994), estas estructuras no se generalizan fácilmente, ya que dependen de convenciones específicas de uso y, en muchos casos, de una fuerte carga pragmática o metafórica que no siempre está disponible para el aprendiz de una L2. En consecuencia, su adquisición requiere no solo reconocimiento pasivo, sino una exposición prolongada a modelos auténticos de lengua y una instrucción explícita orientada al desarrollo de la competencia fraseológica (Jiménez Calderón y Rufat, 2019; Conzett, 2000).

Finalmente, el hecho de que los participantes identificaran correctamente algunas de estas colocaciones en el cuestionario, pero no lograran incorporarlas en sus producciones escritas, refuerza la distinción entre conocimiento receptivo y productivo señalado por Schmitt y Schmitt (2020). La producción de colocaciones complejas implica un mayor grado de automatización, así como un conocimiento más profundo de los patrones idiomáticos y discursivos de la lengua meta. Por tanto, su ausencia en el CACE puede entenderse como resultado tanto de las restricciones temáticas de la tarea, como de una competencia colocacional aún en desarrollo, particularmente en lo que respecta al uso funcional de estructuras complejas en situaciones comunicativas reales.

4.3.3. El conocimiento colocacional vs. producción de las colocaciones gramaticales

Como se ha establecido previamente, las colocaciones gramaticales, identificadas por Travalia (2006), se dividen en seis tipos. Para cada uno de estos, se contrastaron los resultados del conocimiento colocacional y la producción de colocaciones de los participantes para, entonces, hacer observaciones al respecto.

El primer tipo de colocación gramatical es el de *Verbo + Preposición*. Esta categoría presentó un porcentaje alto de respuestas incorrectas en el cuestionario. Sin embargo, tuvo la segunda mayor frecuencia de producción dentro del CACE. Algunos ejemplos de este tipo de colocación dentro del corpus de aprendices son *asistir a*, con 22 ocurrencias, y *perdonar por*, con 12. Asimismo, se registró la combinación *creer que*, también con 12 apariciones. Aunque esta secuencia no corresponde estrictamente a una colocación de tipo *verbo + preposición*, se incluye aquí debido a su alta frecuencia y a su valor representativo dentro del repertorio de combinaciones gramaticales producidas por los aprendices. Se observa que, si bien los

participantes demuestran la capacidad de producir de manera escrita combinaciones de este tipo, las opciones de respuesta proporcionadas en el cuestionario resultaron difíciles de distinguir para ellos.

Vincze et al. (2016) sugieren que es posible que los aprendices estén haciendo uso frecuente de un limitado repertorio de colocaciones que utilizan. Es decir, esta tendencia a repetir un número limitado de colocaciones podría estar influenciada por las L1 de los aprendices. Las colocaciones de *verbo + preposición* pueden variar significativamente entre diferentes idiomas. Por ejemplo, en español, el verbo *asistir* se combina con la preposición *a*, mientras que en inglés se usa *to attend*. Estas diferencias idiomáticas pueden dificultar la adquisición de las colocaciones correctas en español para hablantes de otras lenguas. Además, la frecuencia de uso de ciertas colocaciones en la L1 puede influir en la elección de los aprendices. Si cierta colocación *verbo + preposición* es muy común en la lengua materna, los aprendices podrían transferirla al español, incluso si no es la opción correcta. Este fenómeno de transferencia lingüística podría explicar por qué algunos participantes eligieron opciones incorrectas en el cuestionario o utilizaron un repertorio limitado de colocaciones en sus producciones escritas.

A continuación, las colocaciones gramaticales de tipo *sustantivo + preposición (+ sustantivo / oración subordinada)* y *(artículo +) sustantivo + preposición (+ infinitivo)*, obtuvieron resultados similares. Es decir, no se encontró evidencia de producción en el CACE para estos dos tipos de colocaciones. Por el contrario, en el cuestionario se obtuvo un porcentaje alto de reconocimiento correcto. Con base en ello, se observa que los sujetos participantes cuentan con la competencia para reconocer esta colocación. Sin embargo, en su producción escrita se teoriza que los participantes no emplearon estas colocaciones debido a la naturaleza del

tipo de escrito que se les indicó producir. Algunos ejemplos de estas colocaciones son *miedo de*, *amor por*, *confianza en*, y el *placer de*. Si bien estas combinaciones podrían haberse utilizado en un contexto de disculpa o reflexión personal, su ausencia en los textos sugiere que el tipo de tarea o el registro discursivo pedido no favoreció su aparición.

En cuanto al tipo de colocación *preposición (+ artículo) + sustantivo (+ preposición)*, se trata de una colocación gramatical que obtuvo mayoría de respuestas incorrectas en el cuestionario y tampoco se encontró evidencia de su producción dentro del CACE. Algunos ejemplos de esta colocación son *en comparación con*, *de camino a*, y *de por vida*. Esta última colocación fue identificada de manera errónea dentro del cuestionario por la mayoría de los participantes, quienes seleccionaron incorrectamente la combinación *por la vida*. Este error en particular nos da evidencia de la dificultad que los aprendientes tienen para identificar los colocativos correctos y distinguir entre combinaciones similares. En relación a esto, Vincze et al. (2016) llegaron a la hipótesis de que a los aprendices se les dificulta identificar el colocativo correspondiente a cierta colocación. En cuanto a la producción escrita, si bien, dentro del CACE existe evidencia de que los sujetos participantes tienen conocimiento de la producción de colocaciones gramaticales, se teoriza que esta colocación en particular se dificulta para los aprendices de ELE. Esta dificultad podría deberse a varios factores. Por un lado, es posible que los aprendices no hayan tenido la oportunidad de encontrar contextos que requirieran el uso frecuente de este tipo de colocaciones en sus producciones escritas. Por otro lado, más allá de falta de conocimiento de la lengua, se teoriza que los aprendices simplemente no cuentan con las herramientas para identificar los lazos asociativos entre las palabras.

La última combinación de las colocaciones gramaticales es la de (*verbo*) + *adjetivo* + *preposición*. Dentro del cuestionario, este tipo de colocación obtuvo un nivel de reconocimiento positivo del 56.8%, con las combinaciones *estar seguro de* y *estar convencido de* (para mayores detalles, consultar la tabla 13. Además, también se encontró evidencia de su producción dentro del CACE con los ejemplos *estar lleno de* y *estar presente en*. Esta colocación fue producida, dentro del CACE, por sujetos participantes que declararon tener un nivel básico e intermedio de dominio del español. Con base en estos datos, se infiere que se trata de estructuras fáciles de reconocer para los aprendices desde los primeros niveles.

En términos generales, las colocaciones gramaticales fueron la categoría que mostró mayor reconocimiento, destacando combinaciones como *sustantivo* + *preposición* y *verbo* + *preposición*. Según Benson et al. (1986), su predictibilidad las hace más fáciles de adquirir. Cuatro de los seis tipos de colocaciones gramaticales fueron reconocidos correctamente por la mayoría de los participantes, destacando combinaciones como *estar seguro de* y *miedo de*. Se observó que las colocaciones con estructuras similares en L1 de los participantes tienden a ser reconocidas correctamente. La facilidad de reconocimiento podría deberse a la transparencia de las estructuras y su similitud con la L1.

En este capítulo se abordó el análisis y discusión de los datos obtenidos a través del cuestionario y de las muestras de producción escrita que conforman el CACE. El análisis y discusión de los resultados del cuestionario se presentan por grupos participantes individuales. Es decir, se presentaron los resultados del grupo de redes sociales y los resultados del grupo de estudiantes de UABC por separado. Después, se presentó el concentrado de ambos grupos con el objetivo de hacer un análisis general. A continuación, se presentaron los resultados del CACE.

Se ofreció un análisis y discusión sobre las colocaciones identificadas en el corpus de aprendices, basándonos en la taxonomía de colocaciones. Por último, se ofrece un contraste entre los resultados generales del cuestionario y del CACE, así como observaciones y sugerencias al respecto de los fenómenos que los datos presentan.

En resumen, los resultados presentados y discutidos en este capítulo muestran diferencias importantes entre el reconocimiento y la producción de colocaciones por parte de los aprendientes de ELE. Estas diferencias varían según el tipo de colocación, su nivel de fijación y su función dentro del texto. Las colocaciones más frecuentes o semánticamente más transparentes tendieron a ser reconocidas y producidas con mayor facilidad, mientras que aquellas más complejas o con menor presencia en el input mostraron mayores dificultades, sobre todo en tareas de escritura semidirigida. También se evidenció que la naturaleza de la pauta de escritura influyó en el tipo de estructuras que los participantes decidieron utilizar. En el siguiente capítulo se retoman estos hallazgos para presentar las conclusiones generales del estudio, así como sugerencias para futuras investigaciones.

Capítulo V. Conclusiones

Esta investigación analizó el conocimiento colocacional de los aprendices de español como lengua extranjera (ELE), con base en un cuestionario y un corpus de producción escrita (CACE). Esto con el fin de contribuir en el proceso de la enseñanza aprendizaje de lenguas.

La metodología empleada permitió observar cómo los aprendices relacionan unidades léxicas en la lengua meta, y brindó experiencia empírica sobre la diferencia entre conocimiento receptivo y productivo de las colocaciones. A través del instrumento *Estudio de colocaciones*

ELE, se obtuvo información sobre el reconocimiento de colocaciones, mientras que mediante el CACE se obtuvo información sobre la producción de estas colocaciones. Además, el corpus resultante de la compilación de las producciones de los aprendices, aunque de tamaño reducido, constituye una herramienta relevante para el análisis de la competencia fraseológica en contextos reales.

Aunque de manera descriptiva se observaron tendencias que favorecían al grupo de UABC en ciertos tipos de colocaciones, las pruebas de chi-cuadrado no arrojaron diferencias estadísticamente significativas entre ambos grupos ($p > .05$). Esto implica que las aparentes diferencias deben interpretarse con cautela. No obstante, los ejemplos de colocaciones correctas e incorrectas recopilados permiten visualizar de manera concreta los procesos de transferencia e interferencia, lo que complementa y enriquece los hallazgos cuantitativos.

Entre los principales hallazgos, se encontraron los siguientes:

Sobre las colocaciones léxicas simples, los datos mostraron que el conocimiento colocacional de los aprendientes de *ELE* es parcial y varía según el tipo de combinación. se observó un mayor dominio en combinaciones frecuentes en el input, que presentan una mayor transparencia semántica tal es el caso de las colocaciones *adjetivo + sustantivo* (por ejemplo, *querido amigo*) y *verbo + preposición + sustantivo* (como *poner en práctica*). Además, estas colocaciones suelen tener equivalencias en la lengua materna de los aprendices, lo que sugiere un efecto positivo de la transferencia interlingüística, en línea con lo planteado por Cheikh-Khamis Cases (2013).

En contraste, las colocaciones *verbo + adjetivo* o *verbo + adverbio* presentaron mayores dificultades: obtuvieron bajos niveles de reconocimiento y no fueron producidas, lo cual se

asocia a su complejidad estructural, la menor frecuencia de uso y la opacidad semántica de estas combinaciones, como ya ha sido señalado por Koike (2001) y Corpas Pastor (1996).

Estos resultados confirman que la discrepancia entre el reconocimiento y la producción colocacional continúa siendo un reto en el aprendizaje del ELE. Aunque los aprendices puedan reconocer ciertas combinaciones en contextos controlados, no siempre logran activarlas en situaciones de producción espontánea. Esto sugiere que el conocimiento receptivo no es suficiente para garantizar un uso productivo y automatizado de las colocaciones, especialmente cuando estas requieren un nivel más avanzado de competencia discursiva.

Sobre las colocaciones léxicas complejas, estas representaron un mayor reto para los aprendices de ELE. A pesar de que algunas fueron reconocidas en el cuestionario, no se evidenció su producción en el CACE. Estas colocaciones, al incluir locuciones, muchas de ellas con carga idiomática o metafórica, presentan una mayor complejidad semántica y formal. Esta dificultad concuerda con lo señalado por Koike (2001). Esto sugiere que su adquisición requiere instrucción explícita, exposición intensiva y tareas que favorezcan su activación.

Sobre las colocaciones gramaticales, estas fueron las más empleadas en el corpus, especialmente, las del tipo *verbo + preposición*. lo que sugiere una mayor familiaridad con este patrón, probablemente debido a su frecuencia en el input y a su posible correspondencia con estructuras similares en la L1. Este hallazgo refuerza lo señalado por Travalía (2006), quien afirma que las colocaciones gramaticales, por su carácter más sistemático, pueden ser más accesibles para los aprendices en etapas iniciales o intermedias.

Así, se puede afirmar que el conocimiento colocacional de los aprendientes de español como lengua extranjera (ELE) es parcial, variable y condicionado por una serie de factores lingüísticos, contextuales y pedagógicos. La comprensión y producción de colocaciones en ELE, como se observó en esta investigación, no son uniformes ni automáticas, sino que responden a variables como la complejidad estructural de las combinaciones, su grado de transparencia semántica, su frecuencia en el input y la posibilidad de transferencia desde la lengua materna. Este hallazgo reafirma lo planteado en la literatura sobre la naturaleza multifactorial del aprendizaje de colocaciones, y destaca la importancia de abordarlas de forma integrada en el proceso de enseñanza-aprendizaje.

Con respecto a la clasificación y categorización de los fenómenos observados en la producción de colocaciones, los datos obtenidos permitieron identificar patrones lingüísticos que inciden directamente en el uso adecuado o inadecuado de estas combinaciones. En particular, se identificaron tres fenómenos clave: la interlengua, la transferencia y la interferencia lingüística.

Los fenómenos de interlengua, transferencia e interferencia lingüística desempeñaron un papel relevante en la producción y reconocimiento de colocaciones. En muchos casos, la transferencia positiva desde la lengua materna favoreció el uso adecuado de combinaciones frecuentes, especialmente aquellas con equivalentes directos en la L1, como las colocaciones del tipo *sustantivo + adjetivo* o *verbo + preposición + sustantivo*. Este tipo de combinaciones, al presentar una mayor transparencia semántica y frecuencia de uso, se integraron más fácilmente en la producción escrita de los aprendices. En cambio, se observaron errores recurrentes atribuibles a interferencias negativas, como la selección de colocativos poco naturales en

español, lo cual evidencia la influencia persistente de estructuras propias de la L1 en etapas intermedias del aprendizaje.

Por otro lado, los datos del corpus permitieron identificar ciertas manifestaciones de la interlengua, entendida como un sistema lingüístico en desarrollo (Koike, 2001). En la producción escrita de los aprendices se detectaron patrones que oscilan entre estructuras propias de la lengua meta y rasgos transferidos de la lengua materna. Estos patrones incluyen la omisión o sustitución de preposiciones esperadas, la sobre extensión de colocaciones comunes, y la creación de combinaciones no convencionales. Estos hallazgos confirman que el conocimiento colocacional no se adquiere de forma homogénea, sino que se construye de manera progresiva y variable, influido por factores cognitivos y contextuales. La categorización de estos fenómenos ofrece una base para futuros análisis contrastivos y propuestas didácticas que respondan a las necesidades reales de los aprendices.

Además, uno de los hallazgos más relevantes fue la brecha entre identificación de colocaciones en contextos controlados y su uso en la producción escrita. Este fenómeno sobre el conocimiento receptivo y productivo, ampliamente documentado en estudios previos (Schmitt y Schmitt, 2020; Koike, 2001), sugiere que los aprendices pueden reconocer ciertas combinaciones de manera receptiva sin haberlas integrado plenamente a su competencia productiva.

En términos metodológicos, los resultados de esta investigación muestran que la lingüística de corpus facilitó el analizar cómo los aprendices de ELE combinan palabras en su proceso de construcción léxica.

Cabe señalar, además, que el trabajo se desarrolló en el contexto del confinamiento por la pandemia de COVID-19, lo cual implicó un reto metodológico que se superó mediante el uso de herramientas digitales. El contacto con los participantes, la aplicación del cuestionario y la recolección de producciones escritas, se realizaron de manera remota. Esta experiencia evidenció el potencial de las tecnologías de la información para la investigación lingüística en tiempos de contingencia.

Desde una perspectiva pedagógica, se obtuvieron datos valiosos que reflejan la manera en que los aprendices relacionan palabras de forma significativa, lo que permite observar el desarrollo de fórmulas lingüísticas complejas y su inserción progresiva en el discurso escrito. Estos hallazgos no solo aportan al conocimiento teórico sobre la adquisición de colocaciones, sino que también ofrecen herramientas aplicables al aula para favorecer una producción más fluida y natural. De acuerdo con los hallazgos descritos anteriormente, se recomienda: Incluir colocaciones desde los niveles iniciales, con actividades que promuevan tanto el reconocimiento como la producción. Utilizar corpus auténticos que muestren el uso contextualizado de estas combinaciones. Diseñar tareas variadas (escritas y orales) que estimulen la producción de colocaciones menos frecuentes o más complejas. Contrastar estructuras de la L1 y la L2 para identificar interferencias y aprovechar transferencias positivas.

Finalmente, a partir de los hallazgos obtenidos, se proponen tres líneas principales para futuras investigaciones. Primeramente, la expansión del corpus CACE hacia un modelo sincrónico, que permitirá analizar el uso colocacional en diferentes niveles de competencia, contextos educativos y variedades del español a través del tiempo. La ampliación del corpus

podría facilitar estudios contrastivos, el análisis de errores, el seguimiento de la interlengua y la identificación de patrones de desarrollo en el conocimiento colocacional.

Por otro lado, se propone el diseño e implementación de propuestas didácticas centradas en el desarrollo de la competencia colocacional, basadas en corpus y enfocadas en tareas de producción escrita y oral. Estas propuestas podrían considerar el uso de tecnologías digitales, visualización de datos lingüísticos y metodologías de enseñanza explícita y reflexiva, con el fin de favorecer el paso del reconocimiento al uso automatizado.

Así mismo, se propone continuar con estudios más amplios sobre la naturaleza de la colocación en el español como L2, que incluyan análisis de colocaciones en el discurso oral, exploraciones en distintos géneros textuales y la comparación de patrones colocacionales entre aprendices de diferentes lenguas maternas. Esto permitiría comprender con mayor profundidad los procesos cognitivos, culturales y pragmáticos implicados en la adquisición colocacional.

Si bien el presente estudio no tuvo como propósito el diseño de propuestas didácticas ni la elaboración de materiales de enseñanza, los resultados obtenidos sientan las bases para el desarrollo de futuras estrategias pedagógicas centradas en el componente colocacional. En este sentido, el corpus CACE puede emplearse como recurso para la creación de actividades interactivas o secuencias didácticas basadas en datos reales, aprovechando su potencial para la exploración autónoma y el aprendizaje asistido por tecnología.

Asimismo, la información derivada del corpus tiene el potencial de integrarse en plataformas digitales orientadas a la enseñanza del léxico, favoreciendo el reconocimiento de patrones frecuentes y el análisis contrastivo entre la lengua materna y la lengua meta. Estas

posibles aplicaciones no forman parte de los objetivos del presente trabajo, pero constituyen una línea de investigación complementaria con amplias posibilidades de desarrollo.

En síntesis, esta investigación confirma que las colocaciones constituyen una unidad esencial en la construcción del significado y en la fluidez de la expresión en una lengua extranjera. Su dominio no solo mejora la corrección gramatical, sino también la naturalidad y la adecuación pragmática del discurso. Promover el conocimiento colocacional desde una perspectiva didáctica y científica representa un paso clave hacia una enseñanza más efectiva, centrada en el uso real de la lengua. Se espera que este trabajo sirva de base para futuras iniciativas que fortalezcan el enfoque fraseológico en la enseñanza de ELE y que contribuyan al desarrollo de un conocimiento léxico más rico, preciso y contextualizado por parte de los aprendices.

Referencias

- Alonso Ramos, M. (1994). Hacia una definición del concepto de colocación: de J. R. Firth a I. A. Mel'čuk. *Revista de Lexicografía*, 1(1980), 9–28.
<https://doi.org/10.17979/rlex.1995.1.0.5693>
- Alonso Ramos, M. (2016). Spanish learner corpus research: Achievements and challenges. En 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>
- Alonso Ramos, M., Wanner, L., Vincze, O., De lBosque, G. C., Veiga, N. V., Suárez, E. M., y González, S. P. (2010). Towards a motivated annotation schema of collocation errors in learner corpora. *Proceedings of the 7th International Conference on Language Resources and Evaluation, LREC 2010, May 2014*, 3209–3214.
- Álvarez, I., y 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.
- Bahns, J. (1993). Lexical collocations: a contrastive view. *ELT Journal*, 47(1), 56–63.
<https://doi.org/10.1093/ELT/47.1.56>
- Beaugrande, R. (2001). *Large corpora, small corpora, and the learning of “language”* (E. Tognini-Boneli y W. Teubert, Eds.; pp. 3–30). John Benjamins Publishing Company.
- Benson, Morton., Benson, Evelyn., y Ilson, Robert. (1986). *The BBI combinatory dictionary of English : a guide to word combinations*. J. Benjamins Pub. Co.

- Biskup, D. (1992). L1 Influence on Learners' Renderings of English Collocations: A Polish/German Empirical Study. *Vocabulary and Applied Linguistics*, 85–93.
https://doi.org/10.1007/978-1-349-12396-4_8
- Calero, M. Á., Serrano-Zapata, M., y Gómez-Devís, M. B. (2020). Codificación y etiquetado en los corpus de aprendices y su aplicación didáctica: la propuesta del corpus de interlengua española de aprendices sinohablantes (cineas). *E-Aesla. Revista Digital de Lingüística Aplicada*, 6, 206–222.
<https://repositori.udl.cat/bitstream/handle/10459.1/68354/029863.pdf?sequence=1>
- CEDEL2. (n.d.). *CEDEL2: Corpus Escrito del Español L2 (version 2)*. Retrieved May 27, 2021, from <http://cedel2.learnercorpora.com/>
- Celce-Murcia, M., Dörnyei, Z., y Thurrell, S. (1993). A Pedagogical Framework for Communicative Competence: Content Specifications and Guidelines for Communicative Language Teaching. *Deseret Language and Linguistic Society Symposium*, 19(1), 13–19.
- Cestero, A. M., Penadés, I., Blanco, A., Camargo, L., y Simón, J. (2002). Corpus para el análisis de errores de aprendices de E/LE (CORANE). *Tecnologías de La Información y de Las Comunicaciones En La Enseñanza de ELE. Actas Del XII Congreso Internacional de La Asociación Para La Enseñanza Del Español Como Lengua Extranjera*, 2000, 527–534.
- Cheikh-Khamis Cases, F. (2013a). “ ‘No pienses en un oso blanco ’: una propuesta para la enseñanza de léxico desde el enfoque cognitivo ”.
- Cheikh-Khamis Cases, F. (2013b). *LINGÜÍSTICA COGNITIVA APLICADA A LA ENSEÑANZA DEL LÉXICO EN SEGUNDAS LENGUAS. LAS COLOCACIONES EN CLASE DE ELE*. (Vol. 53, Issue 9) [UNED].
<https://doi.org/10.1017/CBO9781107415324.004>
- Cheng, W., Greaves, C., y Warren, M. (2006). From n-gram to skipgram to concgram. *International Journal of Corpus Linguistics*, 11(4), 411.

- Conrad, S. (2002). Corpus linguistic approaches for discourse analysis. *Annual Review of Applied Linguistics*, 22, 75–95.
- Conzett, J. (2000). Entegrating collocation into a reading and writing course. En M. Lewis (Ed.), *Teaching Collocation - Further developments in the lexical approach* (pp. 70–87). Thomson Heinle.
- Cook, V. (2007). Multi-Competence: Black Hole or Wormhole for Second Language Acquisition Research? In Z. Han (Ed.), *Understanding Second Language Process* (pp. 16–26). Multilingual Matters. <https://doi.org/10.21832/9781847690159>
- CORELE. (2018). http://cartago.lllf.uam.es/corele/home_es.html
- Corpas Pastor, G. (1996). *Manual De Fraseologia 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.
- Cruse, D. A. (1986). *Lexical Semantics*. Cambridge University Press.
<https://www.amazon.com/Semantics-Cambridge-Textbooks-Linguistics-1986-09-26/dp/B01FKTJ128>
- Cruz Piñol, M. (2017). *Lingüística de corpus y enseñanza de español como 2/L*. Arco/Libros.
- Cumming, S. (1986). The lexicon in text generation. *Proceedings of the Workshop on Strategic Computing Natural Language - HLT '86*, 242.
<https://doi.org/10.3115/1077146.1077176>
- Davies, Alan. (2003). *The native speaker : myth and reality*. Multilingual Matters.
- De Alba Quiñones, V. (2012). Lexías simples y colocaciones léxicas en la enseñanza de ELE. En *Lexías simples y colocaciones léxicas en la enseñanza de ELE*. Peter Lang D. <https://doi.org/10.3726/978-3-653-01932-2>

- Deveci, T. (2004). Why and how to teach collocations. En *English Teaching Forum*, 42(2), 16–19.
- Ellis, N. C. (1997). Vocabulary acquisition: Word structure, collocation, word-class, and meaning. En N. Schmitt y M. McCarthy (Eds.), *Vocabulary: Description, acquisition and pedagogy* (pp. 122–140). Cambridge University Press.
- Evert, S. (2009). Corpora and collocations. En A. Lüdeling y M. Kytö (Eds.), *Corpus Linguistics: An International Handbook, Volume 2* (Vol. 2, Issue 1, pp. 1212–1249).
- Farzindar, A., y Inkpen, D. (2018). Natural Language Processing for Social Media, Second Edition. *Synthesis Lectures on Human Language Technologies*, 10(2), 1–197. <https://doi.org/10.2200/S00809ED2V01Y201710HLT038>
- Firth, J. R. (1957). Modes of Meaning. En *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.
- Furman, N., Goldberg, D., y Lusin, N. (2009). Enrollments in Languages Other Than English in United States Institutions of Higher Education, Fall 2009. *Modern Language Association of America*.
- Ghadessy, M., Henry, A., y Roseberry, R. L. (2001). *Small Corpus Studies and ELT* (E. Tognini-Bonelli y W. Teubert, Eds.). John Benjamins Publishing Company.
- Granger, S. (2008). Learner Corpora. En A. Lüdeling y M. Kytö (Eds.), *Corpus Linguistics. An International Handbook*. (pp. 259–275). Walter de Gruyter. <https://doi.org/10.1002/9781405198431.wbeal0669>
- Halliday, M. A. K. (1966). Lexis as a linguistic level. En C. Bazell (Ed.), *In memory of J.R. Firth*. (pp. 148–162). Longman.
- Halliday, M. A. K., y Hassan, R. (1976). *Cohesion in English*. Longman.

- Halliday, M. A. K., Matthiessen, C. M. I. M., Halliday, M., y Matthiessen, C. (2014). *An Introduction to Functional Grammar*. Routledge.
<https://doi.org/10.4324/9780203783771>
- Hasselgård, H., y Johansson, S. (2011). Learner corpora and contrastive interlanguage analysis. En F. Meunier, S. De Cock, G. Gilquin, y M. Paquot (Eds.), *A Taste for Corpora* (pp. 33–61). John Benjamins Publishing Company.
<https://doi.org/10.1017/CBO9781107415324.004>
- 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.
- Heid, U., y Raab, S. (1989). Collocations in Multilingual Generation. *4th Conference of the European Chapter of the Association for Computational Linguistics, EACL 1989 - Proceedings*, 130–136. <https://doi.org/10.3115/976815.976833>
- Hernández Sampieri, R., Fernández Collado, C., y Baptista Lucio, M. del P. (2014). *Metodología de la investigación* (6ta ed.). McGraw Hill.
- Higueras García, M. (2006a). *Las colocaciones y su enseñanza en la clase de ELE*. Arco/Libros.
- Higueras García, Marta. (2006b). *Las colocaciones y su enseñanza en la clase de ELE*. Arco/Libros.
- Hincapié, D. (2018). Corpus De Aprendientes De Español Como Lengua Extranjera Y Segunda Lengua (Caele/2): El Componente Escrito. *Forma y Función*, 31(2), 129–143. <https://doi.org/10.15446/FYF.V31N2.74659>
- Hirschberg, J., y Manning, C. D. (2015). Advances in natural language processing. *Science*, 349(6245), 261–266. <https://doi.org/10.1126/science.aaa8685>
- Hoey, M. (2012). *Lexical Priming*. Routledge. <https://doi.org/10.4324/9780203327630>

- Howarth, P. A. (1996). *Phraseology in English Academic Writing*. Max Niemeyer.
<https://www.amazon.com/Phraseology-English-Academic-Writing-Lexicographica/dp/B01K94AHMC>
- Hunston, S. (2002). *Corpora in Applied Linguistics*. Cambridge University Press.
- Hunston, S., y Francis, G. (2000). Pattern grammar : a corpus-driven approach to the lexical grammar of English. En *Computational Linguistics*. John Benjamins.
<https://doi.org/10.1075/scl.4>
- Hussein, R. F. (1990). Collocations: The missing link in vocabulary acquisition amongst EFL learners. *Papers and Studies in Contrastive Linguistics*, 26, 123–136.
- Ibarretxe Antuñano, I. (2010). Lexicografía y Lingüística Cognitiva. *Revista Española de Linguística Aplicada*, 23, 192–213.
https://www.researchgate.net/publication/268203635_Lexicografia_y_Linguistica_Cognitiva
- Ife, A., y Peña Calvo, A. (n.d.). *The Anglia Polytechnic University (APU) Learner Spanish Corpus*.
- Instituto Cervantes. (n.d.). *Corpus de Aprendices de Español como Lengua Extranjera (CAES)*.
- Irwin, R., y Szurmuk, M. (2009). Cultural Studies and the Field of “Spanish” in the US Academy - Dialnet. *A Contracorriente: Revista de Historia Social y Literatura En América Latina*, 6(3), 36–60.
- Jarvis, S. (2017). Grounding lexical diversity in human judgments. *Language Testing*, 34(4), 537–553. <https://doi.org/10.1177/0265532217710632>
- Jiménez Calderón, F., y Rufat, A. (2019). La enseñanza de verbos frecuentes a partir de enfoques léxicos. *Verba Hispanica*, 27(1), 131–151.
<https://doi.org/10.4312/vh.27.1.131-151>

- Kilgariff, A., Baisa, V., Bušta, J., Jakubíček, M., Kovár, V., Michelfeit, J., Rychlý, P., y Suchomel, V. (2014). The sketch engine: Ten years on. *Lexicography*, 1(1), 7–36. <https://doi.org/10.1007/s40607-014-0009-9>
- Koike, D. (2011). *Spanish Corpus Proficiency Level Training*. University of Texas at Austin Liberal Arts Instructional Technology Center.
- Koike, K. (2001). *Colocaciones léxicas en el español actual: estudio formal y léxico semántico*. Universidad de Alcalá y Universidad de Takushoku.
- Kučera, H., y Francis, W. N. (1967). *Computational analysis of present-day American English*. Brown University Press.
- Lakoff, G., y Johnson, M. (1999). *Philosophy in the flesh : the embodied mind and its challenge to Western thought*. Basic Books.
- Laufer, B., y Waldman, T. (2011). Verb-Noun Collocations in Second Language Writing: A Corpus Analysis of Learners' English. *Language Learning*, 61(2), 647–672. <https://doi.org/10.1111/j.1467-9922.2010.00621.x>
- Learner corpora around the world | UCLouvain*. (n.d.). Retrieved September 23, 2022, from <https://uclouvain.be/en/research-institutes/ilc/cecl/learner-corpora-around-the-world.html>
- Leech, G. (2011). Frequency, corpora and language learning. En F. Meunier, S. De Cock, G. Gilquin, y M. Paquot (Eds.), *A Taste for Corpora* (pp. 7–27). John Benjamins B.V. <https://doi.org/10.1017/CBO9781107415324.004>
- Lewis, M. (1997). *Implementing the Lexical Approach: Putting Theory into Practice*. *Undefined*.
- Lozano, C., y 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.

- Lyons, J. (1966). Firth's Theory of "Meaning." In C. E. Bazell, J. C. Catford, M. A. K. Halliday, y R. H. Robins (Eds.), *In memory of J. R. Firth* (pp. 288–302). Longmans, Green and Co. Ltd. <https://doi.org/10.1017/S0041977X00132756>
- Marton, W. (1977). Foreign vocabulary learning as problem no. 1 of language teaching at the advanced level. *Enterlanguage Studies Bulletin*, 2(1), 33–57.
- Martos Eliche, F., y Contreras Izquierdo, N. M. (2018a). El empleo de corpus para el aprendizaje de secuencias formulaicas en ELE/EL2. La frecuencia de uso en el nivel B2 del PCIC. *CHIMERA: Revista de Corpus de Lenguas Romanes y Estudios Lingüísticos*, 5(1), 1. <https://doi.org/10.15366/chimera2018.5.1.001>
- Martos Eliche, F., y Contreras Izquierdo, N. M. (2018b). El empleo de corpus para el aprendizaje de secuencias formulaicas en ELE/EL2. La frecuencia de uso en el nivel B2 del PCIC. *CHIMERA: Revista de Corpus de Lenguas Romanes y Estudios Lingüísticos*, 5(1), 1. <https://doi.org/10.15366/chimera2018.5.1.001>
- McEnery, T., y Hardie, A. (2012). *Corpus Linguistics: Method, theory and practice*. Cambridge University Press. <https://doi.org/10.4324/9781315107769-2>
- Meara, P. (1980). Vocabulary acquisition: A neglected aspect of language learning. En *Language Teaching* (Vol. 13, Issues 3–4, pp. 221–246). Cambridge University Press. <https://doi.org/10.1017/S0261444800008879>
- Mel'čuk, I. A., y Arbatchewsky-Jumarie, Nadia. (1984). *Dictionnaire explicatif et combinatoire du français contemporain*. Presses de l'Université de Montréal.
- Mellado Blanco, C. (2008). *Colocaciones y fraseología en los diccionarios*. Frankfurt am Main, Peter Lang.
- Milton, J. (2009). *Measuring second language vocabulary acquisition*. Multilingual Matters.

- Mitchell, R., Domínguez, L., Arche, M. J., Myles, F., y Marsden, E. (2008). SPLLOC: A new database for Spanish second language acquisition research. *EUROSLA Yearbook*, 8, 287–304. <https://doi.org/10.1075/eurosla.8.15smit>
- Naeimi, M., y 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>
- Nation, I. S. P. (2001). *Learning Vocabulary in Another Language*. Cambridge University Press. <https://doi.org/10.1017/CBO9781139524759>
- Ñaupas Paitán, H., Mejía Mejía, E., Novoa Ramírez, E., y Villagómez Paucar, A. (2014). *Metodología de la Investigación Cualitativa - Cuantitativa y Redacción de la Tesis*. Ediciones de la U.
- Navajas Algaba, A. (2006). *Las colocaciones en el aula de ELE: actividades para su explotación didáctica*. Universidad Antonio de Nebrija.
- Nesselhauf, N. (2005). *Collocations in a Learner Corpus* (Vol. 14). John Benjamins Publishing Company. <https://doi.org/10.1075/scl.14>
- Nirenburg, S., y Nirenburg, I. (1988). A framework for lexical selection in natural language generation. *Proceedings of the 12th Conference on Computational Linguistics* -, 2, 471–475. <https://doi.org/10.3115/991719.991736>
- Núñez, E. (2019). Pasado, presente y futuro de los corpus de aprendices de ELE. Una revisión bibliográfica. *ReiDoCrea: Revista Electrónica de Investigación Docencia Creativa*, August. <https://doi.org/10.30827/Digibug.58473>
- Oportus Torres, R., y Ferreira Cabrera, A. (2020). Colocaciones gramaticales verbo+preposición en español como lengua extranjera: análisis descriptivo y lingüístico de errores de aprendientes de niveles A2 y B1. *Cuadernos de Lingüística de El Colegio de México*, 1, 1–51. <https://doi.org/10.24201/clecm.v7i0.146>

- Paquot, M., König, A., Stemle, E. W., y Frey, J.-C. (2024). The Core Metadata Schema for Learner Corpora (LC-meta). *International Journal of Learner Corpus Research*, 10(2), 280–300. <https://doi.org/10.1075/ijlcr.24010.paq>
- Parodi, G. (2015). Corpus de aprendices de español (CAES). *Journal of Spanish Language Teaching*, 2(2), 194–200. <https://doi.org/10.1080/23247797.2015.1084685>
- Pérez Serrano, M. (2014). Análisis de errores colocacionales en un corpus de aprendientes de ELE. *MarcoELE Revista de Didáctica ELE*, 19.
- Pottier, B. (1969). *Entroduction à l'étude des structures grammaticales fondamentales*. Publications Linguistiques de la Faculté des Lettres et Sciences Humaines de Nancy.
- Qian, D. D., y Schedl, M. (2004). Evaluation of an in-depth vocabulary knowledge measure for assessing reading performance. *Language Testing*, 21(1), 28–52. <https://doi.org/10.1191/0265532204lt273oa>
- Ragan, P. H. (2001). Classroom use of a systemic functional small learner corpus. En M. Ghadessy, A. Henry, y R. L. Roseberry (Eds.), *Small Corpus Studies and ELT: Theory and practice* (pp. 207–236). John Benjamins Publishing Company. <https://doi.org/10.1075/scl.5.14rag>
- Read, J. (2000). Assessing Vocabulary. En *Assessing Vocabulary*. Cambridge University Press. <https://doi.org/10.1017/cbo9780511732942>
- Rica Peromingo, J. P. (2010). Colocaciones gramaticales en la producción escrita de estudiantes universitarios españoles. *Reduca (Filología)*. *Serie Lengua Inglesa*, 1, 1–25.
- Richards, J. C. (1976). The Role of Vocabulary Teaching. *TESOL Quarterly*, 10(1), 77–89. <https://doi.org/10.2307/3585941>

- Richards, J. C., y Schmidt, R. W. (2013). Longman Dictionary of Language Teaching and Applied Linguistics. En *Longman Dictionary of Language Teaching and Applied Linguistics*. Routledge. <https://doi.org/10.4324/9781315833835>
- Rodríguez-Fernández, S., Carlini, R., y Wanner, L. (2015). Classification of grammatical collocation errors in the writings of learners of Spanish. *Procesamiento Del Lenguaje Natural*, 55, 49–56.
- Sampedro Mella, M. (2022). Corpus de aprendices y cuestiones de discurso en la interlengua: un nuevo acercamiento aplicado a la enseñanza. *Foro de Profesores de E/LE*, 18, 163. <https://doi.org/10.7203/foroele.18.24965>
- San Martín, L. P. B., Cabrera, A. A. F., y San Martín, E. L. B. (2019). Colocaciones léxicas verbo-nominales en un corpus de aprendices de Español como Lengua Extranjera de nivel A2 y B1. *Letras de Hoje*, 54(3), 407. <https://doi.org/10.15448/1984-7726.2019.3.33243>
- Schmitt, N., y Schmitt, D. (2020). *Vocabulary in Language Teaching*. Cambridge University Press. <https://doi.org/10.1017/9781108569057>
- Sierra, G. E. (2015). Entroducción a los corpus lingüísticos. En *Introducción a los corpus lingüísticos*. Universidad Nacional Autónoma de México, Instituto de Ingeniería.
- 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>
- Sinclair, J. (2004). Trust the Text. En R. Carter (Ed.), *Trust the Text*. Routledge. <https://doi.org/10.4324/9780203594070>
- Sinclair, J., Jones, S., y Daley, R. (1970). *English collocation studies : the OSTI report on Project C/ LP/08*. University of Birmingham.

- Stefanowitsch, A., y Gries, S. Th. (2003). Collocations: Investigating the interaction of words and constructions. *International Journal of Corpus Linguistics*, 8(2), 209–243. <https://doi.org/10.1075/ijcl.8.2.03ste>
- Swain, M., y Carroll, S. (1987). Vocabulary instruction in immersion classes. En P. A. B. Harley (Ed.), *The development of bilingual proficiency* (pp. 192–222). Modern Language Centre, Ontario Institute for Studies in Education.
- Tetreault, J., y Leacock, C. (2014). Automated grammatical error correction for language learners. *COLING 2014 - 25th International Conference on Computational Linguistics, Proceedings of Tutorial Abstracts*, 8–10.
- Teubert, W. (2004). Units of meaning, parallel corpora, and their implications for language teaching. En U. Connor y T. A. Upton (Eds.), *Applied Linguistics: A Multidimensional Perspective* (pp. 171–189). Rodopi.
- Tinsley, T., y Board, K. (2017). *Languages for the future*. https://www.britishcouncil.org/sites/default/files/languages_for_the_future_2017.pdf
- Tracy-Ventura, N., Mitchell, R., y McManus, K. (2016). The LANGSNAP longitudinal learner corpus: Design and use. En *Spanish Learner Corpus Research* (pp. 117–142). John Benjamins.
- Travalia, C. (2001). Las colocaciones coloquiales en español. *Congreso Internacional de Fraseología y Paremiología*.
- 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.

- Vazquez, E. C. (2021). Power , Policy , Profit : The Spanish Language in the United States. *Alpenglow : Binghamton University Undergraduate Journal of Research and Creative Activity*, 7(1).
- Viera, C. (2019). *Reseña de Alonso Ramos, Margarita. 2016.*
<https://infoling.org/informacion/Review353.html>
- Villayandre, M., y Llanos, L. (2017). Los corpus electrónicos en la clase de español: reflexiones y aplicaciones. En B. Peña y A. Aguilar (Eds.), *Didáctica de la lengua y la literatura. Buenas prácticas docentes: Vol. I* (pp. 52–79). ACCI.
- Vincze, O., García-Salido, M., Orol, A., y Alonso Ramos, M. (2016). A corpus study of Spanish as a foreign language learners' collocation production. En 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>
- Xi, X. (2008). Encyclopedia of Language and Education. En N. H. Hornberger (Ed.), *Encyclopedia of Language and Education*. Springer US.
<https://doi.org/10.1007/978-0-387-30424-3>
- Zipf, G. K. (1935). *The Psychobiology of Language*. Houghton Mifflin.
- Zipf, G. K. (1949). *Human behavior and the principle of least effort*. Addison-Wesley.

Anexo 1. Perfil de participante

En esta sección se muestran los ejemplos de respuesta, indicadas en letras negritas, proporcionadas por uno de los participantes.

1. Country of origin. [País de origen]

a. Poland

2. Gender. [Género]

a. Mujer

b. Hombre

c. Prefiero no decir

3. Age group. [Rango de edad]

a. 36-40

4. Level of Spanish. [Nivel de español] – How do you rate your own proficiency level in Spanish? [Según tu opinión, ¿cómo clasificarías tu nivel de dominio del español?

a. Básico

b. Intermedio

c. Avanzado

d. Sobresaliente

5. Where did you study Spanish? [¿Dónde estudiaste español?]

a. Escuela

b. Otro:

6. If you answered “other” in the previous question, how did you learn Spanish? [Si respondiste “otra” a la pregunta anterior, ¿cómo aprendiste español?]

7. Where did you mainly study Spanish? [¿En dónde estudiaste español principalmente?]
- a. En mi país**
 - b. En el extranjero
8. Do you have any Spanish language certifications? [¿Tienes alguna certificación de la lengua española?]
- a. Sí**
 - b. No
9. If you answered “yes” to having a Spanish certification, what is its name? [Si respondiste que sí tienes una certificación en español, ¿cuál es su nombre?]
10. Have you ever practiced Spanish abroad? [¿Has practicado el español en el extranjero?]
- a. Sí
 - b. No**
11. If you answered “yes” to studying Spanish abroad, which country did you study in? [Si respondiste que sí has estudiado español en el extranjero, ¿en qué país estudiaste?]
12. How long have you been studying Spanish? [¿Cuántos años has estudiado español?]
- a. 0-1 años
 - b. 2-5 años**
 - c. 5-10 años
 - d. Más de 10 años
13. What is your native language? [¿Cuál es tu lengua nativa?]
- a. Polish**

Anexo 2. El cuestionario

En esta sección se muestran los ejemplos de respuesta, indicadas en letras negritas, proporcionadas por uno de los participantes.

Instructions (instrucciones)

In this section, answer a few multiple-choice questions and write a short paragraph. Do not concern yourself with making mistakes, please express yourself as naturally as possible. (En esta sección responde algunas preguntas de opción múltiple y escribe un párrafo corto. Por favor no te preocupes por cometer errores e intenta expresarse de la manera más natural posible).

1. Laura fue invitada a _____ el cargo de director.

a. **Desempeñar**

b. Cumplir

c. Ejecutar

2. Se _____ falso en la escuela.

a. Esparció un rumor

b. Dijo un rumor

c. **Corrió un rumor**

3. Escuché la noticia de una fuente_____.

a. Veraz

b. **Fidedigna**

c. Verdadera

4. El niño comió ____.
- a. **Una tajada de pan**
 - b. Una rodaja de pan
 - c. Una rebanada de pan
5. El bebé está ____.
- a. **Profundamente dormido**
 - b. Muy dormido
 - c. Realmente dormido
6. Quiero ____ lo que aprendí en el curso
- a. Meter en práctica
 - b. **Poner en práctica**
 - c. Llevar a práctica
7. Se dice que el amor es ____.
- a. Intenso
 - b. Caprichoso
 - c. **Ciego**
8. Jorge está ____ afectado por la muerte de su perro.
- a. Visiblemente
 - b. Evidentemente
 - c. **Claramente**
9. El conductor ____ después del choque.
- a. **Salió ileso**
 - b. Resultó ileso

- c. Quedó ileso
10. Los insurgentes van a _____.
- a. Dar un golpe de estado**
 - b. Poner un golpe de estado
 - c. Atentar un golpe de estado
11. En la universidad _____ un proyecto de negocios.
- a. Llevé a término
 - b. Ejecuté
 - c. Llevé a cabo**
12. Jorge llora _____ por la pérdida de su perro.
- a. A moco tendido
 - b. Echado
 - c. A moco escurrido**
13. El hombre está _____.
- a. Tonto de remate
 - b. Loco de remate**
 - c. Tocado de remate
14. En la conferencia, se habló _____ la bolsa de valores.
- a. Acerca de**
 - b. Con relación a
 - c. Referente a
15. Tengo _____ tropezarme con estos tacones.
- a. Miedo de**

- b. Miedo a
 - c. Miedo por
16. Los niños tienen _____ ser felices.
- a. Derecho de
 - b. Derecho a**
 - c. Derecho por
17. Todo saldrá bien, _____ ello.
- a. Estoy seguro sobre
 - b. Estoy seguro en
 - c. Estoy seguro de**
18. El entrenador _____ a los ganadores por altavoz.
- a. Anunció**
 - b. Informó
 - c. Advirtió
19. El policía _____ sus sospechas.
- a. Verificó
 - b. Constató
 - c. Confirmó**
20. José es más alto _____ Esteban.
- a. En comparación con**
 - b. A comparación de
 - c. En comparación a
21. _____ una linda amistad con mi vecina.

a. Entablé

b. Empecé

c. Emprendí

22. Mi familia y yo tenemos una _____.

a. Relación grande

b. Relación cercana

c. Relación estrecha

23. La universidad ha comenzado _____.

a. Una etapa de conferencias

b. Un ciclo de conferencias

c. Un periodo de conferencias

24. El aprendizaje de lenguas está _____ al conocimiento de su cultura.

a. Grandemente ligado

b. Altamente ligado

c. Estrechamente ligado

25. Juan siempre _____ sus bromas.

a. Anda con

b. Anda en

c. Anda por

26. El niño tiene una conducta _____.

a. Recta

b. Intachable

c. Íntegra

27. Claudia está llorando _____.

- a. **Amargamente**
- b. Tristemente
- c. Dolorosamente

28. El político rechazó _____ la acusación.

- a. Claramente
- b. Terminantemente
- c. Categóricamente

29. Ese chico está _____.

- a. Grandemente enamorado
- b. **Perdidamente enamorado**
- c. Frenéticamente enamorado

30. La profesora preguntó si su explicación _____.

- a. **Quedó clara**
- b. Quedó esclarecida
- c. Quedó aclarada

31. Mi _____ es sencillo.

- a. Modo de vida
- b. Manera de vida
- c. **Estilo de vida**

32. Pedro dejó _____ a Susana, simplemente no llegó.

- a. Abandonada
- b. **Plantada**

- c. Sola
33. Debo _____ para el análisis de sangre.
- a. Estar en ayunas**
 - b. Quedarme en ayunas
 - c. Hacer ayunas
34. La película me hace _____ mi infancia.
- a. Pensar sobre
 - b. Pensar de
 - c. Pensar en**
35. La explicación quedó _____.
- a. Clara como el agua
 - b. Clara como el cristal
 - c. Cristalina como el agua**
36. Demuestra _____ el arte.
- a. Amor para
 - b. Amor por**
 - c. Amor en
37. Pocas veces tengo _____ relajarme.
- a. El placer sobre
 - b. El placer en**
 - c. El placer de
38. Puedes contar con tu familia _____.
- a. De por vida

b. Por la vida

c. Para la vida

39. Todos _____ la noticia en la televisión.

a. Miraron

b. Vieron

c. Observaron

40. El juez no está _____ su inocencia

a. Convencido sobre

b. Convencido con

c. Convencido de

Anexo 3. Colocaciones encontradas en el CACE

a. Colocaciones léxicas simples (Corpas Pastor, 1996; Koike, 2001)

Tipo de colocación	Colocación			IM
	Base	Colocativo	Freq. relativa de la colocación	
Sustantivo + Adjetivo	Año	pasado	2	11.02
	Lugar	primer	3	11.5
	Amigo	querido	19	9.11
	Correo	electrónico	4	11.24
	Abrazo	fuerte	4	11.02
	Asunto	importante	2	11.5
	Hermano	menor	2	8.74
	Fiesta	próxima	13	8.97
	Amigo	mejor	3	10.02
	Día	buen	10	10.09
	Abrazo	gran	4	10.82
	Tarde	buen	3	10.43
	Vez	próximo	8	8.97
	Noche	buen	5	10.69
	Semana	próximo	2	8.97
	cumpleaños	feliz	2	6.32
	Beso	gran	3	10.92

Tipo de colocación	Colocación			IM
	Base	Colocativo	Freq. relativa de la colocación	
Verbo + Sustantivo (Objeto)	pedir	disculpa	13	9.12
	pedir	perdón	4	9.65

Tipo de colocación	Colocación			
	Base	Colocativo	Freq. relativa de la colocación	IM
Verbo + Adjetivo	estar	enfermo	7	5.68
	estar	presente	6	8.54
	estar	mejor	2	8.58

Tipo de colocación	Colocación			
	Base	Colocativo	Freq. relativa de la colocación	IM
Sustantivo + Preposición + Sustantivo	dolor	cabeza	3	11.02
	examen	período	2	10.5
	semana	fin	8	10.02

Tipo de colocación	Colocación			
	Base	Colocativo	Freq. relativa de la colocación	IM
Verbo + Adverbio (mente)	llegar	finalmente	2	11.24

b. Colocaciones gramaticales (Travalia, 2006)

Tipo de colocación	Colocación			IM
	Base	Colocativo	Freq. relativa de la colocación	
verbo + preposición	asistir	a	22	5.47
	perdonar	por	12	6.08
	invitar	a	9	5.3
	pensar	en	3	6.23
	creer	que	12	4.93
	dejar	de	2	5.38
	ayudar	a	6	5.47
	preocupar	por	2	6.08
	olvidar	de	4	5.38
	agradecer	por	2	6.08
	cenar	en	2	6.23
	entregar	a	2	5.47
	despertar	a	2	5.47
	participar	en	2	6.23

Tipo de colocación	Colocación			IM
	Base	Colocativo	Freq. relativa de la colocación	
(verbo) + adjetivo + preposición	lleno de	estar	3	5.38
	presente en	estar	2	9.65