



Análisis de las frecuencias léxicas verbales de redacciones universitarias de estudiantes asesorados de la Clínica de Escritura de la Facultad de Psicología de la UASLP, México

Analysis of the verbal lexical frequencies of university essays by students advised in the Writing Clinic of the School of Psychology of the UASLP, Mexico

Análise das frequências lexicais verbais dos ensaios universitários pelos estudantes aconselhados na Clínica de Escrita da Faculdade de Psicologia da UASLP, Mexico

Marco Pérez

Universidad Autónoma de San Luis Potosí, México

marco.duran@uaslp.mx

<https://orcid.org/0000-0003-0854-3109>

Guadalupe Rojas

Universidad Autónoma de San Luis Potosí, México

rojasgpe@outlook.es

<https://orcid.org/0000-0002-6154-3610>

Resumen

Este estudio de corte cuantitativo descriptivo contrastivo se centra en el análisis de la frecuencia de aparición de los vocablos verbales en sesenta y ocho producciones escritas por estudiantes universitarios de la Facultad de Psicología de la Universidad Autónoma de San Luis Potosí, México, quienes asistieron a sesiones de asesoría en escritura académica antes del periodo de pandemia por COVID-19. El análisis de los datos se realizó mediante la aplicación del programa de acceso libre AntConc. Con este programa se obtuvieron la posición de la palabra (Rank), la frecuencia de repetición (Freq) y las palabras en sí (types y tokens). De los 350 verbos que aparecen, se tomaron los diez más frecuentes (38 % en total) y se contrastaron con el corpus de CREA de la RAE y los corpus de habla adulta, infantil y de textos periodísticos de Moreno y Guirao (2008). Los resultados muestran que no existe variación significativa entre los corpus contrastados. Los estudiantes utilizan con frecuencia verbos atributivos y auxiliares en sus producciones escritas. Los resultados presentados en este trabajo resaltan la necesidad de incluir el uso de verbos de corte más complejos en los currículos de escritura a nivel universitario.

Palabras clave: escrituraescritura académica; frecuencia verbal; lista de palabras; vocablo; unidad verbal.

Abstract

This paper aims to describe the frequency of occurrence of verbal words, in 68 productions written by university students at the School of Psychology of the Universidad Autónoma de San Luis Potosí, México. The students attended counseling sessions on academic writing before the COVID-19 pandemic period. The methodology used was quantitative-contrastive. The data analysis was carried out by applying the AntConc, a free access program. The word position (Rank), the frequency of repetition (Freq) and the words themselves (types and tokens) were obtained. A total of 350 verbs were identified and the 10 most frequent (38 % in total) were taken and contrasted with the CREA corpus of the RAE and the corpus of adult, children, and journalistic texts of Moreno and Guirao (2008). The results show that there is no significant variation between the contrasted corpora. Students frequently use attributive and auxiliary verbs in their text productions. The results discussed in this paper highlight the need to include the use of more complex verbs in writing curricula at the university level.

Keywords: dynamic hypermedia devwriting; verb frequency; word list; word; verbal unit.

Resumo

Este estudo exploratório centra-se na análise da frequência da ocorrência de palavras verbais em 68 produções escritas por estudantes universitários da Faculdade de Psicologia da Universidade Autónoma de San Luis Potosí, México, que participaram em sessões de aconselhamento de escrita académica antes do período pandémico da COVID-19. A metodologia utilizada foi quantitativa-contrastiva. A análise dos dados foi realizada utilizando o programa de acesso aberto AntConc, que foi utilizado para obter a posição da palavra (Rank), a frequência de repetição (Freq) e as próprias palavras (tipos e fichas). Dos 350 verbos que aparecem, os 10 mais frequentes (38 % no total) foram tomados e comparados com o corpus CREA da RAE e o corpus de textos adultos, infantis e jornalísticos de Moreno e Guirao (2008). Os resultados mostram que não há variação significativa entre os corpos contrastados. Os estudantes utilizam frequentemente verbos atributivos e auxiliares nas suas produções escritas. Os resultados apresentados neste artigo sublinham a necessidade de incluir a utilização de verbos mais complexos na escrita de currículos a nível universitário.

Palavras-chave: escrita académica; frequência verbal; lista de palavras; palavra; unidade verbal.

Recibido: 02/04/2023

Aceptado: 01/07/2023

Publicado: 26/09/2023

1. Introducción

El siguiente trabajo es un estudio exploratorio relacionado con la frecuencia de los verbos que aparecen en producciones escritas de estudiantes de primer año de universidad de la Facultad de Psicología de la Universidad Autónoma de San Luis Potosí (UASLP), México, que tomaron el servicio de asesoría en escritura académica dentro de la Clínica de Escritura en la misma facultad durante el periodo de agosto a diciembre de 2018, antes de pandemia. El objetivo de este trabajo de investigación es realizar un estudio descriptivo de frecuencia léxica de voces verbales con la finalidad de contrastar los datos obtenidos de este corpus con los corpus de CREA de la RAE y los corpus de habla adulta, infantil y textos periodísticos de Moreno y Guirao (2008) sobre análisis verbal. La premisa de este estudio es que los verbos, que aparecen en textos escritos por estos estudiantes, coinciden con los resultados de las respuestas verbales de los distintos corpus que se estudian.

El dominio de la escritura en la educación superior es fundamental no solo por su función comunicativa, la cual permite al estudiante universitario integrarse a la comunidad académica de la disciplina que ha elegido, sino por el valor epistémico que conlleva; debido a que durante el proceso de escritura el alumno organiza las ideas y reformula la información que ha leído mediante el léxico verbal y nominal. De esta manera, crea sus posturas y argumentos frente al tema asignado (Hayes y Flower, 1980; Carlino, 2003; Contreras y Aguirre, 2020). El conocimiento y dominio de un vocabulario extenso es un aspecto relevante para la escritura, porque el léxico es un recurso del aspecto cognitivo que regula, en primer momento, la estructuración de las ideas y, después, la construcción

del pensamiento en la escritura. Pedagogos, psicolingüísticas, psicólogos, psicopedagogos, etc., se han encargado de estudiar la escritura como proceso comunicativo y han manifestado preocupación en cuanto a las dificultades que muestran los estudiantes de educación media y superior en México ante el poco dominio de ella. Este bajo dominio repercute de forma directa en la transmisión comunicativa correcta de la información. En este trabajo, se considera que una de las principales problemáticas de la escritura académica reside en el bajo dominio del conocimiento léxico de quienes enseñan y de quienes aprenden. Si bien es cierto que el estudiante debe tener conocimiento de cierto tipo de vocabulario formal cuando ingresa a la educación superior, no siempre es así, o bien, si lo conoce no siempre logra aplicarlo durante su proceso de escritura académica.

Existen pocos estudios relacionados con el análisis de vocabulario académico a nivel superior. En el contexto mexicano, la mayoría de ellos se enfocan en estudiar el reconocimiento del vocabulario académico mediante la aplicación de instrumentos de medición. En la revisión de estudios similares a este, no se encontraron trabajos enfocados en estudiar los verbos de producciones escritas por estudiantes universitarios que tuvieran sesiones de asesoría de escritura. El trabajo más cercano a este es el de las autoras Godínez y Alarcón (2020). En su trabajo, las autoras revisan el léxico en la evaluación y en la didáctica de la escritura de textos académicos de literatura sin enfocarse en una categoría en general como pasa con el verbo en este estudio. Por esta razón, un estudio enfocado en explorar el vocabulario que los estudiantes emplean en sus producciones escritas es relevante, ya que permite conocer el contexto general estadístico relacionado con la lengua (Butler, 1985; Brown, 2008). Además, en este tipo de estudios, se pueden realizar análisis que midan la frecuencia léxica, la longitud de palabra, frases u oraciones, los tipos oracionales: simples o compuestos, la complejidad oracional, la diversidad léxica, la disponibilidad léxica, etc. (López, 2011; Pallotti, 2015; Kurdi, 2017; Lahuerta, 2017; Ochoa, 2021). Para estudiar un corpus desde una perspectiva descriptiva contrastiva se vincula la lingüística cuantitativa con la lingüística del corpus. Gracias a esta última área de estudio, el análisis de los datos se hace de forma más precisa, ya que se pueden estudiar grandes volúmenes de información a través de medios electrónicos. Mancera y Pano (2014) afirman, sobre esto, que el análisis de los corpus ofrece mayor fiabilidad en los análisis cuantitativos y cualitativos que se aplican al estudio de la lengua en todas sus dimensiones. Su resultado estadístico favorece y describe mejor la estructura formal del texto, mismo que es necesario para hacer mejores descripciones lingüísticas.

2. Marco teórico

Dentro de los procesos de escritura académica, la competencia léxica juega un papel relevante, ya que el conocimiento de un amplio espectro léxico brinda al escritor la posibilidad de expresar sus ideas de acuerdo con sus intenciones comunicativas. A pesar de ello, esta última ha sido más estudiada desde el ámbito de la enseñanza de segundas lenguas (ELE) como parte del currículo de enseñanza del español y de la escritura académica (Varela *et al.*, 2013; Riffo *et al.*, 2014; Valdés-León, 2021). La primera se puede entender como la capacidad de cualquier hablante de conocer y saber utilizar de forma clara el vocabulario de su lengua materna dentro de una situación comunicativa establecida. Nation (2001) define tres aspectos generales de la competencia léxica: conocimiento de la forma, del significado y del uso. Esto implica que el informante sea capaz de expresarse adecuadamente en situaciones comunicativas orales y escritas. La segunda, la escritura académica, hace uso de esta competencia con la cual los escritores se aproximan a la comprensión de los textos, a la exposición de sus ideas y la transformación del conocimiento adquirido.

Existen dos tipos de palabras: funcionales y de contenido (Bosque, 2007). Las primeras se refieren a aquellas que sirven de enlace entre las unidades léxicas (preposiciones, conjunciones, etc.), las segundas son las que presentan contenido semántico y que sirven como base para derivación de nuevos términos con diferente categoría gramatical. En este sentido, las palabras son portadoras de la organización del pensamiento y unidades léxicas que pueden ser cuantificables. Sobre la comunicación escrita, la planeación curricular es clave en el desarrollo de esta, porque los estudiantes adquieren las habilidades formales, semánticas y cognitivas que engloban el proceso de escritura académica. En las planeaciones curriculares se debe incluir un apartado dedicado a adquirir, no solo desarrollar, el léxico formal, bases léxicas y su familias.

Al ser unidades cuantificables, pueden formar corpus que servirán de herramienta académica para conocer el estado de la lengua de una comunidad lingüística. Para ello, hay que considerar la relación fundamental entre la educación, la escritura y el léxico. El conocimiento de este último es relevante tanto en el proceso de comprensión lectora (Riffo *et al.*, 2014), como en el proceso de escritura universitaria (Godínez y Alarcón, 2020). De ahí que su estudio haya sido clave a la hora de estudiar la construcción de la escritura académica como fenómeno de análisis. Como ejemplo de esto último, Cerda *et al.* (2017) han estudiado el léxico como un proceso clave en la enseñanza-aprendizaje en la escritura académica a través de metodologías cuantitativas.

Desde el ámbito constructivo del corpus, denominado como complejidad estructural, que es la propiedad formal de los textos (número de elementos lingüísticos) y que sirven de base para el análisis estadístico más específico (Pallotti, 2015), se pueden clasificar las listas de palabras como listas para la elaboración de textos académicos, para aprender una segunda lengua, para mejorar la primera o para conocer un tipo de vocabulario, etc. Además, estas listas se pueden clasificar desde temáticas generales (la Lista de Palabras Generales, GLS, *General Service Lista* de West, 1953) o específicas (la Lista de Palabras Académicas, AWL, *Academic Word List* de Coxhead, 2000). Una de las características que se deben considerar en la construcción de un corpus es la exclusión de las palabras funcionales, coloquiales o especializadas (Nation, 2001; Coxhead y Nation, 2001). Este tipo de palabras no necesariamente sirven como elementos de reconocimiento de la lengua.

A pesar de ser instrumentos de apoyo para el proceso de aprendizaje, Hyland y Tse (2007) y Martínez *et al.* (2009) han afirmado que las listas de palabras presentan un costo muy alto, por la cantidad de horas que requiere su actualización, por la cantidad de significados que una palabra pueda tener en distintas disciplinas, y por la cantidad de cambios que constantemente se dan en la lengua y que son inherentes a ella. Conocer el uso que se hace de los vocablos verbales es importante al momento de elaborar secuencias de intervención en los programas curriculares de escritura académica. De esta manera, se fortalece no solo el aprendizaje del léxico académico, sino el uso de este en los contextos reales de producción.

Cuando se trabaja con un corpus, uno de los muchos estudios que se pueden hacer tiene que ver con su riqueza léxica. Esta riqueza es un indicador cuantificable aplicado a distintos tipos de textos o diferentes grupos comparativos: nominales, verbales, adjetivales o a una mezcla de corte gramatical en distintas listas de palabras. En este sentido, aquello que se toma en cuenta para obtener la riqueza léxica es la extensión o longitud del vocabulario, la frecuencia léxica, el corpus del que se obtiene, el tipo de unidades léxicas, la periodización del corpus o su comportamiento, entre otras. Todas estas son características que se tienen en cuenta en una lista de palabras.

El concepto de frecuencia léxica se ha trabajado desde el ámbito de la lingüística del corpus y se puede aplicar en áreas como los procesos de escritura, el análisis del discurso, la sociolingüística, la fonología, etc. Bybee (2007) ha comentado que la frecuencia es un factor funcional en el análisis y explicación de la estructura de las lenguas naturales. El recuento de las frecuencias es un indicador clave en la medición de la riqueza léxica. Para ello, se hace la división del número de palabras distintas (*type*) sobre el número total de palabras que posee un texto (*tokens*): $type/tokens$. El resultado es la riqueza media del texto (Templin, 1957).

A la relación *type-tokens* se añade la *ratio*, es decir, la relación cuantificada entre dos cantidades que reflejan una proporción matemática. Esta relación se representa así: *TTR* y su valor está entre 0 y 1. Por lo que, a lo largo de un texto, la *TTR* no es un valor constante, sino que va disminuyendo a medida que va aumentando la cantidad de *tokens* (Capsada y Torruella, 2017). Lo anterior permite comprender que la relación que existe entre *type-tokens* no es una relación lineal, si hay más *tokens* en un texto, menos *types* aparecerán en ese mismo. En ese sentido, no hay relación de proporcionalidad, sino que esta relación está vinculada a la longitud del texto. Los *tokens* aparecen cuando se mide la frecuencia de los ítems; los *types* aparecen si se miden los lemas o vocablos. Santos (2017) menciona que cuando se trabaja con las frecuencias de los *types* se pueden conseguir los patrones de frecuencia de los *tokens*, al obtener ambas frecuencias se llega a los ítems.

Desde un criterio semántico, los verbos son unidades léxicas plenas de contenido. Estas permiten al usuario reconocer los argumentos, en el sentido literal o metafórico, al momento de construir una unidad sintáctica. A diferencia de las palabras vacías o funcionales (Bosque, 2007), los verbos pueden ser bases para que mediante procesos morfológicos se obtengan otras categorías derivadas de este que compartan la misma base de significado. Además, el verbo determina la construcción de sentido de la oración en la que se encuentra. De ahí, la relevancia de su estudio.

2.1. Clínica de escritura

Con el fin de dar atención a las necesidades de escritura de los alumnos de la comunidad universitaria de la Facultad de Psicología de la UASLP, México, se creó la Clínica de Escritura a partir del esquema de trabajo aplicado en la Facultad de Educación de la Universidad de Diego Portales en Chile. En la Clínica de Escritura se brinda un paquete de seis asesorías de cincuenta minutos cada una. Los asesores de la Clínica de Escritura son estudiantes de sexto a décimo semestre de la licenciatura en Psicopedagogía con énfasis en lenguaje. Las sesiones se diseñan a partir de una sesión de diagnóstico y de los resultados de evaluación de los textos aportados por los estudiantes. En cada sesión se trabaja aspectos relacionados con el aspectos que van desde lo discursivo, lo macro y microestructural, los recursos del lenguaje para mantener la cohesión y la coherencia, el uso de vocabulario académico y las relaciones estructurales de la lengua. El objetivo de las asesorías es que el asesorado trabaje en sus propias fortalezas y debilidades durante las sesiones y, de esta forma, mejore su escritura en diferentes aspectos del lenguaje.

Para llevar a cabo el diagnóstico, se pidió a los estudiantes que entregarán un texto previamente escrito y que escribieran otro *in situ* a partir de un estímulo. Este último fue el texto que se usó para la elaboración de la base de datos. La consigna para obtener el texto fue la siguiente: Escribe un texto de no más de una cuartilla donde respondas a la siguiente pregunta: ¿Qué has aprendido sobre tu carrera? El texto obtenido se utilizó para identificar los aspectos relevantes sobre los cuales se debe

trabajar durante las asesorías. Una vez finalizado el conjunto de asesorías, se vuelve a aplicar la misma consigna para obtener el texto postintervención.

3. Metodología

La metodología utilizada fue de corte cuantitativa contrastiva aplicada al análisis de las voces verbales de escritos de estudiantes universitarios que asistieron como asesores del centro de prácticas Clínica de Escritura de la Facultad de Psicología de la Universidad Autónoma de San Luis Potosí, México. Los textos pertenecen a estudiantes de Psicología y Psicopedagogía de la misma facultad, y se obtuvieron durante enero y febrero de 2020, antes del inicio de la educación virtual provocada por la pandemia de COVID-19. De estas redacciones, se ubican las voces verbales que aparecen en las redacciones preintervención y postintervención.

3.1. Muestra y corpus

La base datos no se obtuvo mediante la aplicación de un cuestionario como en los trabajos de disponibilidad léxica (López-Morales, 2005; Zambrano, 2021) o la realización de entrevistas orales con formato cerrado, por ejemplo; como en los estudios de PRESEEA (Proyecto para el Estudios Sociolingüístico del Español de España y de América). En este caso, el análisis parte de la idea de riqueza léxica (Ávila, 1988, 1989; López-Morales, 2008; Capsada y Torruella, 2017), los datos se obtuvieron mediante la revisión de sesenta y ocho escritos (treinta y cuatro preintervención y treinta y cuatro postintervención) de estudiantes que asistieron a la Clínica de escritura de la Facultad de Psicología de la Universidad Autónoma de San Luis Potosí, México. Cada uno de los escritos se sometió a la segmentación de sus unidades, palabras, hasta conformar una base de datos de ítem léxicos. Estos ítems léxicos fueron clasificados por categoría gramatical. De esta clasificación se obtuvieron los verbos que sirvieron de corpus para este análisis. Dicho corpus se compone de los verbos que aparecen en los escritos pre y postintervención. En este trabajo se analizaron dos unidades: 1) *type* (ocurrencias), unidad gráfica que corresponde al total de palabras diferentes en un corpus y 2) *tokens* (palabras de un corpus), las palabras que se encuentran repetidas. Al hacer la revisión teórica no se encontraron, hasta este momento, datos de referencia con los cuales comparar los datos obtenidos. Por este motivo, se utilizó el corpus de CREA y el conjunto de corpus de Moreno y Guirao (2008): el corpus de habla adulta, el corpus infantil y el corpus de texto periodístico.

La elección de los corpus obedece a los siguientes criterios: 1) el Corpus de Referencia del Español Actual (CREA) es un corpus que está constituido por un conjunto de diversos textos y tiene como finalidad proporcionar información relevante de todas las variedades del español. Este corpus se distribuye de la siguiente forma: 50 % pertenece a documentos provenientes de España y el 50 % restante a toda América, el 90 % son documentos escritos y el 10 % restante son documentos orales (Molina y Sierra, 2015). Su importancia, en cuanto a que es el corpus rector del español, sirve de base para el contraste; 2) Los otros corpus, el de habla adulta, infantil y de texto periodístico se toman en consideración para el contraste porque los corpus utilizados presentan la lista de los 100 lemas verbales más frecuentes por cada corpus que toman estos autores para el contraste en este estudio. En este trabajo, solo se toman en cuenta los diez verbos lematizados más frecuentes de las redacciones y se comparan según la posición donde se ubican dentro de las listas verbales.

3.2. Software de análisis

El análisis se realizó utilizando el programa AntCon para el análisis de los datos. Este *software* es de acceso libre. Con este programa, se extraen datos de amplios corpus textuales por frecuencia, por patrones de búsqueda, por palabras, etc. Este *software* permite identificar mejor el análisis por frecuencia y pone la información en tablas comparativas de las que se extrae la información necesaria para el análisis de los datos. El uso de este corpus se enfocó en la lematización de verbos y la extracción de las frecuencias.¹

3.3. Procedimiento

Para la parte cuantitativa, se utilizó la noción de variabilidad verbal que se aplica a los trabajos de carácter léxico. Esta es una medida que depende de un alto grado de coincidencias léxicas entre los corpus contrastados aplicados a la categoría verbal y que se utiliza para el cálculo de la riqueza léxica. Este concepto hace referencia a la cantidad de palabras diferentes del corpus. Un texto con mucha riqueza estará formado por muchas palabras diferentes con baja frecuencia y viceversa. Para calcular la riqueza léxica, se emplea la fórmula *type-token ration*. En esta fórmula se hace una división entre el total de palabras distintas (*types*) con el número total de palabras que el texto contiene (*tokens*). El resultado es la riqueza media de cada texto (Capsada y Torruella, 2017). Sobre la fórmula *type-token ration*, el número de *types* siempre será menor a los *tokens* y su rango de variabilidad está relacionado con un rango que va de 0 a 1. En cuanto a la variación, 1 representa que el texto es variable y 0 indica que el corpus es menos variable. En palabras de Sabaj (2004), la variabilidad verbal se refiere al coeficiente entre los tipos de lemas verbales y la cantidad total de las formas verbales del corpus. La fórmula de la variabilidad verbal es la siguiente:

$$\text{Variabilidad verbal} = \text{Tipos de lemas verbales} / \text{total de casos verbales.}$$
$$\text{Porcentaje verbal del corpus} = \text{Total de casos verbales} * 100 / \text{número total de palabras.}$$

Antes de utilizar el *software* seleccionado, se obtuvo la muestra, se depuró y se eliminaron todas las palabras de corte funcional: preposiciones, conjunciones, artículos, etc. del corpus de forma manual. Posterior a esto, se aplicó el *software* AntConc para realizar el análisis. Una vez realizado esto, se obtuvo el primer análisis por extracción de frecuencias. Este procedimiento consiste en contar las palabras de un corpus y mostrarlas en una lista ordenada por frecuencia alta hasta frecuencia baja. Este análisis por frecuencia contiene la posición de la palabra (*Rank*), el número de ocasiones que se repite esa palabra en el corpus (*Freq*) y las palabras en sí (*Word*). En esta lista, aparecen en la parte superior el total de *tokens* y los *types*.

Al hacer el primer análisis por frecuencia, se detectaron errores en la corrección de los datos, por lo que se volvió a depurar el corpus de forma manual hasta alcanzar un corpus limpio de palabras funcionales. Cuando se analizó el corpus, se encontraron palabras con variación morfológica, lo que significa que el verbo se encontraba en forma flexionada, por lo que el *software* lo identificaba como una palabra distinta. Esto obligó a lematizar el corpus. La lematización consiste en un proceso automático mediante el cual cada palabra del corpus (*token*) se vincula con su forma canónica (no marcada o lema). Para este caso, el infinitivo es la forma lematizada del verbo. De esta forma, el cálculo de frecuencias contará todas las apariciones de un verbo como una única forma, independiente de la

¹ Su página web es <https://www.laurenceanthony.net/software/antconcl/>

flexión. El resultado de lematizar el corpus es que hay modificación de las frecuencias en relación con el primer corpus. Para que la lematización fuera precisa, se elaboraron (en otro archivo independiente con el nombre de verbo) todas las entradas de los verbos que aparecen en el corpus. De esta forma, el análisis que se presenta es un trabajo detallado de los verbos que se obtuvieron del corpus.

Las limitaciones de este trabajo tienen que ver con que, al ser una investigación de corte descriptivo contrastivo y por la naturaleza del objeto de estudio, no existen datos comparativos que establezcan un análisis más detallado de los verbos desde el análisis que se hace; sumado a esto, está el tamaño de la muestra y la temporalidad debido a que los datos fueron recolectados antes de la pandemia provocada por el COVID-19 y que es necesario dar a conocer los resultados obtenidos. Estos tres factores incidieron en el análisis contrastivo de los resultados. A pesar de esto, los datos que aquí se presentan pueden funcionar como base de contraste para otras investigaciones en el tema.

4. Análisis

El número total de palabras fue de 24 708. Al procesar la muestra, el número total de vocablos disminuyó a 2 923 del corpus de voces verbales. Esto representa el 8.40 % de la totalidad del corpus. En los estudios sobre frecuencias léxicas, del corpus general, se apreció una disminución de 90 % de las unidades léxicas del corpus que se estudia. Esto se debe a la depuración y aplicación del *software* para el análisis. Sobre el total de vocablos verbales que se obtuvieron, estos ascienden a 358 verbos. Sobre el análisis de los verbos, se presenta la distribución en la Tabla 1.

Tabla 1

Distribución de los datos type, tokens obtenidos después de la aplicación de software AntConc

Número total de palabras (<i>Word tokens</i>)	Número total de vocablos (<i>Word type</i>)	Preintervención		Postintervención	
		<i>Tokens</i>	<i>Types</i>	<i>Tokens</i>	<i>Types</i>
24,708	2,923	12,081	2,915	12,065	2,919

Nota. Elaboración propia a partir de los datos obtenidos.

Del anterior cuadro, el número total de *Tokens* en la preintervención fue de 12 081 y de la postintervención fue de 12 065. Al comparar los datos, se observó que hubo una diferencia entre la muestra preintervención y la postintervención de dieciséis *tokens* en favor de la preintervención. Esto representa el 0,1 % de contracción de *tokens* del corpus, sobre la diferencia de los *Types* entre pre y postintervención, lo que significa que existe una diferencia de cuatro *Types* a favor de la postintervención. Esta es una diferencia mínima que aparece en el corpus. Esto puede observarse en la Tabla 2.

Tabla 2

Datos comparados de la muestra pre y postintervención del estudio

Número total de palabras (<i>Word tokens</i>)	Número total de vocablos (<i>Word type</i>)	Preintervención		Postintervención	
		<i>Lemma tokens</i>	<i>Lemma types</i>	<i>Lemma tokens</i>	<i>Lemma types</i>
24,708	2,923	12,081	2,524	12,065	2,531

Nota. Elaboración propia a partir de los datos obtenidos.

Como se ha mencionado, un lema es una forma que agrupa normalmente todas las posibles variantes flexivas de una determinada palabra. Sobre esto, el número de lemas de preintervención fue de 12 081. Al comparar este resultado con la comparación de la información hay una diferencia de dieciséis *Tokens* a favor de la Preintervención. Sobre el análisis de los verbos, se muestran los resultados en la Tabla 3.

Tabla 3

Resultados obtenidos de las comparaciones del total de vocablos preintervención y postintervención

Número total de vocablos (<i>Word type</i>)	Preintervención		Postintervención	
	<i>Types</i>	Verbos	<i>Types</i>	Verbos
2,923	2,524	317	2,531	350

Nota. Elaboración propia a partir de los datos obtenidos.

Sobre el total de verbos que se analizaron en este corpus, los verbos que aparecen en preintervención representan 12,5 % de la totalidad de este corpus y, por la postintervención genera 13,8 %. Si se comparan los verbos que aparecen pre y postintervención, la diferencia es de apenas treinta y tres verbos a favor de la postintervención, lo que representa 9,4 % de la totalidad verbal aparecida en el corpus. En la tabla 4, se encuentra la lista de los diez verbos frecuentes obtenidos de la muestra.

Tabla 4

Lista de los diez verbos más frecuentes obtenidos en la muestra durante la preintervención y la postintervención

	Posición en la tabla	Verbos Preintervención	Posición en la tabla	Verbos Postintervención
1	1	ser	1	ser
2	2	haber	2	haber
3	4	aprender	4	aprender
4	5	estar	5	estar
5	6	tener	6	tener
6	11	poder	10	poder
7	17	desarrollar	14	desarrollar
8	20	dar	19	dar
9	24	hacer	23	hacer
10	30	deber	29	deber

Nota. Elaboración propia a partir de los datos obtenidos.

Las respuestas de la tabla anterior muestran que, debido a lo compacto de este corpus, se acumulan entre los primeros lugares los mismos verbos frecuentes. La diferencia entre pre y postintervención es la posición en que aparecen los verbos. Por lo que respecta al análisis contrastivo, ahora se hace la comparación entre los diez verbos más frecuentes de este corpus y el corpus CREA. Los resultados se presentan en la tabla 5.

Tabla 5

Lista de comparación entre los 10 verbos más frecuentes de este corpus y el Corpus CREA

Corpus estudiantes			Corpus CREA	
	Posición	Verbo	Posición	Verbo
1	02	haber	39	ser
2	15	ser	110	hacer
3	18	desarrollar	146	decir
4	40	hacer	166	poder
5	47	deber	168	ver
6	56	aprender	191	tener
7	62	dar	209	haber
8	70	poder	240	estar
9	96	tener	279	dar
10	185	estar	307	partir

Nota. Elaboración propia a partir de los datos obtenidos.

Sobre la comparación, el verbo ser se ubicó en los dos primeros lugares de ambos corpus. Sobre el resto de los verbos: haber, hacer, dar, poder, tener y estar se encontraron en los dos corpus, lo que significa que son muy frecuentes en las redacciones. Los verbos desarrollar, decir, ver y aprender son los que no presentan correlato entre los corpus (CREA, habla adulta, infantil y periodística). Por la posición de los datos, el corpus de estudiantes alcanzó casi el 100 % de las respuestas con nueve verbos de las diez respuestas más frecuentes mientras que el corpus CREA, solo tuvo entre las primeras 100 respuestas un solo verbo. En tabla 6, se presentan estas desinencias.

Tabla 6

Número de apariciones por desinencias de la muestra en comparación con el corpus CREA

Corpus estudiantes			Corpus CREA	
	Posición	Verbo	Posición	Verbo
1	1	es	19	es
2	9	son	28	ha
3	14	está	40	son
4	21	esté	43	fue
5	29	puede	44	había
6	31	tiene	45	era
7	41	he	50	está
8	70	hay	55	han
9	77	sea	57	hay
10	80	aprendí	59	puede

Nota. Elaboración propia a partir de los datos obtenidos.

Como se aprecia, por posición, el corpus del CREA llega a las diez desinencias en la posición 59, mientras que los estudiantes alcanzaron las desinencias en la posición 80. Además, las desinencias de los verbos que más aparecieron en el corpus estudiantil son: ser, con tres apariciones; estar y haber, con dos apariciones; y, con una sola aparición, los verbos: poder, tener y aprender. Sobre el corpus

CREA, con cuatro apariciones se encuentran ser (es, son, fue, era) y haber (ha, había, han, hay), y con una aparición se encuentran estar y poder. Al comparar la posición entre los corpus, es más diverso el corpus de los estudiantes que el corpus CREA, ya que este es más compacto, es decir, en el corpus CREA, solo hay cuatro verbos: ser, haber, estar y poder, que acaparan la mayor cantidad de respuestas en la lista, mientras que en el corpus de estudiantes hay seis verbos: ser, estar, poder, tener, haber y aprender. Si se comparan las respuestas, ser, estar, haber y poder aparecen en los dos corpus. Por su parte aprender y tener son los verbos que no aparecen en la lista del corpus CREA. Por último, se contrastan los diez verbos de este corpus con los obtenidos por Moreno y Guirao (2008). Estos datos se observan en la tabla 7.

Tabla 7

Tabla de comparación de los diez verbos obtenidos frente a los verbos del corpus de Moreno y Guirao (2008)

<i>Posición</i>	<i>Escritos universitarios</i>	<i>Habla adulta</i>	<i>Habla infantil</i>	<i>Textos periodísticos</i>
1	haber	ser	ser	ser
2	ser	decir	tener	tener
3	desarrollar	estar	estar	hacer
4	hacer	tener	saber	decir
5	deber	hacer	hacer	haber
6	aprender	haber	ir	estar
7	dar	ir	decir	señalar
8	poder	ver/dar	haber	dar
9	tener	saber	ver	asegurar
10	estar	pasar	llamar	informar

Nota. Elaboración propia a partir de los datos obtenidos.

Al comparar los resultados, hay mucha similitud entre las redacciones universitarias con los resultados del estudio de Moreno y Guirao (2008). En términos generales, los verbos coincidentes en los cuatro corpus son haber, ser, hacer, tener y estar. Son seis verbos que representan el 12,5 % de la totalidad de verbos de este cuadro. Por su posición, ser apareció en segundo lugar en las redacciones académicas mientras que, en los tres distintos corpus en primer lugar, el verbo haber apareció en primer lugar en este corpus de estudiantes y en los otros tres corpus cambió de posición: lugar 6 en el habla adulta, lugar 8 en la infantil y 5 en la periodística. El verbo hacer presenta parecido en cuanto a la posición. En el primer corpus, este verbo apareció en el lugar cuatro en el habla adulta, cinco en el habla infantil y en el lugar tres del corpus periodístico. Sobre el verbo tener, este apareció en el lugar 9 en el corpus de redacciones y en el resto de los corpus se ubica en las primeras posiciones: lugar 4 en el corpus de habla adulta, y en el lugar 2 tanto en el habla infantil como en el texto periodístico.

Dar y decir son verbos que aparecen en tres contextos de uso. El primero, se ubicó en redacciones universitarias en el lugar 7, habla adulta en la posición 8 al igual que en los textos periodísticos. Lo que representa el 2,5 %. Los verbos que aparecen en dos corpus son ir, saber y ver que aparecieron en las hablas adulta e infantil. Esto representa el 7,5 % del total de la muestra. De esta comparación, los verbos que aparecen en un solo corpus son: en el corpus estudiantil, deber, aprender y poder; en el corpus de habla adulta, pasar y en el corpus infantil, llamar. En el corpus periodístico, se ubican señalar, asegurar e informar. Son 8 verbos que representan el 20 % de la totalidad. Esto se puede observar en la tabla 8.

Tabla 8

Contraste de frecuencias de los corpus estudiados

Estudiantes		RAE		Adultos		Infantil		Periodísticos	
Fr.	P	Fr.	P	Fr.	P	Fr.	P	Fr.	P
637	ser	232,924	ser	7,404	ser	509	Ser	98,694	ser
275	haber	96,063	hacer	2,652	decir	330	tener	44,153	tener
165	aprender	76,867	decir	2,404	estar	193	estar	33,509	hacer
162	estar	66,900	poder	2,388	tener	176	saber	30,579	decir
160	tener	66,479	ver	2,220	hacer	172	hacer	23,849	haber
98	poder	58,029	tener	1,456	haber	129	ir	23,038	estar
59	dar	54,150	haber	1,392	ir	118	decir	14,098	señalar
55	hacer	46,282	estar	964	ver	93	haber	13,927	dar
46	deber	37,196	partir	886	dar	89	ver	12,992	asegurar
39	ir	37,054	llegar	865	saber	88	llamar	12,275	informar

Nota. Elaboración propia a partir de los datos obtenidos.

Los verbos en infinitivo que aparecen en los diez primeros lugares más frecuentes son: ser, haber, tener, hacer, deber, ir, aprender, estar, poder, dar, decir, ver, partir, llegar, saber, llamar, señalar, asegurar e informar. Son diecinueve verbos que representan 38 %, si se toma en cuenta que la muestra se conforma por cincuenta verbos, diez por corpus. Los verbos que se encuentran en todos los corpus son: ser, haber, tener, estar y hacer. El verbo que apareció en primer lugar de las tablas es ser. El verbo decir se ubicó en cuatro listas de las cinco posibles, ya que no apareció en la lista de estudiantes. Los verbos dar, ver e ir aparecieron en tres corpus de los cinco. Poder y saber son los verbos que se encontraron en dos listas de palabras.

5. Discusión de los resultados

Los datos muestran que los estudiantes universitarios, que acudieron a sesiones de asesoría en escritura académica, hacen uso de verbos convencionales pertenecientes al léxico general, pero en diferentes posiciones. Al comparar estos datos con los trabajos que se han utilizado para el contraste se observa que: 1) En relación con el corpus CREA, el número de palabras verbales que aparecen en este corpus es de 891 de 5000, lo que representa 17,8 % de la totalidad corpus. Al comparar estos resultados con el corpus de estudiantes, el número de verbos que aparecen en los textos es mayor al que se registra en el CREA. La diferencia es de 2,032 respuestas a favor de las redacciones estudiadas. Sobre los verbos lematizados, en el corpus del CREA se reconocen 333 y en el de estudiantes, 346. Si se observa, la diferencia es de apenas trece verbos a favor de los escritos. 2) Si se comparan los resultados de los estudiantes en relación con las 5000 palabras más frecuentes del corpus del español de Varela *et al.* (2013), estos son muy parecidos a los encontrados del corpus CREA; es decir, hay 827 respuestas verbales. Entre estas respuestas, 316 son vocablos verbales de este corpus. 3) No obstante, los resultados están más cercanos al corpus CREA que a este corpus de escritos de estudiantes de los vocablos verbales sin lematizar. Cuando este corpus ya está lematizado, la diferencia es mínima como se aprecia en los tres corpus analizados. Como parte del análisis presentado, los verbos que aparecen en las redacciones académicas universitarias son muy frecuentes en los demás corpus que sirvieron de base para el contraste.

Sobre los corpus de habla adulta, habla espontánea infantil y de registro periodístico de los autores Moreno y Guirao (2008) que se utilizaron para el contraste se menciona lo siguiente. Al ser estudiantes universitarios, su registro está más relacionado al registro de adultos por lo que el contraste hecho es más que evidente. Este corpus está integrado por tres secciones: el registro informal que contiene 150 000 palabras, 80 000 palabras de registro formal y 70 000 palabras de medios de comunicación. Sobre el corpus de los niños, este está compuesto por 60 000 palabras y el corpus de textos periodísticos está integrado por 15 millones de palabras. De estos corpus, se tomaron en cuenta los 100 vocablos lematizados que sirvieron de contraste para el análisis de los diez vocablos más disponibles. Como se analizó en la Tabla 4, los verbos en las primeras diez posiciones son muy parecidos. Esto indica que estos verbos son muy productivos en el español y, por ende, son verbos compactos. Es decir, estos no permiten el ingreso de otros verbos entre las primeras posiciones. Si se toman las siguientes diez posiciones (11 al 20), aparecen diferencias significativas entre el corpus de estudiantes y el resto de ellos que se utilizan para el contraste. Para ejemplificar esto, obsérvese la Tabla 9:

Tabla 9

Tabla comparativa de la posición 11 a la posición 20 del corpus obtenido y de los corpus de contraste

P.	Estudiantes	CREA	Adulta	Infantil	Periodísticos	E. P.
11	existir	ver	pasar	poner	presentar	trabajar
12	llevar	parecer	poner	jugar	considerar	anotar
13	ayudar	desarrollar	crear	pasar	explicar	comentar
14	ir	dar	venir	dar	afirmar	ver
15	realizar	partir	llamar	venir	llegar	llamar
16	mencionar	llegar	llevar	comer	mantener	comparar
17	hablar	tratar	hablar	caer	indicar	ir
18	placer	saber	quedar	comprar	encontrar	desarrollar
19	leer	pasar	querer	cortar	realizar	crear
20	parecer	querer	llegar	llevar	destacar	saber

Nota. Elaboración propia a partir de los datos obtenidos y los datos de corpus de contraste.

Como se aprecia en la Tabla 9, en los corpus educativos de estudiantes universitarios y de primaria, los verbos son de corte académico como se aprecia en las listas, y en el corpus periodístico, los verbos que aparecen en las posiciones 11 a la 20 son de corte informativo. Se ubican diferencias de voces verbales cuando se contrastan los corpus.

En relación con los corpus CREA y de habla adulta, los verbos encontrados son menos académicos y hacen referencia a situaciones generales, salvo por verbos como saber, que evoca conocimiento, además de llamar y hablar, verbos de enunciación, que pueden ser de corte académico. Esto nos lleva a la reflexión de que los verbos de corte informativo y académico se enseñan en el ámbito académico y no en el ámbito social, prueba de esto es el corpus tanto de adultos como de niños.

Las palabras que se enseñan en el ámbito académico deben formar un corpus cuya aplicación debe estar enfocada para que docentes y programadores de lengua puedan utilizarlas como parte de la programación de planes y estudios. A partir de los corpus, se podría decir que el léxico como conjunto de palabras que constituyen un corpus se clasifica en tres grandes categorías: palabras cotidianas (corpus de adultos y niños), palabras de corte académico (estudiantes) y palabras

informativas (periodistas). El corpus CREA es un corpus híbrido en el que creemos que por su naturaleza agrupa a todos los corpus.

En este estudio, los verbos que aparecen en las primeras diez posiciones son los mismos que en los corpus de contraste. Después de las diez posiciones, las variaciones verbales se aprecian mejor no solo por la posición de los verbos, sino por la frecuencia léxica que acompaña cada respuesta. La frecuencia léxica es un indicador que sirve de base para conocer información como la media, la varianza, la desviación estándar o el coeficiente de variación que junto representa el conocimiento matemático del léxico. Este conocimiento puede servir como punto de partida al momento de diseñar ejercicios de escritura académica donde el objetivo sea el incremento del vocabulario verbal.

6. Conclusiones

Al tomar en cuenta los datos analizados durante esta investigación, se llegó a la conclusión de que los diez primeros verbos de la lista de apariciones que se obtuvieron de las redacciones académicas universitarias de estudiantes de primer año son los mismos que se han encontrado en otros corpus. Esto significa que son verbos altamente disponibles que no presentan variación en cuanto a su grado de aparición. Nuestra expectativa inicial sobre las voces verbales en textos académicos estaba relacionada a la variación de estas respuestas en los textos escritos, lo que supondría la aparición de verbos de corte argumentativo, descriptivo o explicativo, porque son los verbos que forman parte de la escritura académica, no obstante, no se vio esto en las primeras diez posiciones.

La hipótesis y el objetivo se han cumplido cabalmente cuando comparamos las listas de palabras con las otras listas los corpus de contraste, aunque la cantidad de voces no lematizadas es más alta que en los demás corpus. Esto se puede explicar por las diferencias entre los corpus que se analizan; sin embargo, los lemas se mantienen en una proporcionalidad parecidas al resto de los corpus. Solo queda mencionar, en relación con esto, que el hecho de que sean parecidas las voces verbales lematizadas en los corpus distintos no significa que sea una constante si se pretende realizar otros estudios como este.

Si se retoma la premisa de este estudio sobre la idea de que los verbos aparecidos en estas redacciones son coincidentes con los resultados de los demás corpus, la realidad es que hay cinco verbos que tenderán de forma constante a aparecer como los verbos más frecuentes del español en cuanto a la escritura y, posiblemente, en cuanto a la oralidad. Faltaría buscar corpus de voces verbales de otras redacciones universitarias que contribuyeron a estos resultados y faltaría ampliar este tipo de estudios que si bien es cierto tienen una amplia tradición en el ámbito anglosajón, en México, no hay trabajos que permitan tomar como punto de referencia para generar investigación desde distintos enfoques.

A manera de reflexión, esta investigación presenta varias limitaciones. Por un lado, al ser una investigación de corte descriptivo contrastivo y, por la naturaleza del objeto de estudio, no existen datos comparativos que establezcan un análisis más detallado de los verbos desde el análisis que se hace. Por otro lado, el tamaño de la muestra es pequeña, además de que los datos fueron recolectados antes de la pandemia provocada por el COVID-19; por lo tanto, es necesario dar a conocer los resultados obtenidos. A pesar de esto, los datos que aquí se presentan pueden funcionar como base de contraste para otras investigaciones en el tema.

Para futuras investigaciones, consideramos relevante comparar corpus relativos al ámbito académico universitario, de preferencia, en aquellos corpus que se extraigan de centros de práctica académica o de clínicas de escritura que fortalezcan de mejor manera estudios lexicométricos. Además, es necesario generar listas de palabras académicas específicas del español de México que permitan dar respuestas a muchas de las interrogantes que se tienen en relación con el léxico académico. Sobre la riqueza léxica que pudiera describirse, no se logró, porque no hay corpus con estas características y, por el momento, no hay manera de saber si realmente hay riqueza léxica de los estudiantes universitarios.

Referencias bibliográficas

- Ávila, R. (1988). *Lengua hablada y estrato social: un acercamiento léxico-estadístico*. NRFH XXXVI, (1), 131-148. <https://dialnet.unirioja.es/servlet/articulo?codigo=30917>
- Ávila Cell, R. (1989). Las palabras de los niños: tercer grado de primaria. En *Estudios sobre español de América y Lingüística Afroamericana* (pp. 19-39). Instituto Caro y Cuervo.
- Bosque, I. (2008). *Las categorías gramaticales. Relaciones y diferencias*. Editorial Síntesis.
- Brown, J. (2008). Research Methods for Applied Linguistics: Scope, Characteristics, and Standards. En A. Davies y C. Elder (Eds.), *The handbook of applied Linguistics* (pp. 476-500).
- Butler, C. (1985). *Statistics in Linguistics*. Basil Blackwell, Oxford.
- Bybee, J. (2007). *Frequency of use and the organization of language*. Oxford University Press.
- Capsada, R. y Torruella, J. (2017). Métodos para medir la riqueza léxica de los textos. Revisión y propuesta. *Verba: Anuario galego de filoloxia*, 44, 347-408. <https://doi.org/10.15304/verba.44.3155>
- Carlino, P. (2003). Alfabetización académica: Un cambio necesario, algunas alternativas posibles. *Educere*, 6(20), 409-420. <http://www.redalyc.org/articulo.oa?id=35662008>
- Cerda, G., Salcedo, P., Pérez, C. y Marín, V. (2017). Futuros Profesores de Matemáticas: Rol de la Disponibilidad Léxica, Esquemas de Razonamiento Formal en Logros Académicos Durante su Formación Inicial. *Formación Universitaria*, 10(1), 33-46. <http://dx.doi.org/10.4067/S0718-50062017000100005>
- Contreras, I. y Aguirre, R. (2020). Escritura académica y desarrollo del potencial epistémico en estudiantes universitarios. *Educere*, 24(77), 103-116. <http://www.redalyc.org/articulo.oa?id=35663240010>
- Coxhead, A. (2000). A New Academic Word List. *TESOL Quarterly*, 34(2), 213-238. <https://doi.org/10.2307/3587951>
- Coxhead, A. y Nation, P. (2001). The Specialized Vocabulary of English for Academic Purposes. En J. Flowerdew y M. Peacock (Eds.), *Research Perspectives on English for Academic Purposes* (pp. 252-267). Cambridge University Press.
- Flower, L. y Hayes, J. (1980). The cognition of discovery: Defining a rhetorical problem. *College composition and communication*, 31(1), 21-32. <https://doi.org/10.2307/356630>
- Godínez, E. y Alarcón, J. (2020). El léxico en la evaluación y en la didáctica de la escritura de textos de literatura. *Lenguaje y Textos*, 51, 69-79. <https://doi.org/10.4995/lyt.2020.11373>
- Hyland, K. y Tse, P. (2007). Is There an “Academic Vocabulary”? *TESOL Quarterly*, 41(2), 235-253. <https://www.jstor.org/stable/40264352>
- Kurdi, M. (2017). Lexical and syntactic features selection for an adaptive reading recommendation system based on text complexity. En *ICISDM '17: Proceedings of the 2017 International Conference on Information System and Data Mining* (pp. 66-69). Association for Computing Machinery.

- Lahuerta, A. (2017). Syntactic complexity in secondary-level English writing: differences among writers enrolled on bilingual and non-bilingual programmes. *Porta Linguarum*, 28, 67-80. <http://dx.doi.org/10.30827/Digibug.54003>
- López, H. (2008). *Documentos de trabajo del proyecto la enseñanza del español en el Mundo Hispánico*.
- López, H. (2011). Los índices de 'riqueza léxica' y la enseñanza de lenguas. En F. J. Santiago-Guervós, H. Bongaerts, J. J. Sánchez y M. Seseña (Coords.), *Del texto a la lengua: La aplicación de los textos a la enseñanza-aprendizaje del español L2-LE* (pp. 15-28). <https://dialnet.unirioja.es/servlet/articulo?codigo=5419218>
- Mancera, A. y Pano Alamán, A. (2014). Las redes sociales como corpus de estudio para el análisis del discurso mediado por ordenador. *Janus. Humanidades Digitales: desafíos, logros y perspectivas de futuro*, 1, 305-315. <https://dialnet.unirioja.es/servlet/articulo?codigo=5181037>
- Martínez, I., Beck, S. y Panza, C. (2009). Academic vocabulary in agriculture research articles: A corpus-based study. *English for Specific Purposes*, 28(3), 183-198. <http://dx.doi.org/10.1016/j.esp.2009.04.003>
- Molina, C. y Sierra, G. (2015). Hacia una normalización de la frecuencia de los corpus CREA y CORDE. *Revista signos. Estudios de Lingüística*, 48(89), 307-331. <http://dx.doi.org/10.4067/S0718-09342015000300002>
- Moreno, A. y Guirao, J. (2008). *Frecuencia y distintividad en el uso lingüístico: casos tomados de la lematización verbal de corpus de distintos registros*. Actas del I Congreso Internacional de Lingüística de Corpus, Universidad de Murcia.
- Nation, I. (2001). *Learning Vocabulary in Another Language*. Cambridge University Press. <https://doi.org/10.1017/CBO9781139524759>
- Ochoa, L. (2021). Complejidad estructural en textos escolares descriptivo-expositivos: estudio de corpus. *Folios*, 54, 91-110. <https://doi.org/10.17227/folios.54-11977>
- Pallotti, G. (2015). A simple view of linguistic complexity. *Second Language Research*, 31(1), 117-134. <https://doi.org/10.1177/0267658314536435>
- Real Academia Española. (s.f.). *Corpus de Referencia del Español Actual*. http://corpus.rae.es/frec/5000_formas.TXT
- Riffo, B., Reyes, F., Novoa, A., Véliz de Vos, M. y Castro, G. (2014). Competencia léxica, comprensión lectora y rendimiento académico en estudiantes de enseñanza media. *Literatura y Lingüística*, 30, 165-180. <http://dx.doi.org/10.4067/S0716-58112014000200009>
- Sabaj, O. (2004). Especificidad, especialización y variabilidad verbal: Una aproximación computacional en estadística léxica. *Revista Signos*, 37(56), 75-89. <http://www.redalyc.org/articulo.oa?id=157013763006>
- Santos, X. (2017). *Métodos de selección léxica aplicados a la enseñanza. El vocabulario fundamental del español* [Tesis de doctorado, Universidad de Salamanca]. <http://hdl.handle.net/10366/137266>

- Templin, M. (1957). *Certain Language Skills in Children: Their Development and Interrelationships*. (Vol. 26). University of Minnesota Press. <https://www.jstor.org/stable/10.5749/j.ctttv2st>
- Valdés-León, G. (2021). Competencia léxica y escritura académica: analíticas de aprendizaje en un curso de escritura universitaria. *Texto Livre: Linguagem e Tecnologia*, 14(1), 1-13. <https://doi.org/10.35699/1983-3652.2021.24560>
- Varela, J., Cabrera, F., Zarabozo, D., Larios, Y. y González, M. (2013). Las 5000 palabras más frecuentes en los libros de texto oficiales de la educación básica en México. *Revista electrónica de investigación educativa*, 15(3), 114-120. http://www.scielo.org.mx/scielo.php?script=sci_arttext&pid=S1607-40412013000300008&lng=es&tlng=es
- West, M. (Ed.). (1953). *A general service List of English words: with semantic frequencies and a supplementary word-list for the writing of popular science and technology*. Longman.

Contribución del autor

Ma. Guadalupe Rojas Corona participó en la integración y selección de la base de datos, así como en la redacción del manuscrito e interpretación de los datos. Marco Antonio Pérez Duran participó en la depuración, procesamiento e interpretación de los datos, al igual que en la redacción del manuscrito y revisión de la bibliografía. Los autores dan aprobación a la versión que se publica en la revista.

Agradecimientos

Los investigadores de este artículo agradecen al centro de prácticas “Clínica de Escritura” de la Facultad de Psicología de la Universidad Autónoma de San Luis Potosí por permitirnos desarrollar la investigación interdisciplinaria mediante el acceso a su población escolar y, con ello, a la base de datos de esta investigación.

Financiamiento

Esta investigación se realizó sin financiamiento.

Conflicto de intereses

Los autores no presentan conflicto de interés

Correspondencia: marco.duran@uaslp.mx

Trayectoria Académica de los autores

Marco Antonio Pérez Durán es doctor en Lingüística por la Universidad Nacional Autónoma de México (UNAM) y catedrático de Lengua en la Facultad de Ciencias Sociales y Humanidades y en la Facultad de Psicología de la Universidad Autónoma de San Luis Potosí, México. Pertenece al Sistema Nacional de Investigadores y al Programa de Fortalecimiento del Profesorado (PRODEP). Su investigación se ha centrado fundamentalmente en lexicología, enfocada en la fraseología, léxico del tabú o altisonante, disponibilidad léxica y análisis del léxico para el contexto escolar.

Ma. Guadalupe Rojas Corona es licenciada en Lingüística Aplicada por la Universidad Autónoma de Tlaxcala. Maestra en Ciencias de Lenguaje por la Benemérita Universidad Autónoma de Puebla y doctorante en Psicología (énfasis en educación) por la Universidad Autónoma de San Luis Potosí. Es becaria del Consejo Nacional de Ciencia y Tecnología. Sus áreas de investigación se centran en los siguientes temas: habilidades lingüísticas y cognitivas que intervienen en los procesos de escritura académica y disciplinar, aprovechamiento académico, desarrollo de intervenciones en escritura y análisis del discurso. Recientemente se ha adentrado, en colaboración con otros colegas, en estudiar formas de intervenir en la comprensión lectora y las habilidades del lenguaje que intervienen en la escritura académica. Además, es cofundadora y responsable de la “Clínica de Escritura”, centro enfocado en dar asesoría personal en los procesos de escritura.