

ARTÍCULO

Colocaciones gramaticales verbo+preposición en español como lengua extranjera: análisis descriptivo y lingüístico de errores de aprendientes de niveles A2 y B1

Grammatical collocations verb+preposition in Spanish as a foreign language: descriptive and linguistic error analysis of learners of levels A2 and B1

René Oportus Torres

reneoportus@gmail.com

Anita Ferreira Cabrera

aferreir@udec.cl

Universidad de Concepción, Chile

Original recibido: 2019/10/06

Dictamen enviado al autor: 2019/22/10

Aceptado: 2020/08/01

Abstract

This article informs of the findings of a study of Spanish as a foreign language learners' corpus of levels A2 and B1 whose focus examines errors in the use of grammatical collocations of the type verb+preposition (GramCol v+p). Its aim is to determine and compare their frequency in writing according to competence and under location, descriptive and syntactic-semantic criteria. This study conceptualizes GramCol from a phraseological perspective, and is based on the models of Error Analysis, and Contrastive Interlanguage Analysis. The results show that the most frequent errors: (1) affect the whole collocation, (2) are of the addition type, and (3) concentrate on those units whose collocates alternate

with others expressing the same meaning, and on those whose collocates may appear implicit. Moreover, the data point out a similar frequency in the errors of levels A2 and B1 regarding the three analyzed criteria. The method and design of this study contribute to the research of intermediate fixedness phraseological units, specifically in respect to the nature of errors that affect ColGram v+p.

Keywords: grammatical collocations verb+preposition; error analysis; interlanguage contrastive analysis; phraseological units of intermediate fixedness

Resumen

Este artículo reporta los hallazgos de un estudio de corpus de aprendientes anglófonos de español como lengua extranjera (ELE) de niveles A2 y B1 cuyo foco examina los errores en el uso de colocaciones gramaticales verbo + preposición (ColGram v+p). El objetivo es determinar y comparar la frecuencia en la escritura según la competencia y bajo los criterios de localización, descriptivo y sintáctico-semántico. El trabajo conceptualiza las ColGram desde una perspectiva fraseológica y se sustenta en los modelos de Análisis de Errores y del Análisis Contrastivo de la Interlengua. Los resultados muestran que los errores más frecuentes: (1) afectan a la colocación completa, (2) son del tipo de adición, y (3) se concentran en las unidades cuyos colocativos alternan entre sí para expresar el mismo significado y en aquellas cuyo colocativo puede aparecer implícito. Por otra parte, los datos apuntan a una frecuencia similar en los niveles A2 y B1 respecto de los tres criterios analizados. La metodología y diseño de este trabajo contribuyen a la investigación de unidades fraseológicas de fijación intermedia, específicamente en lo que respecta a la naturaleza de los errores que afectan a las ColGram v+p.

Palabras clave: colocaciones gramaticales verbo+preposición; análisis de errores; análisis contrastivo de la interlengua; unidades fraseológicas de fijación intermedia

1. INTRODUCCIÓN¹

En los estudios léxicos en Adquisición de Segundas Lenguas (ASL), una línea de investigación que ha suscitado un interés creciente de los especialistas se ha focalizado en las unidades fraseológicas de fijación intermedia, vale decir, las colocaciones. Actualmente, existe un acuerdo amplio sobre la importancia del desarrollo de la competencia colocacional respecto de que su manejo y comprensión contribuyen a la naturalidad, precisión, fluidez y eficiencia en la producción y procesamiento de la L2 (Moreno 2009; Cheikh-Khamis 2013; Quiñones 2016). Así también, desde una perspectiva pedagógica, se recalca que conocer una palabra implica no solo aprender sus diferentes acepciones, usos y áreas temáticas; sino que también los contextos lingüísticos en los que aparece; es decir, los términos con los que combina frecuentemente (Lu 2016).

En español como lengua extranjera (ELE), un área particular de la temática colocacional que ha recibido escasa atención ha sido el estudio de las ColGram, en contraste con otras expresiones fraseológicas, o con el grupo específico de las colocaciones léxicas (Travalia 2006; Domínguez 2014). Esto derivaría, primero, de los múltiples planteamientos conceptuales disponibles en la literatura especializada en torno a los rasgos que definen a las colocaciones como tales; lo que ha conducido a una evidente confusión terminológica (Sánchez 2010; Gysllrad & Wolter 2015). En

¹ Este artículo se ha desarrollado en el contexto del proyecto de investigación FONDECYT N° 1180974 “Diseño e implementación de un corpus escrito de aprendientes de ELE en formato computacional para el análisis de la interlengua”, cuya investigadora responsable es la Dra. Anita Ferreira Cabrera.

segundo lugar, aparece el cuestionamiento hacia el grupo específico de las ColGram, en términos de si estas deben ser efectivamente incluidas y examinadas desde la perspectiva fraseológica, o si bien corresponderían a fenómenos abordados por otras parcelas lingüísticas (Fernández 2014; Vincze 2015).

Dentro de las ColGram, Rica (2010) destaca el tipo verbo+preposición y la necesidad de su estudio, pues estos patrones constituyen la categoría más amplia dentro de dichas unidades. Por su parte, Henriksen (2013: 36) subraya la necesidad de emprender estudios específicos, considerando que se han realizado pocos intentos de clasificación sistemática de las diferentes subcategorías colocacionales. Así también, debido a su uso idiosincrático, las ColGram constituyen un área de dificultad para el aprendizaje de una L2 (Ferraro et al. 2011). Lo cierto es que los antecedentes disponibles en la literatura especializada son limitados; a saber, solo contamos con los trabajos de Domínguez (2014), y Oportus y Ferreira (2019; 2020), por lo que se hace necesario recabar evidencia empírica sobre el manejo y errores en el uso de las ColGram v+p en ELE.

Este trabajo se enmarca dentro de una investigación mayor que aborda el manejo y procesamiento de las ColGram v+p en aprendientes de ELE.² Específicamente, el presente estudio corresponde a la etapa que responde a la necesidad de examinar la naturaleza de los errores en la escritura colectiva de aprendientes anglófonos en el marco de la función diag-

² Este estudio forma parte de la investigación de la tesis doctoral *Influencia de la L1 en el Procesamiento y Errores en el Uso de Colocaciones Gramaticales en ELE* del autor principal, bajo la dirección de la Dra. Anita Ferreira Cabrera.

nóstica del Análisis de la Interlengua (ACI) (Sánchez 2015a).³ La interrogante que se intenta responder es si los niveles de competencia (A2 y B2) difieren en la frecuencia de errores de acuerdo a los criterios de localización, patrón sintáctico-semántico, y mecanismos superficiales que caracterizan a estas desviaciones. Para este fin, los errores son analizados aplicando una taxonomía específica, representativa de dichos criterios. En este proceso, se adopta una perspectiva fraseológica sobre el concepto de ColGram v+p; en donde la combinación v+p contrasta con las combinaciones libres y combinaciones fijas de la lengua (Buyse 2003).

Este manuscrito se organiza de la siguiente manera. La sección 2 desarrolla sucintamente la evolución del estudio colocacional y en particular del constructo de ColGram v+p en los estudios en ELE. En este apartado se incluye una descripción de los modelos relativos al estudio de la interlengua (IL) y los errores; así como de las herramientas utilizadas. Por último, en esta sección, se incluyen los antecedentes investigativos previos vinculados a este estudio. La sección 3 incluye los datos más relevantes del estudio: objetivos, metodología, el corpus y la muestra, la conceptualización del objeto de estudio y su delimitación; así como, las especificaciones de la taxonomía y análisis empleados. La cuarta sección reporta los resultados y discusión. Finalmente, la última sección comprende las conclusiones del trabajo.

³ El uso de ColGram v+p aparece como una problemática de dichos aprendientes, la que se deduce de trabajos previos relativos al manejo preposicional en aprendientes de ELE (Ferreira et al. 2014; Ferreira & Lafleur 2015; Ferreira et al. 2016; Ferreira & Elejalde 2017; Ferreira & Oportus 2018).

2. MARCO TEÓRICO

2.1 *Enfoques de estudio colocacional*

El estudio del léxico durante largo tiempo ocupó un lugar secundario en la lingüística teórica. Bajo la influencia del paradigma generativista la preocupación central recaía en la oración y aspectos que rompían su gramaticalidad; en contraste, el léxico permanecía relegado al estudio de la semántica y filosofía del lenguaje. Bajo este marco, surge el interés por el ámbito no creativo de la lengua referido a los patrones repetidos a lo largo de diferentes tipos textuales, como es el caso de las colocaciones. En esta línea, el léxico adquirió importancia dentro de la *competencia lingüística*, en donde léxico y gramática aparecen integrados e interconectados; perspectiva a la que adhiere el concepto de ColGram v+p de este trabajo.

Dos enfoques lexicológicos principales cimentaron el estudio y conceptualización actual del término colocación, las corrientes semántico-fraseológica y la estadística; siendo la primera la más cercana a ASL (Nesselhauf 2005; Gyllstad 2007; Moreno 2009; Henriksen 2013; Santillán 2015). Dicha corriente se basa en relaciones sintáctico-semánticas, en donde los constituyentes de una colocación expresan una relación típica, de fijación intermedia, y de significado relativamente literal y transparente (Buyse 2003).

En cambio, la corriente estadística determina la identificación de las colocaciones de acuerdo a la probabilidad estadística de que dos palabras ocurran a corta distancia en el discurso (Sinclair 1991); es decir, se

basa en criterios objetivos, tales como, la frecuencia y el rango de amplitud colocacional. Sin embargo, debido a que opera sobre los datos duros, requiere también del análisis semántico; de otra manera, es conducente a detectar combinaciones de alta frecuencia, pero de escasa validez psicolingüística. A su vez, la corriente semántico-fraseológica, de carácter más subjetivo, permite identificar unidades con clara relación semántica, pero falla en dar cuenta de su frecuencia real; lo que reduce la validez de los datos en pos de tomar decisiones relativas a la enseñanza.

Una tendencia reciente a la aproximación del fenómeno colocacional combina las ventajas de ambos enfoques bajo un esquema híbrido, por ejemplo, Nesselhauf (2005) y Moreno (2009). Siguiendo esta senda, el presente estudio posee un carácter polivalente. Se alinea eminentemente con el enfoque lexicográfico de la corriente semántico-fraseológica, lo que presenta una desventaja a nivel operativo, pues depende del trabajo manual para la identificación y clasificación de las unidades (Moreno 2009). No obstante, también se emplean procedimientos de la corriente estadística, en lo referido a la obtención de datos de frecuencia y medidas de probabilidad estadística (Henriksen 2013).

2.2 Conceptualización de las ColGram v+p en ELE

Desde un punto de vista semántico fraseológico, el concepto de colocación es de naturaleza difusa y compleja, sin acuerdo unánime entre los especialistas; rasgo que ha generado dificultades en su investigación (Alonso 2010). Numerosos trabajos han intentado clarificar los aspectos terminológicos y conceptuales contrastando sus rasgos distintivos respec-

to de otras unidades fraseológicas. No obstante, estos han resultado abstractos y subjetivos; por ejemplo, los referidos al grado de fijación entre los constituyentes colocacionales.

El constructo tradicional de ColGram se ha desarrollado naturalmente bajo el alero de las colocaciones léxicas, por lo que las consideraciones anteriores también tienen su aplicación en estas unidades. Adicionalmente, ha existido un intenso debate en torno a qué marco disciplinario debería encargarse de su estudio (Quiñones 2016). Se argumenta que como la concepción de las ColGram implica relaciones sintagmáticas, estas deberían ser estudiadas desde parcelas lingüísticas distintas de la léxica. Específicamente, se ha planteado que el fenómeno de las ColGram corresponde a patrones de régimen y, por ende, debería ser tratado bajo el prisma de los estudios sintácticos. Esta apreciación se contrapone a la visión imperante en el ámbito de ASL, compartida por este estudio.

La identificación de las ColGram v+p como un fenómeno fraseológico se origina en la propuesta tipológica de Benson et al. (1986) para el inglés. En español, algunos autores han validado por defecto dicha clasificación, mientras que otros han ajustado dicho esquema a los rasgos particulares de la lengua. En particular, la propuesta de Travalía (2006) profundiza en consideraciones conceptuales que señalan como característica fundamental la fijación intermedia entre un término léxico y uno gramatical, en contraste con las combinaciones obligatorias —como son los complementos de régimen verbal (Zato 2011). También la investigadora propone una clasificación que incluye distinciones de diferentes tipos de ColGram, identificando grupos y subgrupos de patrones colocacionales descritos en base a consideraciones sintáctico-semánticas.

Tomando en consideración lo anterior, se puede proponer que las Col-Gram v+p para el español tendrían como correlato gramatical a algunos sintagmas preposicionales que funcionan como complementos de régimen preposicional (CPREP), o como complementos del predicativo no obligatorios (RAE & AAL 2009; Vaisberg 2013: 10). Esto facilita la operacionalización de nuestro objeto de estudio.

2.3 El estudio de la interlengua y Corpus Electrónicos de Aprendientes

El objetivo principal de la investigación en ASL es develar los principios que rigen el aprendizaje de una segunda lengua. Dado que este es un proceso mental no observable, debe estudiarse indirectamente; por ejemplo, a través de su resultado, el desempeño lingüístico (Granger 1998). Lo anterior implica llevar a cabo el análisis y descripción de la IL del aprendiente, de la forma más completa posible, incluyendo los elementos producidos correcta o incorrectamente. A partir de estos datos, el investigador puede inferir cómo proceden las etapas del desarrollo de la IL.

Los procedimientos de la Lingüística de Corpus (LC) ponen su foco en la naturaleza sistemática de la lengua, es decir, el campo conceptual de la actuación (Bolaños 2015). Teóricamente se orienta por dos planteamientos fundamentales: la función social de la lengua y la importancia del contexto situacional. En consecuencia, el significado de las palabras no puede separarse de su uso; por lo que el estudio lingüístico debe basarse en el análisis de la lengua en su uso real (Santillán 2015). Esta aproximación metodológica implica la recolección de muestras extensas de textos producidos en actividades de comunicación auténtica; de lo que se

desprende la mayor ventaja de la LC, su objetividad, que deriva de datos empíricos (Bolaños 2015; Santillán 2015).

La investigación bajo la línea de la LC comienza a mostrar su impacto en ASL a partir de la década de los noventa con la compilación de corpora de aprendientes de inglés (Granger 1998). Posteriormente, esta hace su llegada a ELE con la implementación de varios corpus informatizados de aprendientes de español, por ejemplo, CEDEL2 (Lozano & Medndi-koetxea 2013) y CAELE (Ferreira & Elejalde 2017), entre otros.

El área investigativa de corpus electrónicos de aprendientes de lenguas reúne los planteamientos de la LC y el ámbito de ASL con el objetivo de producir descripciones enriquecidas de la lengua de aprendientes (Granger 2002; 2004; Sánchez 2015a). En la actualidad, como resultado de la innovación tecnológica y las nuevas técnicas computacionales, surgen los corpus electrónicos de aprendientes que facilitan la recolección, procesamiento y análisis de datos mediante el uso de programas computacionales especializados (Granger 2004; Ferreira et al., 2014; López 2014; Elejalde & Ferreira 2016).

2.4 Los modelos de Análisis de Errores y de Análisis Contrastivo de la Interlengua

En las décadas de los sesenta y setenta surge el modelo de Análisis de Errores (AE). Este se focaliza en la producción concreta del aprendiz; lo que implica un cambio de foco, desde los modelos abstractos previos hacia la obtención de evidencia empírica para caracterizar y explicar los errores en base a su propio análisis (Quiñones 2009; Alexopoulou

2010). En este marco, la adquisición se concibe cognitivamente como un proceso creativo de acercamiento a la lengua objeto por medio de la comprobación de hipótesis conducentes a la generación de reglas. En definitiva, el AE intenta recoger evidencia sobre los errores de los estudiantes en un momento determinado, considerando que son inevitables y necesarios para el aprendizaje (Corder 1967; 1971).

Los datos lingüísticos obtenidos mediante el análisis de corpus electrónicos permiten realizar varios contrastes para identificar el manejo y dificultades en los aprendientes. El ACI (Granger 1998) se ha convertido en el procedimiento estándar para llevar a cabo estos contrastes (Ishikawa 2013). El ACI trata a la IL como una entidad autónoma e independiente (Granger 2015), y su enfoque comparativo de naturaleza cuantitativa y cualitativa permite identificar elementos caracterizadores de variedades de la IL que comparten rasgos en común (Sánchez 2015a); por ejemplo, entre distintas lenguas maternas o entre grupos de diferentes niveles de competencia; además de la incorporación de otros factores inherentes al proceso de adquisición con el fin de determinar su contribución a un estadio particular.

2.5 Concepto de error

La evolución del concepto de error a lo largo de los diferentes paradigmas ha resultado en una visión positiva sobre este, de gran valor para la investigación desde un punto de vista psicolingüístico; lo que, a su vez, sirve como fuente para apoyar el ámbito pedagógico (Vásquez 2009). En la formulación de Corder (1967), el error se define como una desvia-

ción sistemática que determina el sistema lingüístico de un aprendiz en una etapa concreta de aprendizaje, en contraste con aquellas desviaciones no sistemáticas: las faltas y lapsus. Las faltas comprenden las desviaciones no sistemáticas producidas en forma inconsciente y esporádica y de fácil corrección; mientras que el lapsus, también de ocurrencia no sistemática, tendría su origen en factores extralingüísticos como la falta de concentración o restricciones de memoria bajo la presión de la comunicación. Como es esperable, en ASL, así como en el desarrollo de este trabajo, las desviaciones de mayor interés investigativo son aquellas de naturaleza sistemática. No obstante, una de las críticas recurrentes al AE es su debilidad para distinguir entre errores y faltas (Quiñones 2009).

Complementariamente, para Ellis (1997), el error se entiende como una desviación de la norma estandarizada de la lengua meta,⁴ referida a los usos lingüísticos con los que los datos de la IL son comparados y contrastados. En este sentido, la norma puede adoptar una orientación descriptiva, relacionada con los usos considerados naturales por una comunidad de hablantes; o una orientación prescriptiva, relativa a las pautas que regulan el uso correcto de una variante considerada culta (Sánchez 2015b). En los estudios de la IL, la propuesta más actual integra ambos enfoques, en donde la norma descriptiva toma como base

⁴ Debido a las múltiples normas lingüísticas que conviven en el mundo hispanohablante, la tendencia en los estudios de la interlengua actuales es adoptar como variante estándar el referente lingüístico de carácter panhispánico (Vásquez 2009; Sánchez & Jiménez 2013). Esta opción también ha sido escogida en las fuentes de referencia, como son la Nueva Gramática de la Lengua Española (RAE & AALE 2009) y los corpus CREA (RAE & AAL 2018) y CORPES (RAE & AAL 2018).

la comparación de frecuencias de un corpus de aprendientes con uno de hablantes nativos (HN) de características similares;⁵ mientras que la norma prescriptiva, referida al aspecto de corrección, es definida eminentemente con base en el criterio valorativo de los HN (D'Aquino & Ribas 2001).

Respecto de la orientación prescriptiva, Sánchez y Jiménez (2013) proponen para el análisis de estructuras sintácticas de la IL —en donde podríamos situar este estudio— seguir la siguiente secuencia: (1) realizar un ejercicio de introspección para identificar las expresiones bajo examinación; (2) comprobar el uso que hacen los HN de estas unidades en un corpus control; (3) en caso que el corpus control no sea lo suficientemente amplio para contener una expresión dada, se debe consultar un corpus de referencia más extenso; y (4) si los datos obtenidos del corpus de referencia no permiten realizar un juicio de gramaticalidad, es recomendable consultar una fuente normativa estandarizada, como es la Nueva Gramática de la Lengua Española (NGLA) (RAE & AAL 2009).

2.6 Análisis de errores asistido por computador

La disponibilidad de recursos computacionales ha dado un nuevo giro al AE, sustentado en la rigurosidad metodológica de la LC. Sus procedimientos contemplan las etapas de Corder (1967; 1971) para la identificación,

⁵ Los resultados del análisis bajo una orientación descriptiva han sido presentados en un manuscrito anterior para su publicación.

descripción, clasificación y explicación del error.⁶ Lo anterior se realiza sometiendo los corpus a un proceso de anotación de acuerdo a criterios representativos de dichas etapas, los que son integrados al procesamiento de datos mediante esquemas taxonómicos específicos al fenómeno observado. Al respecto, una crítica frecuente al AE alude a las inconsistencias en los esquemas taxonómicos que mezclan niveles distintos de análisis, por ejemplo, los descriptivos y lingüísticos, o los criterios lingüístico y explicativo⁷ (Sánchez 2015d: 68).

2.7 Evidencia empírica vinculada a los errores de las ColGram v+p

Los antecedentes sobre la temática de las ColGram en la investigación de ELE son limitados. Oportus y Ferreira (2019; 2020) examinan la naturaleza del uso global de las ColGram v+p, con un énfasis en las diferencias de frecuencia de los usos correctos, entre aprendientes anglófonos (A2 y B1) y versus la producción de un corpus control de hablantes nativos. Estos trabajos concluyen que los aprendientes muestran una producción de naturalidad restringida debido al uso recurrente de algunas ColGram v+p en particular. Por su parte, Domínguez (2014) compara la producción de colocaciones léxicas versus las ColGram, como una categoría fraseológica general, en aprendientes de L1 valón y flamenco;

⁶ Los resultados del *análisis explicativo* serán presentados en un manuscrito distinto para su publicación.

⁷ Un ejemplo de ello, es la propuesta taxonómica de Vázquez (1992:31), considerada una de la más completas para el AE (Quiñones 2009), pero que no contempla con claridad la dimensión lingüística del error dentro de sus seis niveles taxonómicos.

indagando en la etiología de los errores, pero sin abordar el análisis de los niveles de localización, sintáctico-semántico y descriptivo. Adicionalmente, disponemos de algunos antecedentes sobre las ColGram de un grupo reducido de estudios en inglés como L2. Entre estos se reporta que la proporción de unidades correctas supera a las incorrectas (Bahardoust & Moeini 2012); los errores tienden a ser más numerosos en un contexto de enseñanza de lengua extranjera que en uno de segundas lenguas (Alsulayyi 2015); y que el mayor número de errores que afecta a las ColGram (sin considerar las ColGram v+p) ocurre en casos en que el colocativo es una preposición: sustantivo/adjetivo + preposición (Alotaibi & Alotaibi 2015; Alsulayyi 2015). En definitiva, el análisis sobre los errores se ha circunscrito mayormente al contexto del inglés como L2, con un enfoque global sobre distintos tipos de ColGram, del cual virtualmente se ha excluido el tipo v+p.

3. EL ESTUDIO

El objetivo general del estudio es determinar y comparar la frecuencia de errores en la producción de ColGram v+p según la competencia (niveles A2 y B1) acorde con los criterios de localización, sintáctico-semántico y descriptivo del error. Esta investigación es de tipo transversal y alcance descriptivo. Su metodología es mixta, con un enfoque cuantitativo basado en el índice de frecuencia, y cualitativo respecto de la identificación, clasificación y determinación de gramaticalidad de las unidades colocacionales.

3.1 Pregunta de investigación

¿Cómo difiere la frecuencia de errores en el uso de ColGram v+p de los aprendientes anglófonos acorde con la competencia en los criterios de localización, descriptivo y sintáctico-semántico del error?

3.2 Corpus de aprendientes y muestra

Los textos analizados corresponden a un subcorpus del Corpus Especializado de Aprendientes de Español como Lengua Extranjera CAELE (Ferreira 2014-2018) (Ferreira & Elejalde 2017).⁸ Este es un inventario abierto que cuenta con más de 750 textos redactados por aprendientes de diferentes orígenes lingüísticos, estudiantes de programas de intercambio de nivel de pregrado y posgrado. Las tareas fueron realizadas durante las clases de español y supervisadas por el docente a cargo. Estas corresponden a una primera instancia de escritura ingresada directamente en el bloc de notas; sin la ayuda del editor en línea ni diccionarios.

El subcorpus de aprendientes corresponde a la producción de 33 aprendientes anglófonos (19 del nivel A2, y 14 de B1). El nivel de competencia fue determinado previamente al ingreso de los cursos de español con la Prueba de Multinivel con Fines Específicos Académicos (Ferreira 2016), evaluación alineada con el Marco Común Europeo de Referencia (Consejo de Europa 2001). El subcorpus comprendió 207 textos

⁸ Corpus desarrollado en los proyectos FONDECYT N° 1140651 (Ferreira, 2014) y N° 1180974 (Ferreira, 2018).

(116 del nivel A2 y 91 de B1); con una longitud promedio de 193 palabras en el nivel A2, y 244 en B1. El total de palabras en el nivel A2 fue de 24790, y en el nivel B1 de 24774. Las temáticas de escritura se centraron en aspectos culturales y problemáticas actuales en la sociedad chilena y países de origen de los estudiantes, totalizando 20 tópicos; 16 de ellos temas comunes a los niveles A2 y B1. La motivación de escritura estuvo marcada por la función comunicativa, mientras que el género discursivo puede ser descrito como expositivo en los estilos descriptivo, narrativo y argumentativo.

3.3 Metodología de la investigación

El estudio se basa en el enfoque ACI (Granger 2015), y el modelo del AE asistido por el computador (Corder 1967; 1971). En cuanto a los procedimientos se adopta el modelo metodológico de Análisis de Corpus Electrónicos de Aprendientes de los proyectos FONDECYT N° 1140651 (Ferreira 2014) y N° 1180974 (Ferreira 2018), en lo que concierne a la recopilación, anotación y procesamiento de datos.

3.4 Concepto, identificación y gramaticalidad de las unidades colocacionales

El trabajo adopta una perspectiva fraseológica sobre el concepto de ColGram v+p; en donde el sintagma preposicional seleccionado por el verbo para completar su significado representa una combinación con un grado intermedio de fijación (Benson et al. 1986; Travalía

2006). Para facilitar la identificación de las unidades estudiadas, se elaboró su definición operacional como la combinación de un verbo pleno y una preposición (o locución preposicional) que introduce un CPREP no obligatorio. No obstante, debido a la forma en que las ColGram v+p se solapan con otros fenómenos lingüísticos, se identificaron aquellas secuencias verbo + preposición que no corresponden a casos colocacionales: la estructura interna en locuciones verbales y cuasilocuciones con verbo de apoyo; marcadores funcionales del complemento directo (CD), complemento indirecto (CI) y complemento agente; CPREP obligatorios; perífrasis; complementos argumentales de ubicación, complementos preposicionales adjuntos; oraciones copulativas; y secuencias verbo + locuciones adverbiales introducidas por preposición.

Debido a que no se contó con un programa computacional que pudiese detectar las ColGram v+p acorde con la conceptualización del estudio, la identificación y etiquetamiento de las ocurrencias se efectuó primero manualmente en un archivo .doc y luego en el software de gestión de corpus UAM (ver sección 1.7). Siguiendo a Ducrot & Todorov (1986: 153) y Sánchez (2015b: 180), este proceso se basó en la introspección del investigador responsable, HN de español. No obstante, para obtener mayor precisión en las unidades identificadas, se contó con la revisión externa de ocho especialistas en español. Del total de 207 textos del subcorpus de estudio, los especialistas revisaron 119 de ellos (57.5% del total) e indicaron si las unidades que se les señalaban eran gramaticalmente correctas o no. En caso de error, debían explicar el mecanismo de error y sugerir una posible corrección.

Adicionalmente, debido a que es común que un HN no posea el dominio completo de la lengua (Cervantes 2009), o que se presenten diferencias en el juicio de expertos, se complementó este procedimiento, siguiendo a Sánchez & Jiménez (2013), y Vincze (2015: 194), obteniendo datos de frecuencia mediante reportes de concordancias del corpus de referencia CORPES (RAE & AAL 2018). Este inventario es representativo de la norma panhispánica del sistema lingüístico actual y ayudó a realizar el juicio de gramaticalidad de aquellas combinaciones dudosas en su corrección. Adicionalmente, siguiendo las recomendaciones especializadas, se utilizó la NGLA (RAE & AAA 2009) como fuente normativa de la lengua en casos puntuales.

En definitiva, se consideraron correctas las secuencias que no presentaban anomalías combinatorias en el contexto comunicativo en que aparecían los enunciados. Por el contrario, las combinaciones incorrectas correspondían a secuencias que, con base en diferentes mecanismos (adición, falsa selección, omisión), resultaban agramaticales a razón que no concordaban con los patrones sintáctico-semánticos de la lengua.

A continuación, se presenta un ejemplo del proceso de identificación de las ColGram v+p (ejemplo 1). Los segmentos que contienen la unidad colocacional están destacados en fuente negrita, mientras que la anotación respecto de la gramaticalidad y detalle de la clasificación según los niveles taxonómicos aparece entre paréntesis cuadrados. Se muestra el parámetro de gramaticalidad: CORRECTO (1 y 2), o ERROR (3), y en este último caso se registraba el patrón sintáctico-semántico, la localización del error, y la descripción.

- (1) * “La literatura de EEUU tiene su basa en política y correspondencias al medio de los colonistas y Britancia. La verdad es que no (a) **se mucho sobre** [CORRECTO] la literatura de mi país, es una verguenza pequeña. Sí, hay peomas, leyendas, etc., pero no se como se lama y tampoco quien escribió. Me parece durante el naciamiento del EEUU hay un autor que (b) **escribió mucho, Benjamín Franklin, sobre políticos** [CORRECTO] y negocios, pero no (c) **se otros** [ERROR: V + (CD) + CPREP ALTERN SIGNIF CERCANO – COLOCATIF – OMI - INTRAL – NEUTR].” (S11A2T2)⁹

3.5 Delimitación del rango de errores

Debido a la naturaleza colocacional del estudio, los casos son analizados desde una perspectiva fraseológica, y, por tanto, de la combinatoria de la unidad léxica (el verbo) con el colocativo (la preposición) para expresar un significado dado. Al considerar que no existe un “estándar uniforme sobre lo que es correcto e incorrecto” (Sánchez 2015c: 202), se adoptó el enfoque de Howarth (1996 en Vincze 2015), el análisis se focalizó exclusivamente en los errores que rompían con el vínculo colocacional; ya fuese, debido a: (1) la adición de rasgos colocacionales mediante la creación de un verbo inexistente en español en base a un morfema conocido por el aprendiente, (2) la omisión o reemplazo del colocativo o (3) la adición de un rasgo colocacional mediante la creación de una combinación v+p de naturaleza colocacional inaceptable en la lengua

⁹ Fragmento extraído del corpus CAELE (Ferreira, 2014; 2018). El asterisco al inicio del extracto indica que este contiene agramaticalidades. El código entre paréntesis al final del extracto representa el identificador del texto.

mediante la adición o falsa selección de una preposición. De esta manera, se excluyeron los errores relativos a otros aspectos manifiestos en el contexto de los enunciados donde ocurría la ColGram v+p; por ejemplo, errores de voz, tiempo y modo verbal; formas finitas; concordancia sujeto-predicado y de género/número; uso de determinantes; ortografía literal y acentual; y léxico y registro.

3.6 Clasificación de errores y niveles taxonómicos

La propuesta taxonómica fue desarrollada luego de la revisión preliminar del subcorpus tomando como base: (1) el modelo metodológico de los Proyectos FONDECYT N° 1140651 (Ferreira 2014) y N° 1180974 (Ferreira 2018) en lo que respecta a los procedimientos y etiquetas. (2) el AE asistido por el computador (Ferreira & Lafleur 2015; Ferreira & Elejalde 2017) en lo que compete a las criterios y categorías de errores. (3) el AE de colocaciones léxicas (Orol & Alonso 2013; Vincze 2015), en lo que se refiere al criterio de localización. (4) los planteamientos teóricos sobre ColGram para ELE (Travalia 2006), y los CPrep del español (RAE & AAL 2009; Jiménez 2011; Vaisberg 2013), en lo que dice en relación con el criterio sintáctico-semántico.

La taxonomía, representada en la Figura 1, incluyó tres niveles correspondientes a los tres criterios de análisis de los errores. El primer nivel corresponde al criterio de localización que indica si el error se origina en el uso inadecuado del colocativo (la preposición), o si este corresponde a uno que afecta a la colocación completa. El segundo se refiere al nivel descriptivo respecto de la forma en que se manifiesta superficialmen-

te el error. Finalmente, el tercer nivel especifica la clasificación según el patrón sintáctico-semántico que describe las ColGram producidas. Este esquema taxonómico intentó incorporar categorías definidas y cerradas, con una delimitación tajante entre ellas. En la práctica, a partir de la evidencia lingüística examinada fue imposible plantear categorías totalmente cerradas, especialmente en el criterio sintáctico-semántico. Al respecto, Fellbaum (2007) destaca el papel de los análisis cuantitativos basados en la probabilidad estadística, como los utilizados en este estudio, en el mejoramiento de estas deficiencias.

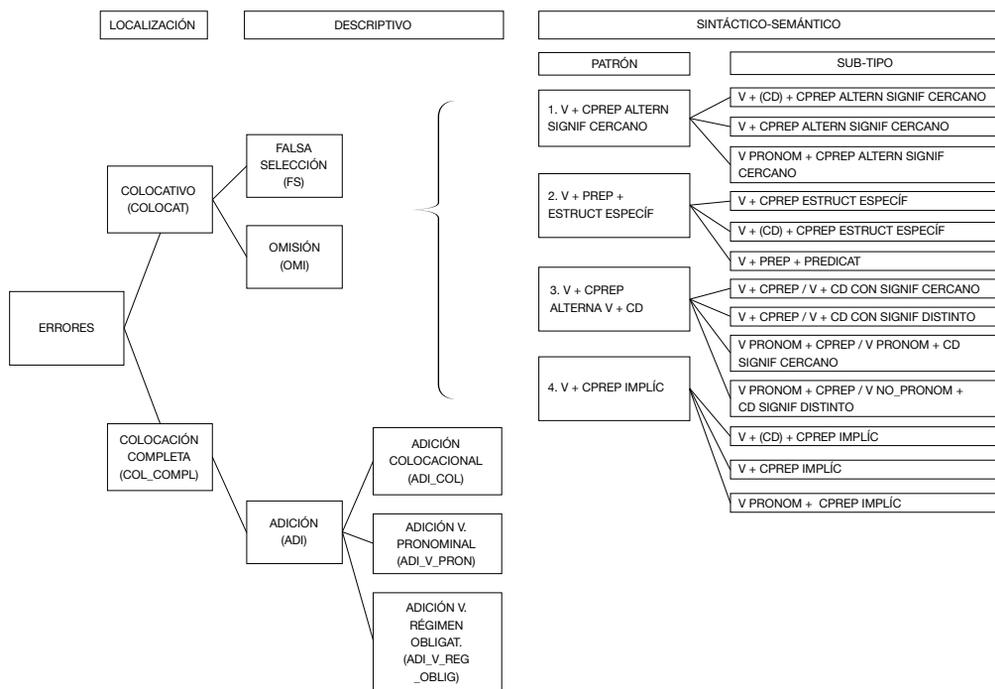


Figura 1. Esquema taxonómico de criterios de análisis de errores

3.6.1 Criterio de localización

- Error del colocativo (COLOCAT): deriva de un uso indebido del colocativo.

- (2) *se enojan cuando **reír sobre** su preciosa cueca. (S11A2POSDIF)
 ‘Se enojan cuando se ríen de su preciosa cueca’.

- Error de la colocación completa (COL_COMPL): corresponde a ColGram agramaticales debido a que se establece un vínculo inaceptable entre *verbo + preposición + término de preposición* para el español.

- (3) *a mi no suelo **escuchar** mucho **a** este tipo de música. (S23B1POSDIF)
 ‘No suelo escuchar mucho este tipo de música’.

3.6.2 Criterio descriptivo

- Error de falsa selección (FS): se utiliza una preposición inaceptable para el significado expresado en el contexto del enunciado, en vez de otra.

- (4) *¿que **opinas a** la música típica de Chile? (S18A2+POSIN)
 ‘¿Qué opinas de la música típica de Chile?’.

- Error de omisión (OMI): se omite la preposición exigida por el verbo en un contexto en que debería aparecer.

- (5) *yo también **viaje un** sector famoso del vinos de Chile. (S11A2T1)
 ‘Yo también viajé a un sector famoso de vinos en Chile’.

- Error de adición (ADI): enunciados que contienen secuencias verbo + preposición + término de preposición que constituyen combinaciones que se asemejan a una ColGram v+p, pero que son incorrectas o inexistentes en español puesto que el contenido semántico se expresa mediante otra función sintáctica o mecanismo o mediante la creación de un verbo inexistente en español. Se distinguen tres sub-tipos:

a) Adición colocacional (ADI_COL): corresponde a producciones en donde se inserta erróneamente una preposición a continuación de un verbo, véase ejemplo (6), o bien se crea un verbo inexistente en español, lo que produce una secuencia colocacional agramatical, véase ejemplo (7).

(6) *me **gusta a** trabajar con mis manos. (S14B1PRE)

‘Me gusta trabajar con mis manos’.

(7) *costumbristas chileans son **pintureando sobre** los temas.
(S15A2+T3)

‘Los pintores costumbristas chilenos pintan sobre los temas’.

b) Adición con verbo pronominal (ADI_V_PROM): error caracterizado por un empleo erróneo de un pronominal, lo que crea una secuencia colocacional inaceptable en español.

(8) *muchos paisajes diferentes que **se varían entre** montañas y desiertos. (S41POSDIF)

‘Muchos paisajes diferentes que varían entre montañas y desiertos’.

c) Adición con verbo de régimen obligatorio (ADI_V_REG_OBLIG): se produce una secuencia colocacional inaceptable modificando la preposición única exigida por un verbo de *régimen preposicional obligatorio*.

(9) *mucho de la economía **depende a** las turistas. (S48B1PRE)

‘Gran parte de la economía depende de los turistas’.

3.6.3 Criterio sintáctico semántico

Patrón 1. Verbo + 2 o más preposiciones alternantes con un significado cercano (V + CPREP ALTERN SIGNIF CERCANO).

- Subtipo 1.1 Verbo + (CD) + CPREP con preposiciones alternantes (V + (CD) + CPREP ALTERN SIGNIF CERCANO). Ej., hablar + sobre / de / acerca de.

- Subtipo 1.2 Verbo + CPREP con preposiciones alternantes (V + CPREP ALTERN SIGNIF CERCANO). Ej., ir + a / hacia / para / hasta.

- Subtipo 1.3 Verbo pronominal + CPREP con preposiciones alternantes (V PRONOM + CPREP ALTERN SIGNIF CERCANO). Ej., tratarse + de / sobre / acerca de.

Patrón 2. Verbo + 1 o más de 1 preposición + estructura sintáctica específica (V + PREP + ESTRUCT ESPECÍF).

- Subtipo 2.1 Verbo + CPREP con estructura sintáctica específica (V + CPREP ESTRUCT ESPECÍF).

(10) *y **aprendes a** aplicar esas ideas. (S17B1POSIN).

- Subtipo 2.2 Verbo + CD + CPREP con estructura sintáctica específica (V + (CD) + CPREP ESTRUCT ESPECÍF).

(11) para **apoyar a** los músicos chilenos **a difundir** mejor su música. (S14B1POSDIF).

- Subtipo 2.3 Verbo + sintagma preposicional predicativo del sujeto o del CD (V + PREP + PREDICAT).

(12) *me **llevaría** ropa y cerveza **a** recuerdo. (S25A2+T1)

‘Me llevaría ropa y cerveza de recuerdo’.

Patrón 3. Verbo + CPREP alternante con CD con diferencias semánticas leves o marcadas entre las alternancias (V + CPREP ALTERNA V + CD).

- Subtipo 3.1 Verbo no pronominal + CPREP que alterna con verbo no pronominal + CD con significados cercanos (V + CPREP / V + CD CON SIGNIF CERCANO). Ej., *acabar* puede indistintamente ser seguido de CD, o de un CPREP introducido por la preposición *con* manteniendo un significado similar: *acabar algo* versus *acabar con algo*.

- Subtipo 3.2 Verbo no pronominal + CPREP que alterna con verbo no pronominal + CD, con diferencias semánticas importantes (V + CPREP / V + CD CON SIGNIF DISTINTO). Ej., *creer* cuando aparece junto a un CPREP conlleva el significado de tener fe o creencia; mientras que al ser seguido de CD, su significado es tener a algo por veraz; por ejemplo, *creí la versión sin dudar*.

- Subtipo 3.3 Verbo pronominal + CPREP que alterna con verbo no pronominal + CD, con significados cercanos (V PRONOM + CPREP / V PRONOM + CD SIGNIF CERCANO). Ej., el verbo *encontrar* puede expresar un significado parecido si aparece seguido de un CD, o si en su uso pronominal selecciona un CPREP introducido por la preposición *con*.

- Subtipo 3.4 Verbo pronominal + CPREP que alterna con verbo no pronominal + CD, con diferencias semánticas importantes (V PRONOM + CPREP / V NO_PRONOM + CD SIGNIF DISTINTO). Ej., el verbo *enfocar* seguido de CD difiere en forma importante en su significado cuando este es usado pronominalmente seguido de CPREP con la preposición *en*.

Patrón 4. Verbo + CPREP que puede aparecer tácito o implícito (V + CPREP IMPLÍC).

- Subtipo 4.1 Verbo no pronominal + CD + CPREP que puede aparecer tácito (V + (CD) + CPREP IMPLÍC).

(13) obras lindas literarias que deberíamos **compartir con** el mundo. (S16B1PRE), comparado con el CPREP implícito: ‘obras lindas literarias que deberíamos compartir’.

- Subtipo 4.2 Verbo no pronominal + CPREP que puede aparecer tácito (V + CPREP IMPLÍC).

(14) *Mi preferencia de musica **cambia con** las emociones mas fuerte en mi vida. (S21A2+POSDIF), comparado con el CPREP implícito: ‘mi preferencia de música cambia en mi vida’.

- Subtipo 4.3 Verbo pronominal + CPREP que puede aparecer tácito (V PRONOM + CPREP IMPLÍC).

- (15) *Chilenos les gusta **divertirse** mucho **con** carrete y aventura. (S56A2+P1), comparado con el CPREP implícito: ‘a los chilenos les gusta divertirse mucho’.

3.7 Anotación del subcorpus

Las anotaciones del subcorpus se realizaron utilizando el software de distribución libre UAM Corpus Tool 3 (O'Donnell 2013). Esta herramienta permite registrar manualmente, de acuerdo a un esquema de anotación definido por el investigador, distintos atributos en diferentes niveles dentro de un mismo texto sin alterar su contenido original. Esto presenta la ventaja de poder editar el marco de etiquetas en pos de una mayor consistencia en la anotación. Así también, permite recuperar datos definiendo parámetros de búsqueda, y obtener informes estadísticos para comparar la relación de dependencia entre dos bases de datos nominales. Los textos fueron ingresados en formato .txt con codificación UTF-8, conjuntamente con los datos de competencia y origen lingüístico de los sujetos. Con esta herramienta, se implementó el esquema taxonómico descrito anteriormente el cual fue utilizado para etiquetar cada unidad identificada con los atributos correspondientes de gramaticalidad, y nivel de análisis en el caso de los errores. Una vez que se finalizó con la revisión de las anotaciones, se obtuvo reportes estadísticos y de frecuencia necesarios para responder a la pregunta de investigación.

3.8 *Análisis*

Siguiendo estudios sobre colocaciones léxicas (Orol & Alonso 2013; Vincze 2015), se calcularon porcentajes y proporciones de frecuencia y se utilizó la prueba *prop test* del programa R (R Core Team 2013) para determinar la igualdad entre estos (nivel alfa de 0.05). Por otra parte, se aplicó el estadístico de Chi Cuadrado para determinar la independencia entre las frecuencias de los dos niveles de competencia. Dicha prueba, incorporada al programa UAM Corpus, es utilizada con frecuencia en estudios lingüísticos e informa sobre la significancia estadística de las diferencias de frecuencia en los siguientes niveles: (1) +, leve; significativo en el nivel del 90%, 10% probabilidad de error, (2) ++, moderada; significativo en el 95%, 5% probabilidad de error, y (3) (+++), alta; significativo en el 98%, con 2% probabilidad de error (Choura 2017).

4. RESULTADOS Y DISCUSIÓN

Para responder a la pregunta de investigación, en primer lugar, se identificó el número de ColGram v+p correctas e incorrectas del total producido en cada nivel. Se contabilizaron como unidades distintas los verbos que alternaban con preposiciones sin producir cambios semánticos importantes; tales como, *hablar de algo / hablar sobre algo* (Zato 2011); interpretación también aplicada a casos diferenciados por el incremento reflexivo, ej., *ir a / irse a* (Martínez 1987: 85).

Específicamente, en el nivel A2 el total de ColGram v+p identificadas fue de 376; mientras que en el nivel B1 fue de 403. Enseguida, se contabilizaron los usos acertados e incorrectos del total de ocurrencias en cada nivel. El número de usos correctos correspondió a 217 (57.7%) en A2, y 264 (65.5%) en B1; mientras que el número de colocaciones incorrectas fue de 159 (42.3%) en A2, y 139 (34.5%) en B1, como se resume en la Tabla 1. Al comparar las proporciones de usos correctos versus los incorrectos en cada grupo, se obtuvo una diferencia estadística significativa dentro de cada subcorpus (A2: $z=4.303$, $p=0.000$; B1: $z=8.806$, $p=0.000$). Esto va en línea con los hallazgos sobre colocaciones léxicas en L2 inglés de Bahardoust & Moeini (2012), en donde la proporción de unidades correctas supera los errores.

Tabla 1. Total de ColGram v+p correctas e incorrectas según subcorpus

Índice	A2	B1
Tamaño sub-corpus (núm. de palabras)	24790	24774
Núm. de colocaciones correctas	217	264
Núm. usos colocaciones incorrectas	159	139
Núm. total de ColGram v+p	376	403

Posteriormente, al hacer la comparación entre subcorpus respecto de los usos colocacionales adecuados e incorrectos, se pudo observar que existían diferencias estadísticas significativas entre los dos niveles de competencia; como se observa en la Tabla 2. Los usos correctos del nivel B1 superan a los del nivel A2, y consecuentemente, los usos incorrectos de A2 son más numerosos que en el nivel B1. Lo anterior sería espera-

ble en términos de que el avance en la competencia debería ir acompañado de una mayor precisión y una disminución del error en el uso de estas unidades.

Tabla 2. Comparación de usos correctos e incorrectos según subcorpus

	A2 (N=376)		B1 (N=403)		Chisqu	Signif.
	Frecuencia	%	Frecuencia	%		
Usos correctos	217	57.7	264	65.5	5.01	++
Usos incorrectos	159	42.3	139	34.5	4.68	++

4.1 Usos incorrectos acordes con la localización del error

Con el objetivo de responder a la pregunta sobre los errores a lo largo de los tres criterios, en un primer análisis, se recabaron los datos de localización del error respecto de si este aparece como un uso inadecuado del colocativo, o si es un error asociado a la unidad colocacional completa, como ilustra el Gráfico 1. Al considerar los datos conjuntos de ambos niveles, se obtuvo que, de un total de 298 errores, 104 derivan del mal uso del colocativo, los que corresponden a los errores relativos a los cuatro patrones sintáctico-semánticos identificados en la taxonomía. Por otra parte, 194 usos incorrectos constituyen errores de mayor envergadura que comprometen la gramaticalidad de toda la combinación colocacional debido a que corresponden a secuencias que muestran superficialmente la forma de combinaciones colocacionales v+p, pero que resultan agramaticales en español. Con los datos anteriores, se obtiene una diferencia estadísticamente significativa al comparar las proporciones

de las frecuencias de los errores del colocativo, y de la colocación completa ($z=7.303$; $p=0.000$), lo que arroja fehacientemente que los errores de la colocación completa son los que muestran la mayor frecuencia estadística y por ende los errores más numerosos de los aprendientes.

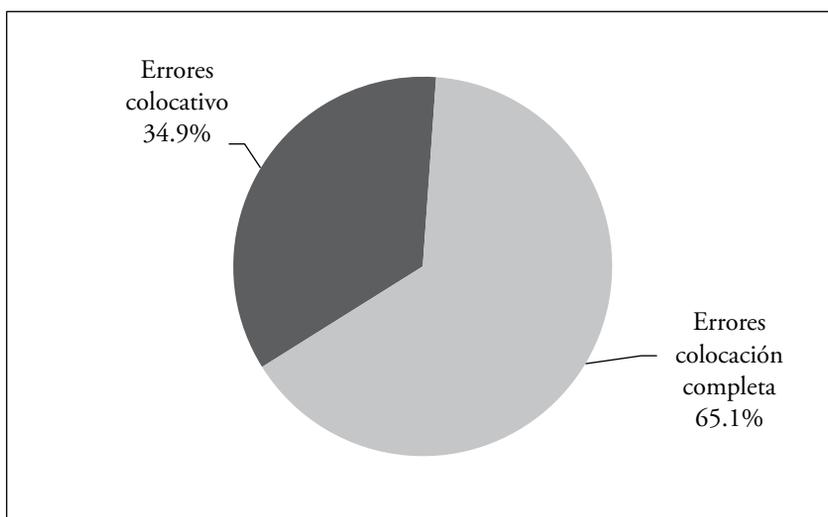


Gráfico 1. Porcentaje de errores según la *localización* del error

Respecto de dichos resultados, es difícil establecer un contraste con estudios sobre colocaciones léxicas; por ejemplo, Vincze (2015); puesto que, de acuerdo a la naturaleza léxica de los dos constituyentes que forman las colocaciones léxicas, se pueden distinguir claramente 3 tipos de errores: los de la base, del colocativo y de la colocación completa; lo que no es aplicable a las ColGram. Sin entrar en mayor detalle en estas distinciones, la proporción de errores que afecta al colocativo en el caso de Vincze (2015) es mucho más alta en su caso (51.6%) que

en este estudio (35.0%). Algo similar ocurre también con la colocación completa, en Vinzce (2015) la proporción de frecuencia equivale al 13.1%, porcentaje considerablemente menor que el de 65% alcanzado en este estudio.

Posteriormente, al comparar la localización del error según la competencia, aparecen 55 errores en el uso del colocativo en el nivel A2 en la Tabla 3, los que superan la frecuencia de 49 errores del nivel B1; mientras que, en el caso de los errores de la colocación completa, la frecuencia de 103 del nivel A2 también sobrepasa al nivel B1 con 90 errores. A pesar de esta diferencia numérica, no se obtienen diferencias de significancia estadística entre ambos niveles de competencia, ya sea en los errores que afectan al colocativo o a la colocación completa. Es decir, aunque los aprendientes del nivel A2 cometieron un número levemente superior de errores que el nivel B1, esta diferencia parecería explicarse por el azar. Los datos señalados nos llevarían a concluir que tanto los sujetos del nivel A2 y B1 tienden a producir proporcionalmente una cantidad similar de errores que afectan al colocativo y a la colocación completa.

Tabla 3. Frecuencia de errores según localización del error y nivel de competencia

Localización del error	A2 (N=376)		B1 (N=403)		Chisqu	Signif.
	Frecuencia	%	Frecuencia	%		
Colocativo	55	14.6	49	12.2	1.38	
Colocación Completa	104	27.7	90	22.3	1.01	
Total errores	159		139			

4.2 Usos incorrectos acorde el nivel descriptivo

En el nivel descriptivo del error distinguimos tres tipos de mecanismos formales de error: (1) falsa selección (FS), (2) omisión (OMI), y (3) de adición (ADI). En un primer análisis, se comparó la frecuencia de estos errores en forma global, como se presenta en el Gráfico 2. Del total de 298 errores producidos, la mayor frecuencia se presentó en los errores de ADI con 194 errores; luego, con una frecuencia de 61, se ubicaron los errores de FS y, en tercer lugar, con la menor frecuencia aparecen los errores de OMI con 43 ocurrencias. Al comparar las proporciones de frecuencia entre sí, se obtuvo una diferencia estadística significativa entre los errores de ADI y los de FS ($z=10.948$; $p=0.000$), así como entre los de ADI y los de OMI ($z=12.577$; $p=0.000$). Por otra parte, entre las proporciones de los errores de FS y OMI, no se encontraron diferencias estadísticas significativas. Estos datos apuntan a que en un nivel global los mecanismos de FS y OMI son proporcionalmente iguales entre sí y que su frecuencia es altamente superada por los errores de ADI cuyos procesos específicos implican mecanismos de adición o modificación de algún aspecto, ya sea, del verbo o el colocativo.

Posteriormente, analizamos el mecanismo de error según la competencia en la Tabla 4. Dentro de los errores de ADI, con las mayores frecuencias globales, aparecen 104 errores en el nivel A2, y 90 en B1. Descendiendo, en segundo lugar, los errores de FS incluyeron 34 errores en el nivel A2, y 27 en B1. Luego, con la menor frecuencia, aparecen los errores de OMI distribuidos en 21 errores en el grupo A2, y 22 en B1. En ninguno de los tres tipos de errores bajo el nivel descriptivo se obtuvo diferencia estadística significativa entre el nivel A2 y B1. De acuerdo

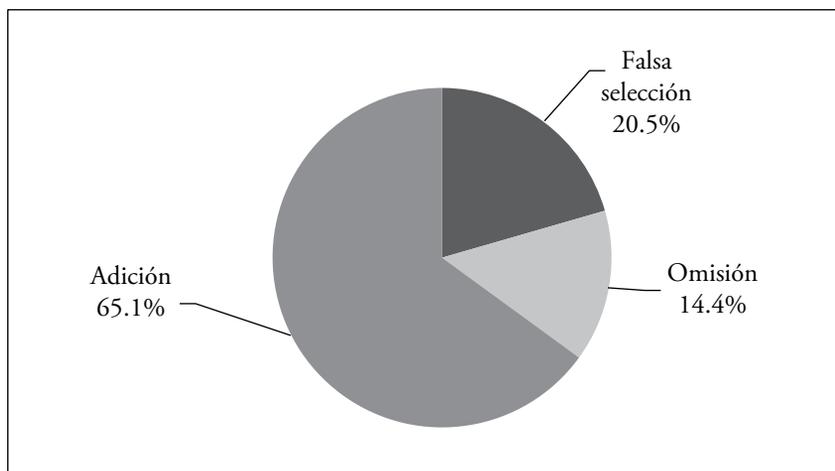


Gráfico 2. Porcentajes de frecuencia de errores según criterio descriptivo

a lo anterior, la frecuencia en orden descendiente de los tres mecanismos de error se presenta en ambos grupos de manera similar, observándose también una distribución similar entre los niveles A2 y B1.

Tabla 4. Frecuencia de errores según el criterio descriptivo y la competencia

Descripción del error	A2 (N=376)		B1 (N=403)		Chisqu	Signif.
	Frecuencia	%	Frecuencia	%		
Falsa Selección	34	9.0	27	6.7	0.01	
Omisión	21	5.6	22	5.5	0.03	
Adición	104	27.7	90	22.3	2.67	
Total errores	159		139			

Adicionalmente, bajo el criterio descriptivo, también se examinó la frecuencia global de errores de ADI en relación con los tres subtipos espe-

cíficos identificados en esta categoría: (1) Adición Colocacional, 155 errores; (2) Adición V. Pronominal, 22 errores; y (3) Adición V. de Régimen Obligat, 17 ocurrencias, como muestra el Gráfico 3. Al comparar las proporciones entre los subtipos, se obtuvo una diferencia estadísticamente significativa entre los errores de Adición Colocacional versus los de Adición V. Pronominal ($z=1.348$; $p=0.000$), y entre los errores de Adición Colocacional versus los de Adición V. Régimen Obligat. ($z=1.403$; $p=0.000$). Por otra parte, no se observó una diferencia de significancia estadística entre las proporciones de los errores de Adición V. Régimen Obligat., y los errores de Adición V. Pronominal. De acuerdo a estos datos, la frecuencia global de errores de Adición Colocacional supe-

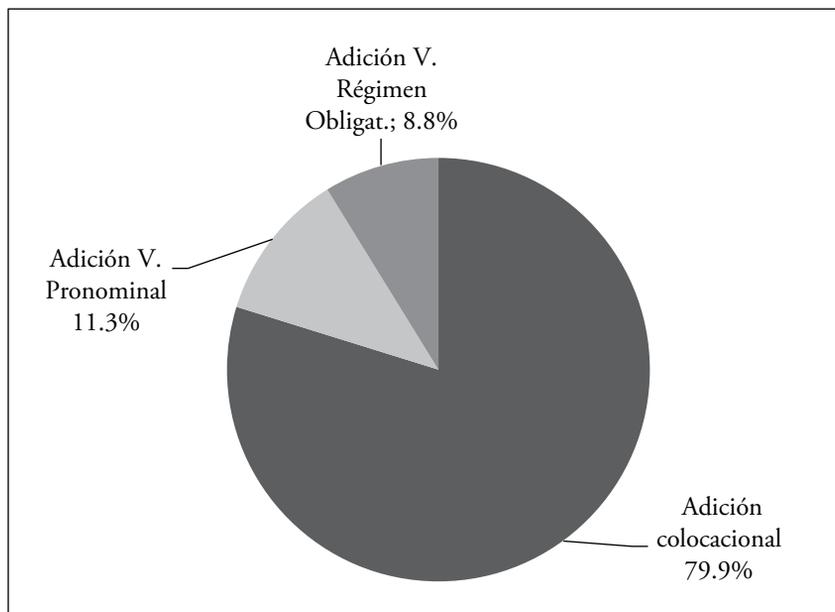


Gráfico 3. Porcentaje de frecuencia de errores de adición según subtipo

ra en forma clara la frecuencia de los errores de Adición V. Pronominal, y los de Adición V. de Régimen Obligat.; mostrando estos dos últimos una frecuencia inferior y estadísticamente similar entre sí.

Enseguida, al analizar específicamente los errores de ADI acorde la competencia, obtenemos que los errores de Adición Colocacional son los que alcanzan las mayores frecuencias en ambos niveles, con 78 ocurrencias en el nivel A2 y 77 en B1, como se presenta en la Tabla 5. Con menores frecuencias se ubican los errores de Adición V. Pronominal, con 13 errores en el nivel A2 y 9 errores en el nivel B1, y los errores de Adición V. Régimen Obligat., con 13 errores en el nivel A2, y solo 4 errores en el nivel B1. Este último subtipo es el único que presenta diferencia de alta significancia estadística entre ambos niveles. Lo anterior sería explicable debido a que la producción de complementos de régimen obligatorio implica el manejo de unidades fijas de la lengua, en donde los aprendientes de menor nivel mostrarían un menor repertorio de unidades internalizadas, lo que se traduciría en errores colocaciones durante sus intentos comunicativos.

Tabla 5. Frecuencia de subtipos errores de adición según competencia

Subtipo error de adición	A2 (N=376)		B1 (N=403)		Chisqu	Signif.
	Frecuencia	%	Frecuencia	%		
De adición colocacional	78	20.7	77	19.1	0.23	
De uso de pronominal	13	3.5	9	2.2	1.06	
Con verbos de régimen obligatorio	13	3.5	4	1.0	5.54	+++
Total errores	104		90			

4.3 Usos incorrectos acorde el nivel sintáctico-semántico

A continuación, nos focalizamos exclusivamente en los errores de FS y OMI, como se ilustra en la Tabla 6, para describir su frecuencia según el patrón sintáctico-semántico. En el análisis global se obtuvo diferencias con significancia estadística entre las proporciones del patrón 1 y los patrones 2 y 3; alcanzando el patrón 1 la mayor frecuencia de errores (43.3%), y superando a los patrones 2 ($z=4.244$; $p=0.000$), y 3 ($z=5.693$; $p=0.000$). Luego, le sigue el patrón 4 con la segunda mayor frecuencia (31.7%), siendo mayor que el patrón 2 ($z=2.596$ $p=0.009$), y que el patrón 3 ($z=4.145$; $p=0.000$). Finalmente se ubican los patrones 2 (16.3%) y 3 (8.7%), los que no muestran diferencia estadística entre ellos.

Tabla 6. Frecuencia de errores FS y OMI según patrón sintáctico-semántico

Patrón sintáctico semántico	Frecuencia	%
Patrón 1	45	43.3
Patrón 2	17	16.3
Patrón 3	9	8.7
Patrón 4	33	31.7
Total	104	100

Estos hallazgos tendrían relación con la naturaleza clasificatoria de la tipología de patrones sintáctico-semánticos empleada. Vale decir, que en los patrones 1 y 4 se categorizan algunos verbos léxicos de uso frecuente en la lengua (Aghmiri 2014). Por ejemplo, en el patrón 1 se incluirían colocaciones como *ir/ venir del/desde... a/hasta/para, ver en/por*

(medio de comunicación); *hablar/opinar de/sobre*; etcétera. Así también en el patrón 4, se clasifican colocaciones frecuentes tales como: *parecerse en algo a*; *ir por*; *usar/utilizarse para algo*; *prepararse para*; *comparar(se) con*; *trabajar para (alguien)*; *trabajar por (algo)*, etc. Por otra parte, en los patrones 2 (V + PREP + ESTRUCT ESPECÍF) y 3 (V + CPREP ALTERNA V + CD), se incluyen colocaciones bajo un marco más restrictivo desde la perspectiva de frecuencia léxica. Por ello, parecería existir una relación entre la frecuencia de los verbos colocacionales y la frecuencia de errores de OMI y FS. Es decir, aquellos patrones de mayor frecuencia, que serían también los más usados, serían también los más susceptibles a mostrar errores. En este sentido, podría ser que la baja frecuencia en los patrones 2 y 3 pueda ser explicado por estrategias de evitación o porque no eran necesarios en la comunicación, no necesariamente por una mayor precisión.

Posteriormente, se analizaron globalmente los errores de FS y OMI a lo largo de los cuatro patrones según la competencia, como se presenta en la Tabla 7. Se puede observar que no existían diferencias estadísticas significativas entre las frecuencias de los niveles A2 y B1 en ninguno de los cuatro patrones sintáctico-semánticos. De acuerdo a lo anterior, los errores de FS y OMI tenderían a distribuirse en cada patrón en forma equitativa en ambos niveles de competencia.

Luego comparamos separadamente las frecuencias de los errores de FS de los de OMI acorde con el patrón sintáctico-semántico. En este análisis más fino, como presenta la Tabla 8, en el caso de los errores de FS no se observan diferencias entre los dos niveles de competencia en los patrones sintáctico-semánticos, exceptuando el patrón 4, el cual muestra una

Tabla 7. Frecuencia de errores de FS y OMI según patrón y nivel de competencia

Patrón sintáctico semántico	A2 (N=376)		B1 (N=403)		Chisqu	Signif.
	Frecuencia	%	Frecuencia	%		
Patrón 1	23	6.1	22	5.5	0.69	
Patrón 2	8	2.1	9	2.2	0.01	
Patrón 3	5	1.3	4	1.0	0.19	
Patrón 4	19	5.1	14	3.5	1.12	
Total	55		49			

Tabla 8. Frecuencia de errores de FS según patrón y nivel de competencia

Patrón sintáctico semántico	A2 (N=376)		B1 (N=403)		Chisqu	Signif.
	Frecuencia	%	Frecuencia	%		
Patrón 1	10	2.7	14	3.5	0.43	
Patrón 2	4	1.1	4	1.0	0.01	
Patrón 3	5	1.3	3	0.7	0.41	
Patrón 4	15	4.0	6	1.7	4.39	++
Total	34		27			

diferencia estadística moderada entre los niveles A2 y B1, con una frecuencia mayor en el nivel A2.

Enseguida, al analizar solo los errores de OMI en la Tabla 9, no se obtienen diferencias estadísticas de significancia entre las frecuencias de los niveles A2 y B1. En definitiva, considerando los datos de los errores de FS y OMI por separado, se observa que estos se distribuyen en forma

Tabla 9. Frecuencia de errores de OMI según patrón y nivel de competencia

Patrón sintáctico semántico	A2 (N=376)		B1 (N=403)		Chisqu	Signif.
	Frecuencia	%	Frecuencia	%		
Patrón 1	13	3.5	8	2.0	1.60	
Patrón 2	4	1.1	5	1.2	0.06	
Patrón 3	0	0.0	1	0.2	0.93	
Patrón 4	4	1.1	8	2.0	1.09	
Total	21		22			

similar entre ambos niveles, siendo la única excepción los errores de FS en el patrón 4, en donde el nivel A2 comete más errores que el nivel B1.

5 CONCLUSIONES

El estudio presentado responde a la necesidad de examinar las ColGram v+p en la escritura en ELE, específicamente en lo que se refiere al análisis de errores y a cómo se presentan en los dos niveles de competencia examinados (A2 y B1), respecto de su descripción formal y lingüística. Para ello, nos situamos en el ámbito de la fraseología, examinando estas combinaciones caracterizadas por grados intermedios de fijación y de transparencia semántica. Bajo este marco conceptual se implementó un esquema taxonómico que incluyó tres criterios de análisis: localización, descripción y el patrón sintáctico-semántico que caracterizan a estos errores. Mediante esta taxonomía y la aplicación de los modelos del ACI y del

AE se analizó la frecuencia colocacional, siguiendo la línea de investigación de Corpus Electrónicos de Aprendientes.

Desde una perspectiva prescriptiva; es decir, distinguiendo entre usos correctos y errores, se observa como un rasgo común a los dos niveles observados, que estos producen una mayor frecuencia de aciertos colocacionales que de errores, coincidiendo con los estudios sobre colocaciones léxicas en L2 inglés. No obstante, la tasa relativamente alta de errores de los aprendientes, que alcanza el 42.3% (A2) y 34.5% (B1) de las ocurrencias, aparece como un rasgo distintivo de su producción escrita. Al respecto, la competencia ayudaría a explicar la disminución de estos errores en el nivel B1. Sin desmedro de la anterior, la alta frecuencia relativa de errores en el nivel B1 mostraría que el manejo productivo de las Col-Gram v+p es una problemática lingüística importante de la IL colectiva de los estudiantes anglófonos de ambos niveles de competencia.

En el análisis de errores bajo los 3 niveles taxonómicos, en el criterio de localización destaca que los errores que afectan a la colocación completa alcanzan mayor frecuencia que los errores del colocativo; y que ambos tipos de errores se distribuyen en forma similar entre los dos niveles de competencia.

Bajo el criterio descriptivo de los errores, de los 3 mecanismos principales identificados, se observó que los errores más frecuentes son los de ADI, los que superan ampliamente a los de OMI y FS; estos últimos con la misma frecuencia relativa a nivel global. Luego, considerando los tres subtipos de errores de ADI, el que concentró la mayor frecuencia fue el de adición colocacional superando ampliamente a los de adición con pronominal o adición con verbo de régimen obligatorio. Debido a que los

errores de adición implican el cruce de diferentes funciones o aspectos lingüísticos, se puede deducir que estos representan una gran dificultad para los aprendientes, más que el uso del colocativo correcto; por lo que su mejoramiento pasaría por un enfoque de enseñanza que debiese integrar el uso colocacional con los otros aspectos lingüísticos relacionados.

Un aspecto interesante bajo el criterio descriptivo es que la frecuencia de todos los tipos y subtipos de errores se distribuyen en la misma proporción en ambos niveles de competencia, con la excepción del subtipo Adición V. Régimen Obligat., el que disminuye su frecuencia en el nivel B1. De acuerdo a lo anterior, exceptuando este último caso, no es posible advertir un descenso estadístico a lo largo de los tipos y subtipos de mecanismos formales en los que se presentan los errores en el nivel B1. A razón de ello, se presenta como un punto interesante de ser investigado el determinar cómo procede la reducción de errores en estos tipos en niveles más altos, superiores al B1.

Finalmente, en el análisis bajo el criterio sintáctico-semántico de los errores del colocativo, aparece como un primer hallazgo que los errores se concentran en los patrones 1 (V + CPREP ALTERN SIGNIF CERCANO) y 4 (V + CPREP IMPLÍC), coincidiendo con los patrones que incluyen potencialmente a verbos de mayor frecuencia léxica; y que por el contrario, la cifra de errores es menor en los patrones 2 y 3, los que corresponden a estructuras más complejas y que incluyen potencialmente a verbos menos frecuentes. En segundo lugar, no aparecen diferencias de frecuencia a lo largo de los patrones entre los dos niveles de competencia, ya sean estos, errores de FS o de OMI, exceptuando los errores de FS con una reducción de frecuencia en el patrón 4 en el nivel B1.

Como se desarrolló anteriormente, el foco de este estudio recae en el ámbito descriptivo del error (criterios de localización y descriptivo) y en el aspecto lingüístico (criterio sintáctico-semántico). Por consiguiente, el análisis etiológico, en torno a los mecanismos psicolingüísticos que los originan, así como el análisis cualitativo e integrado sobre los errores más recurrentes en la producción de ColGram v+p fueron excluidos de este trabajo y quedan como una materia pendiente a ser divulgada en un trabajo posterior.

FUENTES

- Ferreira, Anita. 2014. *Proyecto de investigación FONDECYT – CONICYT N°1140651: El feedback correctivo escrito directo e indirecto en la adquisición y aprendizaje del español como lengua extranjera*. Universidad de Concepción, Chile 2014–2017.
- Ferreira, Anita. 2016. *Prueba de multinivel con fines específicos académicos*. Universidad de Concepción, Chile.
- Ferreira, Anita. 2018. *Proyecto de investigación FONDECYT – CONICYT N°1180974: Diseño e implementación de un corpus escrito de aprendientes de ELE en formato computacional para el análisis de la interlengua*. Universidad de Concepción, Chile.
- Ferreira, Anita. 2014–2018. *Corpus de aprendices de español lengua extranjera, CAELE*, Universidad de Concepción, Chile.
- Real Academia Española & Asociación de Academias de la Lengua. 2018. Banco de datos CREA. Corpus de referencia del español actual.

Real Academia Española & Asociación de Academias de la Lengua 2018.
Banco de datos CORPES XXI. Corpus del español del siglo XXI.

REFERENCIAS

- Aghmiri, Rachid. 2014. *Los usos de los 10 verbos más frecuentes del español comparados con el amazigh*. Girona: Universidad de Girona. (Tesis de máster).
- Alexopoulou, Angélica. 2010. Errores intralinguales e interlinguales en la interlengua escrita de aprendientes griegos de ELE. Madrid: Iberoamericana. (*Actas del XVI Congreso de la Asociación Internacional de Hispanistas*).
- Alonso, Margarita. 2010. No importa si la llamas o no colocación, descríbela. En Mellado, Carmen & Buján, Patricia & Herrero, Claudia & Iglesias, Nelly & Mansilla, Ana (eds.), *La fraseografía del s. XXI: Nuevas propuestas para el español y el alemán*, 55–80. Berlín: Frank & Timme.
- Alotaibi, Abdullah & Alotaibi, Maye. 2015. The Comprehension and production of English grammatical collocations by Kuwaiti EFL learners. *International Journal of English Language Teaching* 13(3). 26–39.
- Alsulayyi, Marzoug. 2015. The use of grammatical collocations by advanced Saudi EFL learners in the UK and KSA. *International Journal of English Linguistics* 5(1). 32–43.
- Bahardoust, Maryam & Moeini, Mohammad. 2012. Lexical and grammatical collocations in writing production of EFL learners. *The Journal of Applied Linguistics* 5(1). 61–86.

- Benson, Morton & Benson, Evelyn & Ilson, Robert. 1986. *BB1 combinatory dictionary of English. A guide to word combinations*. Amsterdam: John Benjamins.
- Bolaños, Sergio. 2015. La lingüística de corpus: Perspectivas para la investigación lingüística contemporánea. *Forma y Función* 28(1). 31–54.
- Buyse, Kris. 2003. Colocaciones léxicas: pistas y trampas. *Mosaico* 10. 10–17.
- Cervantes, Instituto. 2009. *Guía práctica del español correcto*. Madrid: Espasa.
- Cheikh-Khamis, Fátima. 2013. *Lingüística cognitiva aplicada a la enseñanza del léxico en segundas lenguas. Las colocaciones en clase de ELE*. Madrid: UNED. (Trabajo fin de máster).
- Choura, Sabiha. 2017. Ditransitive complementation per clause type in academic research article: Medical science and sociology as case studies. *IOSR Journal of Humanities and Social Science* 22(12). 19–29.
- Consejo de Europa. 2001. *Marco común europeo de referencia para las lenguas: Aprendizaje, enseñanza, evaluación*. Estrasburgo: Consejo de Europa.
- Corder, Pit. 1967. The significance of learners' errors. *International Review of Applied Linguistics* 5(4). 161–170.
- Corder, Pit. 1971. Idiosyncratic dialects and error analysis. *International Review of Applied Linguistics* 9(2). 147–160.
- D'Aquino, Alessandra & Ribas, Rosa. 2001. Interferencias morfosintácticas entre italiano y español L2: Un estudio empírico. *Atti del XIX Convegno Associazione IspanistiItaliani* 2. 119–130.

- Domínguez, Uriel. 2014. Las colocaciones en un corpus de aprendices valón y flamenco. En Morimoto, Yuko; Pavón, María & Santamaría, Rocío (eds.), *La enseñanza de ELE centrada en el alumno*, 977–986. Asociación para la Enseñanza del Español como Lengua Extranjera. (Presentación en Congreso Internacional, Getafe).
- Ducrot, Oswald & Todorov, Tzvetan. 1986. *Nuevo diccionario enciclopédico de las ciencias del lenguaje*. México: Siglo XXI Editores.
- Elejalde, Jessica & Ferreira, Anita. 2016. Errores de transferencia en comunidades de aprendizaje en línea por aprendientes de español como lengua extranjera (ELE). *Trabalhos em Lingüística* 55(3). 619–650.
- Ellis, Rod. 1997. *Second language acquisition*. Oxford: Oxford University Press.
- Fellbaum, Christiane. 2007. *Collocations and idioms: corpus-based linguistics and lexicographic studies*. Birmingham: Continuum Press.
- Fernández, Gisele. 2014. Enseñanza-aprendizaje de las colocaciones en el nivel inicial (A1-A2). *Marcoele* 19. 1–19.
- Ferreira, Anita & Elejalde, Jessica. 2017. Análisis de errores recurrentes en un Corpus de aprendices de español como lengua extranjera (Corpus CAELE). *Revista Brasileira de Lingüística Aplicada* 17(3). 509–538.
- Ferreira, Anita; Elejalde, Jessica & Vine, Anita. 2014. Análisis de errores asistido por computador basado en un corpus de aprendientes de español como lengua extranjera. *Signos* 46(86). 385–411.
- Ferreira, Anita & Lafleur, Naum. 2015. Analyse et description des erreurs prépositionnelles les plus fréquentes en espagnol L2. *Lingüística y Literatura* 68. 57–79.

- Ferreira, Anita & Oportus, René. 2018. Procesamiento cognitivo del feedback correctivo escrito indirecto en los errores preposicionales en ELE. *Boletín de Filología* 53(1). 83–108.
- Ferreira, Anita; Oportus, René & Fuentes, Karina. 2016. Tiempos de respuesta y feedback correctivo escrito en aprendientes de ELE. *Revue Française de Linguistique Appliquée* 22(2). 109–122.
- Granger, Sylviane. 1998. The computer learner corpus: a versatile new source of data for SLA research. En Granger, Sylviane (ed.), *Learner English on computer*, 3–18. Londres: Longman.
- Granger, Sylviane. 2002. A bird's-eye view of learner corpus research. En S. Granger, Sylviane & Hung, Joseph & Petch-Tyson, Stephanie (eds.). *Computer corpora, second language acquisition and foreign language teaching*, 3–33. Ámsterdam: Benjamins.
- Granger, Sylviane. 2004. Computer learner corpus research: Current status and future prospects. *Language and Computers* 52(1). 123–145.
- Granger, Sylviane. 2015. Contrastive interlanguage analysis: A reappraisal, *International Journal of Learner Corpus Research* 1(1). 7–24.
- Gyllstad, Henrik. 2007. *Testing English collocations. Developing receptive tests for use with advanced Swedish learners*. Lund: Lunds universitet. (Tesis doctoral).
- Gyllstad, Henrik & Wolter Brent. 2015. Collocational processing in light of the phraseological continuum model: Does semantic transparency matter? *Language Learning* 66(2). 296–323.
- Henriksen, Birgit. 2013. Research on L2 learners' collocational competence and development – a progress report. En Bardel, Camilla; Lindquist, Christine & Laufer, Batia (eds.), *L2 vocabulary acquisition*,

- knowledge and use. New perspectives on assessment and corpus analysis*, 29-57. Ámsterdam: Eurosla.
- Howarth, Peter. 1996. *Phraseology in English Academic Writing*. Tübingen: Max Niemeyer Verlag.
- Ishikawa, Shin'ichiro. 2013. The ICNALE and sophisticated contrastive interlanguage analysis of Asian learners of English. En Shin'ichiro Ishikawa (ed.), *Learner corpus studies in Asia and the world*, 91-118. Kobe: Kobe University.
- Jiménez, Laura. 2011. El estatus argumental del complemento de régimen prepositivo en español. *Signos Lingüísticos* 7(14). 81-114.
- López Noguira, Lucía. 2014. *La interlingua de aprendices suecos de español: Uso, desuso y sobreuso de la preposición de*. Estocolmo: Stockholm University. (Tesis de máster).
- Lozano, Cristobal & Mendikoetxea, Amaya. 2013. Learner corpora and SLA: The design and collection of CEDEL2. En Díaz-Negrillo, Ana; Ballier, Nicolas & Thompson, Paul (eds.), *Automatic Treatment and Analysis of Learner Corpus Data*, 65-100. Ámsterdam: John Benjamins.
- Lu, Lo. 2016. Colocaciones usadas en las pruebas de lectura DELE B1. *Círculo de Lingüística Aplicada a la Comunicación* 66. 210-243.
- Martínez, Hortensia. 1987. Sobre la rección y el régimen preposicional. *Archivum. Revista de la Facultad de Filología*. 75-87.
- Moreno, María. 2009. *Recopilación, desarrollo pedagógico y evaluación de un banco de colocaciones frecuentes de lengua inglesa a través de la lingüística de corpus y computacional*. Granada: Universidad de Granada. (Tesis doctoral).

- Nesselhauf, Nadja. 2005. *Collocations in a learner corpus*. Amsterdam: John Benjamins.
- O'Donnell, Mick. 2013. *UAM CorpusTool 3.0 tutorial*. Madrid: Universidad Autónoma de Madrid.
- Oportus, René & Ferreira, Anita. 2019. Colocaciones gramaticales verbo + preposición en ELE: aspectos de naturalidad en aprendientes anglófonos de nivel A2 y B1. *Trabalhos em Linguística Aplicada* 58(2) 826–858.
- Oportus, René & Ferreira, Anita. 2020. Colocaciones gramaticales verbo + preposición en ELE: análisis contrastivo de la interlengua en aprendientes de nivel A2 y B1. *Alpha* 58(2). (Artículo aceptado).
- Orol, Ana & Alonso, Margarita. 2013. A comparative study of collocations in a native corpus and a learner corpus of Spanish. *Procedia Social and Behavioral Sciences* 95. 563–570.
- Quiñones, Virginia de Alba. 2009. El análisis de errores en el campo del español como lengua extranjera: Algunas cuestiones metodológicas. *Revista Nebrija de Lingüística Aplicada* 5(3). 1–16.
- Quiñones, Virginia de Alba. 2016. Variable de género y competencia colocacional léxica en el aula de ELE. *Porta Linguarum* 26. 201–215.
- R Core Team. 2013. *R: A language and environment for statistical computing*. Viena: R Foundation for Statistical Computing.
- Real Academia Española & Asociación de Academias de la Lengua. 2009. *Nueva gramática de la lengua española. Manual*. Madrid: Espasa.
- Rica, Juan. 2010. Colocaciones gramaticales en la producción escrita de estudiantes universitarios españoles. *Reduca (Filología)* 2(1). 1–25.

- Sánchez, Anna. 2010. Apuntes sobre las colocaciones léxicas y el concepto de colocación. *Anuario de Estudios Filológicos* 33. 291–306.
- Sánchez, Anna. 2015a. Las funciones diagnóstica y evaluativa del análisis contrastivo de la interlengua del español basado en corpus. *Linred* 13.
- Sánchez, Anna. 2015b. *El verbo dar en el español escrito de aprendientes de L1 inglés: Estudio comparativo entre hablantes no nativos y hablantes nativos basado en corpus*. Cáceres: Universidad de Extremadura. (Tesis doctoral).
- Sánchez, Anna. 2015c. Análisis contrastivo de interlengua y corpus de aprendientes: Precisiones metodológicas. *Pragmalingüística* 23. 191–210.
- Sánchez, Anna. 2015d. La investigación de corpus de aprendientes y el desarrollo de los estudios de la interlengua del español. *Language Design* 17. 57–48.
- Sánchez, Anna & Jiménez, Francisco. 2013. Apreciaciones sobre la cuestión de la norma en el análisis de la interlengua. *Normas. Revista de estudios hispánicos* 3. 183–204.
- Santillán, Paula. 2015. Nueva taxonomía de las colocaciones en árabe: Aspectos sintáctico, léxicos y semánticos. Granada: Universidad de Granada. (Tesis doctoral).
- Sinclair, John. 1991. *Corpus, concordance, collocation: Describing English language*. Oxford: Oxford University Press.
- Travalia, Carolina. 2006. Las colocaciones gramaticales en español. *Anuario de estudios filológicos* 29. 279–295.

- Vaisberg, Rebeca. 2013. *Cinco nuevas funciones sintácticas en el español*. Caracas: Universidad Católica Andrés Bello. (Trabajo de ascenso para optar al escalafón de Profesor Agregado.)
- Vásquez, Graciela. 1992. *Análisis de errores y aprendizaje de español / lengua extranjera*. Fráncfort: Peter Long.
- Vásquez, Graciela. 2009. El concepto de error: Estado de la cuestión y posibles investigaciones. *Revista Nebrija de Lingüística Aplicada* 5(3). 104–112.
- Vincze, Orsolya. 2015. *Learning multiword expression from corpora and dictionaries*. La Coruña: Universidade da Coruña. (Tesis doctoral).
- Zato, Zoltan. 2011. Alternancia preposicional en los complementos preposicionales regidos en español: Aproximación subléxica. *Círculo de Lingüística Aplicada a la Comunicación* 48. 41–76.