



ANÁLISIS DE LAS FRECUENCIAS LÉXICAS NOMINALES, ADJETIVALES Y ADVERBIALES PRESENTES EN ESCRITOS DE ESTUDIANTES UNIVERSITARIOS DE LA CLÍNICA DE ESCRITURA DE LA FACULTAD DE PSICOLOGÍA DE LA UASLP¹

ANALYSIS OF THE NOMINAL, ADJECTIVAL AND ADVERBIAL LEXICAL FREQUENCIES PRESENT IN WRITINGS OF UNIVERSITY STUDENTS OF THE WRITING CLINIC OF THE FACULTY OF PSYCHOLOGY OF THE UASLP

Marco Antonio Pérez Durán
Universidad Autónoma de San Luis Potosí
marco.duran@uaslp.mx

Ma. Guadalupe Rojas Corona
Universidad Autónoma de San Luis Potosí
rojasgpe@outlook.es

RESUMEN

Este estudio léxico-matemático analiza la frecuencia léxica de vocablos nominales, adjetivales y adverbiales de 68 producciones de estudiantes universitarios que asistieron a la Clínica de Escritura de la Facultad de Psicología de la Universidad Autónoma de San Luis Potosí, México, antes de pandemia por COVID-19. El objetivo es realizar un estudio exploratorio-descriptivo para obtener un panorama general de las frecuencias léxicas de dichos vocablos. El método de análisis es de corte cuantitativo descriptivo de tipo exploratorio contrastivo con un enfoque empírico no transversal. Para ello, se utilizaron el programa libre de cómputo *AntConc*, las fórmulas matemáticas Porcentaje de Vocablos (PV) y Fórmula de Palabras Nacionales (PN) (López Morales, 2011). Los datos obtenidos se contrastaron con el corpus especializado *Academic Vocabulary in Psychology Articles: A Corpus-Based Study* (2020). Los resultados indican que de los 100 vocablos obtenidos en la muestra solo 10 aparecen entre las primeras 50 posiciones de dicho corpus. Esto revela la necesidad por desarrollar intervenciones enfocadas en incrementar el vocabulario disciplinar.

Palabras clave: escritura académica, frecuencia léxica, léxico académico, léxico general, vocablo.

ABSTRACT

This lexical-statistical study analyzes the lexical frequency of nominal, adjectival and adverbial words of 68 productions of college students who attended the Writing Clinic of the Faculty of Psychology of Autonomous University of San Luis Potosi, Mexico, before the COVID-19 pandemic. The objective is to carry out an exploratory-descriptive study to obtain a general overview of the lexical frequencies of these words. The analysis method is quantitative, descriptive, exploratory-contrastive with a non-cross-sectional empirical approach. To do this, the free computing program *AntConc*, the mathematical formulas Percentage of Words (PV) and Formula of Notional Words (PN) (Lopez Morales, 2011) were used. The data obtained were contrasted with the specialized corpus *Academic Vocabulary in Psychology Articles: A Corpus-Based Study* (2020). The results indicate that the 100 words obtained in the sample, only 10 appear among the first 50 positions of this corpus. The data reveals the need to increase the disciplinary vocabulary.

Keywords: academic writing, academic lexicon, general lexicon, lexical frequency, word.

¹ La realización del presente estudio no hubiera sido posible sin la información de los textos de escritura que nos facilitó la Clínica de Escritura de la Facultad de Psicología de la Universidad Autónoma de San Luis Potosí.

Recibido: 12-04-2024
Aceptado: 23-07-2024



1. INTRODUCCIÓN

La escritura es una habilidad compleja que contribuye al fortalecimiento de la comunicación en el ámbito académico. Esta habilidad se conforma de otras competencias básicas, por ejemplo, el conocimiento del vocabulario, la ortografía y la organización discursiva (Montersen, Smith-Lock & Nickels, 2009, como se ha citado en Gómez Vera, Sotomayor, Bedwell, Domínguez & Jéldrez, 2014; Rojas Corona, Silva-Maceda & Romero-Contreras, 2024). Dichas habilidades son antecesoras del desarrollo de la escritura, lo que supone que su dominio impacta en la producción escrita y, por lo tanto, en el desarrollo académico de los estudiantes dentro de la disciplina que han decidido estudiar (Rojas Corona, Silva-Maceda & Romero-Contreras, 2024). Conocer cuál es el vocabulario que los estudiantes emplean puede ayudar a los docentes a desarrollar estrategias para dominar el vocabulario académico y disciplinar. El objetivo de este trabajo es hacer un estudio exploratorio de las frecuencias léxicas de vocablos nominales, adjetivales y adverbiales extraídos de redacciones universitarias de estudiantes que asistieron a la Clínica de Escritura de la Facultad de Psicología de la Universidad Autónoma de San Luis Potosí (UASLP), México, antes de la pandemia COVID-19. La premisa de la que parte este estudio es que los estudiantes que asistieron a este centro ya evidenciaban uso del vocabulario académico propio de su disciplina, pero no un dominio satisfactorio.

La literacidad académica se define como una práctica de interacción social oral o escrita que se ubica y se produce dentro de las actividades académicas. A través de ella, se observa la capacidad del estudiantado de poder interactuar con la información académica a partir de la manipulación, uso y estructuración del discurso académico (Vanegas, 2020). En estos estudios, se define como texto académico a toda producción escrita que se realiza en el ámbito universitario (Martín-Sánchez *et al.*, 2015). Dentro de la escritura académica, el dominio del léxico se percibe como unidad formal, semántica y pragmática que cumple con una función comunicativa

específica. Haché de Yunén (1991) y Sánchez-Avendaño (2005) han afirmado que, en cuanto al vocabulario, los estudiantes manifiestan resultados poco favorables de uso y dominio, los cuales se observan cuando tratan de leer o escribir (transcribir) lo que han leído, por lo que fallan en la transformación del conocimiento. Esto significa una desventaja para el alumnado, ya que es por medio de la escritura que pueden validar los conocimientos de su disciplina.

Para mejorar los procesos de escritura, en muchas universidades en Latinoamérica, se han abierto centros cuyo énfasis está en ayudar a los estudiantes a mejorar sus habilidades básicas y necesarias de escritura, por ejemplo, el dominio del vocabulario académico y el vocabulario de su disciplina (Molina-Natera, 2016; Starkie & Sidotti, 2017). En México, la Facultad de Psicología de la Universidad Autónoma de San Luis Potosí cuenta con un centro de escritura denominado Clínica de Escritura, el cual se fundó en 2017 y tomó como referencia el esquema de trabajo aplicado en la Facultad de Educación de la Universidad de Diego Portales en Chile. Este es un centro estratégico enfocado en la mejora y atención personalizada a estudiantes universitarios que requieran de apoyo educativo en una habilidad tan indispensable y necesaria como la escritura. Este apoyo se logra mediante la autorevisión y corrección de su texto, así como la práctica oral de los recursos del lenguaje y de discurso antes de la escritura.

En este trabajo, se estudia el léxico de las redacciones de un grupo de estudiantes a través de la frecuencia léxica (Fl desde este momento). Según Bargetto Fernández y Riffo Ocares (2019), la Fl se refiere a una palabra que por su uso se vuelve frecuente frente a otro conjunto de palabras que también aparecen en el texto de uso menor o mínimo. A través de la frecuencia léxica, se pueden conocer tres aspectos relevantes sobre el dominio del léxico por parte del estudiante: a) la comprensión sobre la relación interna de las palabras frecuentes en el corpus estudiado y su determinación de las

causas léxicas asociativas que pudieran explicar mejor el comportamiento léxico del hablante y de la lengua en general; b) la idea de que las palabras que alcanzan una frecuencia alta son unidades léxicas iniciales del aprendizaje, por lo que son ítems que frecuentemente aparecerán en el léxico de los informantes, en los textos académicos y en la comunicación discursiva sin importar su posición; y c) la idea de que el léxico es pieza angular en la transmisión y recepción del conocimiento, por lo que con su estudio se pueden explicar mejor las variaciones léxicas de los textos y discursos académicos.

En el siguiente apartado se presenta la base teórica sobre la cual se sustenta este trabajo de corte léxicométrico enfocado en el estudio del vocabulario académico y vocabulario disciplinar.

1.1. MARCO TEÓRICO

En general, la escritura es 1) una habilidad clave para el desarrollo del aprendizaje y la comunicación dentro de las diferentes disciplinas, 2) el vehículo por el cual se accede al conocimiento, 3) una estructura mediante la cual se organiza el pensamiento, 4) un sistema secundario del pensamiento y 5) un sistema mediante el cual se crea, se transmite y se conserva el conocimiento. Además, la escritura se construye con los recursos lingüísticos adecuados a nivel académico, es decir, dominio de la ortografía acentual, literal y puntual, de los aspectos morfológicos y sintácticos, del léxico académico y disciplinar, conocimiento de los géneros discursivos, entre otros (Snow & Ucelli, 2009; Ucelli, Dobbs & Scott, 2013).

Enseñar a leer y escribir en la universidad significa ayudar institucionalmente a los estudiantes a ingresar y participar en las prácticas discursivas propias de las profesiones para las que se preparan y en las que se van a especializar (Narvaja de Arnoux, 2011; Carlino, 2013). Uno de los aspectos relevantes en la producción del texto escrito es el léxico, el cual se define como el conjunto de lexías obtenidas de producciones textuales (Wagner, 2000) que se usan de forma recurrente en los géneros textuales dentro del

ámbito académico (Parodi, 2008; Castillo Fadić & Sologuren, 2020). Este último sirve como mecanismo para el fortalecimiento de la comprensión lectora y la escritura académica (Romero-Contreras, Silva-Maceda & Snow, 2021). Conocer el conjunto de lexías que comúnmente aparecen en determinados géneros textuales puede ayudar a los docentes a decidir sobre el léxico que se debe incorporar en la planeación de sus clases.

Por su parte, la escritura académica es un tipo de escritura que se enseña a partir de los dos últimos grados del nivel educativo básico, y los grados de la educación media, pero que se utiliza más en el ámbito universitario. De acuerdo con Vanegas (2020) esta escritura permite construir el conocimiento, conocer y aprender sobre la disciplina a través de prácticas letradas con características y formas de escribir del área de conocimiento. Esto se debe a que todas las tareas y trabajos que forman parte de la vida académica universitaria están contruidos con base en la escritura del tipo académico. Se reconocen varias características de este tipo de escritura 1) ser clara, entendible y puntual, 2) contar con un propósito académico, 3) tener un tono impersonal, 4) tener argumentos sólidos que validen el propósito del trabajo y, lo principal, 5) generar o renovar el conocimiento en torno al área de estudio. Un texto académico se construye a partir de un léxico del tipo general y del tipo académico o disciplinar. El primero se entiende como el vocabulario que se utiliza para cualquier situación comunicativa, mientras que el segundo se centra más en el ámbito especializado. Para la construcción de un escrito académico se requiere de ambos en el sentido de que el primero ayuda a soportar al segundo.

Aunque en un contexto general, los términos léxico y vocabulario pueden tomarse como sinónimos, para los estudios de corte lexicológico como es el caso de este trabajo, esto no sucede así. El léxico es un término más abstracto que se refiere a la totalidad de palabras que el locutor (escritor en este caso) sabe en un momento determinado, de ahí su valor global y general de la lengua, mientras que el término vocabulario alude a los vocablos propios

de la lengua que se han usado en un acto de habla particular, por tanto, es un concepto más material (Madrigal & Vargas, 2016). Dicho de otro modo, el vocabulario es un hecho concreto que sirve para construir un texto, su unidad de análisis es el *type* o palabra diferente, mientras que el léxico es algo universal que pertenece a la lengua, su unidad de estudio es la palabra repetida o *token*.

Los *types* y *tokens* son unidades que se pueden encontrar en un corpus, o bien que pueden conformar uno. Un corpus se define como el conjunto general de las muestras orales o escritas que se han recogido en un contexto y momento determinado y que pueden servir para una investigación. Para estudiar los corpus académicos se utilizan pruebas de riqueza léxica en las que los *types* y los *tokens* son las unidades de análisis. Estas pruebas son una fuente de información valiosa para los especialistas porque se analiza el desarrollo cuantitativo del léxico de cualquier corpus, en este caso, de los estudiantes universitarios. Tomando en cuenta los resultados de estas pruebas se puede obtener la riqueza léxica que es el grado de madurez y de dominio léxico general de un estudiante medidos en su producción escrita por el porcentaje de vocablos de contenido nocional (Álvarez Martín, 2019) y por el intervalo de palabras y vocablos de contenido nocional (Cuba Vega, 2007).

En el caso del léxico académico, el corpus más conocido es la lista de vocabulario académico elaborada y evaluada por Coxhead en el año 2000 (para el contexto español esta lista fue traducida y adaptada por Juan C. Ripoll). Dicho corpus se obtuvo de muestras que pertenecían a diferentes disciplinas universitarias, entre 15 a 28 disciplinas diferentes. Como resultado la autora extrajo las 2.000 palabras más frecuentes del inglés que se repetían por lo menos 100 veces en el corpus. Sin embargo, esta lista de vocabulario académico no presenta la clasificación de vocabulario por disciplina que es otro tipo de estudio que se puede realizar.

El vocabulario disciplinar pertenece al ámbito de la especialización académica. En este tipo de léxico, se pueden encontrar palabras cuyo significado cambie en relación con la disciplina que haga uso de ellas. Un corpus de esta naturaleza es *Academic Vocabulary in Psychology Articles: A Corpus-Based Study* que está compuesto por 143 millones de palabras. Este se obtuvo de 20.000 textos sobre investigaciones de psicología, por lo que pertenece a un solo ámbito de estudio. Este último se ha utilizado para esta investigación como corpus de contraste debido al tipo de población que permitió obtener la muestra del corpus de este estudio, escritos realizados por estudiantes de la Facultad de Psicología.

2. METODOLOGÍA

La presente investigación se enmarca en los estudios léxico-matemáticos de tipo cuantitativo descriptivo de tipo exploratorio contrastivo, con un enfoque empírico no transversal por la naturaleza del estudio. En este se busca cuantificar y describir las voces nominales, adjetivales y adverbiales que se obtuvieron de textos escritos por estudiantes de una universidad estatal y pública de México.

2.1. MUESTRA

La muestra está compuesta por 68 escritos realizados por estudiantes que asistieron a tomar sesiones de asesoría en escritura académica en el centro de prácticas Clínica de Escritura de la Facultad de Psicología de la Universidad Autónoma de San Luis Potosí. La recolección de estos escritos se realizó en dos momentos. El primero se refiere a los textos que se recogieron antes de que los estudiantes tomaran sus sesiones de asesoría, 34 textos pre-asesoría. Una vez que los estudiantes terminaron dichas sesiones, ellos elaboraron otro escrito denominado textos post-asesoría, 34 en total. La muestra se obtuvo durante enero y febrero de 2020, antes del inicio de la pandemia de COVID-19 y el cierre de escuelas en México. Con los datos obtenidos de

la muestra pre y post-intervención se conformó el corpus de esta investigación. Para su estudio, se realizó el análisis del corpus por relación: *types* y *tokens*, lematización de las unidades léxicas y N-grama de los *types*.

2.2. LA CLÍNICA DE ESCRITURA

La Clínica de Escritura es un centro de prácticas que se encuentra ubicado dentro de la Facultad de Psicología de la Universidad Autónoma de San Luis Potosí. Este centro cuenta con asesores, alumnos de la licenciatura en Psicopedagogía, con énfasis en lengua, que realizan sus prácticas profesionales. Estos asesores son capacitados en procesos de escritura académica y recursos del lenguaje académico para brindar apoyo en el desarrollo de las habilidades de escritura académica mediante sesiones individuales de 50 minutos. Para poder obtener este tipo de asesoría, el alumno interesado la solicita mediante la página de Facebook². Una vez hecho esto, se le agenda una fecha de asistencia. Durante la primera cita los estudiantes deben realizar un escrito *in situ* a partir de la pregunta de elicitación: *¿Qué has aprendido sobre tu carrera?* De esta forma se busca que el alumno escriba un texto real. Este sirve de base para la detección de los errores de escritura. Por lo regular, la extensión de este debe ser mayor a una cuartilla (entre 200 y 250 palabras en promedio).

Una vez obtenido el texto, se somete a una evaluación ciega por tres asesores con el propósito de focalizar y jerarquizar de forma objetiva los problemas de escritura y planear las asesorías de acuerdo con los problemas más relevantes. Al terminar las sesiones, se solicita al estudiante redactar, nuevamente, un texto, para ello se toma de base la misma pregunta del inicio. A este paso se le denomina evaluación post-asesoría.

2.3. HERRAMIENTAS DE ANÁLISIS

Para el análisis de los datos, se utilizó el programa *AntConc*³, software de acceso libre. Con este programa, se identificaron los *types* y *tokens* de la muestra. La

² https://www.facebook.com/ClinicaEscrituraUASLP/?locale=es_LA

³ <https://www.laurenceanthony.net/software/antconc/>.

relación entre *types* y *tokens* se denomina *ratio* y se representa con las siglas TTR. La relevancia de establecer la relación, *ratio* radica en que esta es necesaria para determinar la riqueza léxica de la muestra. Para encontrar la riqueza léxica, se utiliza la fórmula: $TTR = \text{Número de } types * 100 / \text{número de } tokens$ (Capsada Blanch & Torruella Casañas, 2017). Como se sabe, el TTR presenta la desventaja de que el cálculo depende de la longitud del texto, ya que existe una relación inversa entre la cantidad de *tokens* y *ratio*. Además de esta fórmula, se utilizó la de Honoré (1979), la cual se puede relacionar con la Ley de Zipf (1949) que indica que con cada aparición de una palabra esta desarrollará un patrón de frecuencia exponencial en un contexto de uso en relación con la palabra que le antecede (Rodríguez Velásquez, Prieto Bohórquez, Ortiz Salamanca, Bautista Charry, Bernal & Avilán Vargas, 2006). A continuación, se presenta la ejemplificación de la fórmula de Honoré (Imagen 1).

$R = \frac{100 \log N}{1 - (V_1 / V)}$	donde	<p>R = riqueza léxica N = n° palabras en el texto V = n° de formas en el texto V₁ = n° de palabras usadas una sola vez</p>
--	-------	---

IMAGEN 1: Ejemplificación de fórmula de Honoré

El análisis que se presenta a través de esta fórmula parte de la idea de que, en la palabra con más apariciones, se tiene una frecuencia igual al doble de la frecuencia de la segunda palabra y así sucesivamente. Los resultados obtenidos se organizan en tablas de rasgos y frecuencias. Una vez aplicada la fórmula se obtienen los lemas, es decir, la unidad autónoma que constituye el léxico de un idioma.

Otro análisis que se obtuvo mediante el software *AntConc* fue la organización de la información en N-gramas o *clústers* de los *types*. Un N-grama es un conjunto de *n* elementos consecutivos que se relacionan a partir de las relaciones sintácticas entre las unidades léxicas. Estos N-gramas pueden incluir palabras, números, símbolos. En este trabajo, se utilizaron los N-gramas para comprender las relaciones asociativas estructurales entre las palabras del

corpus. Para ello se tomó un ítem que sirve de núcleo con la cual se asocia el resto de las respuestas. A esto se le denominó familia léxica. El concepto de familia léxica es relevante en estos estudios porque funciona como la precursora de la organización de la información.

Además de lo anterior, se aplicaron las fórmulas Porcentaje de Vocablos (pv) y de Palabras Nocionales (pn), propuestas por López Morales (2011). Para encontrar el resultado de la fórmula Porcentaje de Vocablos (pv), se requiere que se divida el total de vocablos (v) entre el total de las unidades léxicas comprendidas en el texto (n). Después el resultado de la división se multiplica por 100, su representación es $pv = v \cdot 100 / n$. Esta fórmula se usa para encontrar el porcentaje de vocablos, es decir, el indicador grueso para la riqueza léxica. Por último, se hizo el contraste entre los vocablos del corpus de estudiantes universitarios con el de *Academic Vocabulary in Psychology Articles: A Corpus-Based Study* (Xodabande & Xodabande, 2020). Para este análisis, se tomaron las diez unidades léxicas más frecuentes del corpus de estudiantes y se buscaron por su posición en el corpus de corte psicológico. La posición de estas es una característica relevante en el conocimiento del vocablo, ya que permite conocer la activación del léxico de los informantes de forma clara. Por último, el corpus se redujo a dos subclasificaciones: 1) léxico general, y 2) léxico especializado.

2.4. PROCEDIMIENTO DE ANÁLISIS DEL CORPUS

Una vez obtenidas las muestras de escritura pre y post-intervención, se conformó el corpus de este estudio y se procedió a su análisis. Para ello, se llevaron a cabo los siguientes pasos:

- 1) Se procedió a la anonimización de los textos según el reglamento de ética de la Facultad de Psicología.⁴
- 2) Posteriormente, se segmentó cada una de las redacciones en unidades léxicas y se creó una sola lista de palabras.

⁴Este reglamento se puede consultar en <https://www.psicologia.uaslp.mx/Paginas/Servicios/3737#gsc.tab=0>.

- 3) Mediante el proceso de lematización, se separaron las palabras funcionales (sin contenido semántico) de aquellas que tuvieran contenido semántico. Una vez conformado este conjunto de palabras, se quitó las voces verbales, ya que esta submuestra forma parte de otro estudio, por lo que solo se conservaron voces adverbiales, adjetivales y nominales.
- 4) Posterior a la selección de las voces nominales, adjetivales y adverbiales de los escritos donados por los estudiantes que asistieron a la Clínica de Escritura, estas se organizaron por familias léxicas, lo cual facilitó el conteo final y puso en evidencia todas las relaciones entre vocablos. Cabe mencionar que no se consideraron para el análisis las preposiciones, onomatopeyas, artículos y conjunciones.
- 5) Después, se hizo la corrección tanto de forma como de contenido de cada unidad léxica que compone el corpus. Se encontraron y se corrigieron errores de escritura y de acentuación comunes en las unidades léxicas que pertenecen a la escritura hecha a lápiz y papel. La depuración de la información tuvo una duración de 200 días por el tipo de procedimiento manual que se implementó.
- 6) Posterior a la depuración de la información, se organizó el corpus en Excel y en formato txt. En este último, se pudieron quitar acentos del corpus para mejorar el análisis de la información en el software *Ant-Conc*. Después de procesar la información en este software, se utilizó Excel para ingresar la fórmula de Honoré (1979) y poder generar los resultados que se exponen. Todo el procedimiento de análisis tuvo una duración de 280 días de trabajo.
- 7) Se compararon los datos obtenidos con el corpus especializado de *Academic Vocabulary in Psychology Articles: A Corpus-Based Study* (2020). Para ello, se tomaron las diez unidades léxicas más frecuentes que aparecieron en ambos corpus.
- 8) La información se organizó en tablas para un mejor manejo de los datos y para poder hacer más evidente el análisis.

3. RESULTADOS

Para identificar los resultados generales, se presenta la siguiente tabla con los datos obtenidos del corpus distribuido en total de palabras (tp), vocablos (v), *types tokens ratio* (ttr), valor de riqueza léxica (vrl) y palabras de contenido nocional (iat), que aparecieron durante el análisis. Esto se puede observar en la Tabla 1.

Total	tp	v	ttr	vrl	iat
Muestra general	17,167	2,306	13.43	6.32	7.44
Preasesoría	16,890	2,290	13.55	6.28	7.49
Postasesoría	17,050	2,297	13.47	6.23	7.49

TABLA 1: Resultados generales del análisis del corpus. *Elaboración propia*

Como se aprecia en la Tabla 1, los resultados muestran que el número total de palabras del corpus asciende a 17,167. Al analizarlas por pre-asesoría y post-asesoría se obtuvo lo siguiente: el total de términos por pre-asesoría es de 16,890 y post-asesoría es de 17,050. Al comparar los resultados se encontró una diferencia a favor del corpus general de 277 unidades léxicas, en relación con los resultados pre-asesoría y de 117, en relación con post-asesoría.

Sobre el número de vocablos, el corpus ascendió a 2,306. Al contrastar este resultado, el número total de términos pre-asesoría es de 2,290 y el de post-asesoría es de 2,297. En relación con el análisis ttr, parámetro primordial del análisis de la variación léxica, la muestra general es de 13.43 del corpus. Vinculado al análisis del ttr, se encuentra el intervalo de aparición de contenido nocional (iat) que refleja, como se ha mencionado, las palabras nocionales en el texto. El corpus general presenta un iat de 7.44. En contraste, la muestra pre y post-asesoría tiene un iat de 7.49. Esto significa que el corpus analizado presenta un alto nivel de palabras nocionales.

Además de los indicadores anteriores, el promedio de palabras asciende a 14.24, la longitud del corpus es de 8.61 términos y la media es de 10. Sobre el análisis de lemas o vocablos lematizados del corpus, hay 1,153 vocablos lematizados y 2,324 vocablos totales. Al comparar estos dos resultados, hay una reducción de 1,171 vocablos (50%) de la totalidad. A su vez, el corpus lematizado se dividió en 682 nominales (58.2%), 386 adjetivos (32.9%) y 85 adverbios (7.2%).

La información de los vocablos más frecuentes por distribución de las diez respuestas se organiza usando las siguientes clasificaciones: lugar de aparición (n), vocablo (v), frecuencia (f), frecuencia acumulada (Fi) y la frecuencia absoluta (Hi). Todo esto se expone en la Tabla 2.

N	Vocablo	Frecuencia	Fi	Hi
1	<i>Psicología</i>	343	343	0.03695
2	<i>No</i>	342	685	0.02807
3	<i>Niño</i>	278	963	0.02232
4	<i>Mucha</i>	260	1223	0.02091
5	<i>Importancia</i>	208	1431	0.02091
6	<i>Carrera</i>	197	1628	0.01864
7	<i>Su</i>	196	1824	0.01506
8	<i>Materia</i>	193	2017	0.01441
9	<i>Todo</i>	179	2196	0.01355
10	<i>Mi</i>	178	2374	0.01355

TABLA 2: Orden de la frecuencia por distribución de los diez vocablos más frecuentes. Elaboración propia

En las primeras diez respuestas, hay cinco nominales, cuatro adjetivos y un solo adverbio. En relación con los nominales, el vocablo *psicología* acumuló el mayor número de frecuencias con 343, le siguió el adjetivo *mucha* con 260 y el adverbio *no* con 342 apariciones. Los mismos resultados se aprecian tanto en la frecuencia acumulada como en la absoluta. La categoría que menos apariciones presentó en las 100 palabras de las listas fue la adverbial, aunque el adverbio negativo no haya aparecido en segundo lugar de la lista de vocablos, de 2,306. La categoría que presentó mayor número de vocablos es la nominal y después la adjetival. En la Tabla 3, se presenta la distribución de las diez respuestas según sea su categoría gramatical.

	Nominales	Adjetivales	Adverbios
1	<i>Psicología</i>	<i>Mucha</i>	<i>No</i>
2	<i>Niño</i>	<i>Su</i>	<i>Muy</i>
3	<i>Importancia</i>	<i>Todo</i>	<i>Sí</i>
4	<i>Carrera</i>	<i>Mi</i>	<i>Siempre</i>
5	<i>Materia</i>	<i>Este</i>	<i>Bien</i>
6	<i>Infante</i>	<i>Conocida</i>	<i>Ahora</i>
7	<i>Personas</i>	<i>Esta</i>	<i>Aquí</i>
8	<i>Diferencia</i>	<i>Sus</i>	<i>Actualmente</i>
9	<i>Patología</i>	<i>Otros</i>	<i>Abí</i>
10	<i>Aprendizaje</i>	<i>Cada</i>	<i>Principalmente</i>

TABLA 3: Distribución de las diez respuestas según su categoría gramatical.
Elaboración propia

En el conjunto de vocablos nominales, aparecieron términos como: *psicología*, *carrera*, *materia*, *patología* y *aprendizaje*. Estas respuestas aparecieron

como esperadas, ya que estos nominales están vinculados al área de psicología como *patología* o *psicología*, o bien pertenecen estas unidades al vocabulario académico: *materia*, *carrera* o *aprendizaje*. En cuanto a la aparición de los adjetivos, estos se clasifican por la ubicación en relación con el nominal, la mayoría de ellos se usan antepuestos al nominal: adj+nominal. Solo dos adjetivos presentan un uso menos restrictivo, el adjetivo *mucha* por lo regular acompaña a un adverbio o nominal y el vocablo *conocido(a)* es una palabra deverbal. Sobre los adverbios, estos pueden ser de tipo modal (*muy*, *siempre*, *bien*, *principalmente*); o de negación (*no*), afirmación (*sí*), temporal (*ahora*, *actualmente*) y lugar (*aquí*, *ahí*).

Como se mencionó en la metodología, se sometió el corpus de vocablos a un nuevo procesamiento estadístico para obtener los vocablos lema de sus familias léxicas. Así pues, de los 2,306 se redujo a 1,152 vocablos lematizados (50 %) de 199 familias léxicas. Para el análisis, se presentan, en la Tabla 4, los seis vocablos lema que aparecen en la familia léxica con mayor producción.

	Vocablo lematizado	Vocablos	Familia léxica
1	Psicología	21	Psicopedagogía 81, psicopatología 75, psicólogo 61, psicóloga 26, psicopedagogo 24, psicológico 17, psíquica 9, psicosis 8, psicótica 6, psíquico 6, psicométrica 6, psicopedagógico 5, psicótico 4, psicoterapia 3, psicomotora 2, psicóticas 2, psicopatológico 2, psicoanalítica 2, psicoterapia 2, psicopatología 1, psicobiólogo 1.
2	Mucho	6	Mucho 128, mucha 68, muchísimo 8, muchísimas 2, muchísima 1, muchísimo 1.
3	Patología	6	Patología 74, patológico 37, patológicamente 7, patológica 4, patólogo 3, patologías 1.
4	Profesional	6	Profesional 51, profesor 21, profesión 8, profesora 6, profesionalidad 2, profesionalmente 2.
5	Conocimiento	4	Conocimiento 119, conocida 3, conocido 3, conociendo 1.
6	Todo	4	Todo 89, todos 44 toda 13, todas 32.

TABLA 4: Familia léxica de los seis vocablos lemas con mayor producción de vocablos.
Elaboración propia

La información que se desprende del cuadro anterior permite observar las relaciones entre el vocablo lema y el vocablo. Nótese que las familias léxicas más productivas del corpus están vinculadas al léxico de tipo nominal: *psicología*, *patología* y *conocimiento*, adjetival: *profesional* y *todo* y adverbial: *mucho*. De estas familias léxicas, las más productivas fueron: *psicología* con 21 vocablos, *mucho*, *patología* y *profesional* con seis y *conocimiento* y *todo* con solo dos unidades.

Sobre el análisis de los N-gramas o clúster, el número total asciende a 9,072 de un total de 17,166. El clúster nominal-nominal fue el que más apareció en el corpus con 5,789 ocasiones, seguido del adjetivo-nombre con 2,779, adjetivo-adjetivo o adverbio con 378 y adverbio-adverbio con 126. En la Tabla 5, se ubican los clústeres que más apariciones tuvieron en el corpus.

P	Relación gramatical	Frecuencia	Ejemplos
1	N-n	0.097	<i>niños adolescentes</i> <i>infancia adolescencia</i> <i>enseñanza aprendizaje</i> <i>comportamiento humano</i> <i>énfasis lengua</i>
2	Adj-n N-adj	0.086	<i>estas etapas</i> <i>esta carrera</i> <i>mi carrera</i> <i>objeto amoroso</i>
3	Adv-adj Adj-adv	0.042	<i>no sólo</i>

TABLA 5: Clústeres con más apariciones en el corpus de esta investigación.
Elaboración propia

En el análisis, se divide el corpus de estudiantes en dos grupos: léxico general y léxico académico. En este último, se engloban unidades léxicas denominadas especializadas. Para este análisis, se tomaron de nuevo el número de vocablos lematizados de 1,152. Se observó que el vocabulario general asciende a 807 (70%) vocablos y el académico a 345 (29.7%). La organización de esta información se puede apreciar en la Tabla 6.

Vocabulario	Ejemplos
General	No, niño, mucha, importancia, su, mi(s), persona, diferencia, esto, otro, alguna, muy, cada, formación, desarrollo, observable, normal, manera, algo, nueva.
Académico	Psicólogo(a), carrera, materia, aprendizaje, lengua, humano, profesión, semestre, teoría, maestro(a), estudio, comportamiento, conducta, contexto, adolescente, clases, área, tema, educación, variable, factor, metodología, enseñanza, clínica, ideales, literaria, alumno(a), historia, curso, característica, licenciatura, escolar, proceso, concepto, experiencia, universidad, base, académica, autor(a), redacción, prueba, función, escrita, herramienta, ámbito, comunicación, enfermedad, identidad, institución, tarea, especializado, estructura, lectura, paciencia, frecuencia, texto, docente, duración, mental, atención, crítica, dinámica, ejemplo, interés, escuela, proyecto, arqueología, sistema, cultura, plan, salud, entorno, perspectiva, técnica, asignatura, artificial, conclusión, duda, facultad, análisis, búsqueda, consecuencia, doctor, economía, información, mente, suma, jurídica, pensamiento, sujeto, biología, nivel, percepción, personalidad, relevancia, castración, genética, interacción, manual.

TABLA 6: Representación de las familias léxica de este corpus de investigación.
Elaboración propia

Por último, para el análisis contrastivo, se utilizó el corpus de vocabulario especializado de psicología de *Academic Vocabylary in Psychology Articles: A Corpus-Based Study* (2020) que contiene 1,526 términos y es parecido al corpus de estudiantes universitarios. En este corpus, de los 100 vocablos contrastados, 10 vocablos de los estudiantes aparecen en el listado especializado. Esto se puede observar en la Tabla 7.

P	Corpus estudiantes	Corpus especializado (<i>Academic Vocabylary in Psychology Art</i>)
1	Psicólogo (1)	Psychological (75)
2	Conducta (48)	Conducted (61)
3	Contexto (50)	Context (55)

(Continúa)

P	Corpus estudiantes	Corpus especializado (<i>Academic Vocabylary in Psychology Art</i>)
4	<i>Variables</i> (65)	<i>Variables</i> (65)
5	<i>Factor(es)</i> (69)	<i>Factors</i> (29)
6	<i>Metodología</i> (70)	<i>Methods</i> (62)
7	<i>Tiempo</i> (71)	<i>Temporal</i> (78)
8	<i>Clínica</i> (76)	<i>Clinical</i> (67)
9	<i>Características</i> (87)	<i>Features</i> (95)
10	<i>Proceso</i> (96)	<i>Process</i> (54)

TABLA 7: Análisis contrastivo del corpus universitario y el corpus de vocabulario especializado de psicología de *Academic Vocabylary in Psychology Articles: A Corpus-Based Study* (2020). Elaboración propia

4. CONCLUSIÓN

Sobre los resultados que se han encontrado a lo largo de esta investigación, se ha observado que independientemente de la extensión de los escritos que conformaron el corpus, los vocablos lema guardan un parecido con la lista de palabras del corpus *Academic Vocabylary in Psychology Articles: A Corpus-Based Study*, aunque no en el mismo orden y frecuencia. Llama la atención que los vocablos con más aparición por frecuencia, frecuencia acumulada y absoluta sean unidades pertenecientes a categorías como adverbios y adjetivos, unidades que cumplen la función de determinantes. Después aparecen con alto grado de frecuencia términos de categoría nominal, pero de vocabulario general como *niño*, *importancia*, *carrera* y *materia* y solo después aparecen vocablos que pertenecen al vocabulario

académico y muy pocos al disciplinar como *conocimiento*, *psicología*, *pato-
logía*, entre otras. Se observó una disminución del 10% de la totalidad del
corpus de este trabajo en relación con otras investigaciones realizadas sobre
el mismo tema (Read, 2000; Vine Jara, Fuentes Riffo, Neira Martínez & Poza
Molina, 2021).

La escritura académica es una habilidad compleja que se solicita al estu-
diantado universitario, su estructura está basada en un dominio de léxico de
corte académico y de corte disciplinar. El conocimiento de este es básico
para que los estudiantes puedan comprender lo que leen y para escribir
sobre un tema. Cuando el docente fortalece el dominio del vocabulario aca-
démico y el vocabulario disciplinar ayuda al estudiante en el proceso de
adquisición y comprensión de las familias de palabras y del desarrollo de
una escritura académica.

Como se ha comentado en este trabajo, la escritura académica, al ser un
proceso complejo, se debe fortalecer desde el ámbito de la práctica de las
habilidades que la componen, entre estas, la adquisición del vocabulario. La
corrección del sistema y la revisión de las normas se harán en un momento
posterior, es decir, en una revisión de estilo. Sobre el estudio que se llevó
a cabo en este trabajo, la importancia que tiene el léxico es más que evi-
dente porque este es la unidad de análisis y comunicación que se utiliza
para todo.

Así, los resultados de esta investigación constituyen un aporte valioso,
puesto que, no se han encontrado trabajos de léxico universitario donde se
hayan obtenido y analizado tres categorías gramaticales (nominales, adjeti-
vales y adverbiales) después de la pandemia por COVID-19 y mucho menos
investigaciones que hayan sido obtenidas de textos como los que se presen-
taron durante el análisis de este artículo. Si bien es cierto que los resultados
han aparecido después del COVID-19, todavía hay mucha información que
sigue fortaleciendo la descripción del léxico dentro de la escritura académica

universitaria, nivel educativo en el que la cantidad de estudios, por lo menos en México, todavía no son suficientes.

El conocimiento de este tipo de información es significativo, ya que la presencia alta de los nominales es característica de la escritura en general. Este resultado es relevante para efectos didácticos porque se observan que, de las seis palabras más frecuentes, dos pertenecen al ámbito disciplinar (*psicología* y *patología*), dos al léxico académico (*profesional*, y *conocimiento*) y dos más al léxico general (*mucho* y *todo*). En este caso, la tendencia es que el léxico general sea más que el académico como se aprecian en los resultados. Sobre el léxico académico y el comparativo que se hace con el especializado de *Academic Vocabulary in Psychology Articles: A Corpus-Based Study*, se observó como de las 10 palabras, *psicología* en el corpus de los estudiantes ocupa la primera posición, mientras que en otro corpus aparece en el lugar 75. Esto indica que existen otras palabras que están por debajo de esta posición en el corpus especializado. Sobre el resto de las unidades, la mayoría ocupa posiciones por encima del lugar 50.

Los resultados que se han visto en este artículo son un preámbulo para los interesados en la didáctica de la escritura disciplinar, o bien, en la producción de materiales administrativos de esta área. Además, los vocablos académicos identificados podrían relacionarse con los textos disciplinares por semestre, lo que ayudaría a la integración y dominio de estas respuestas como parte del repertorio léxico y como parte de los géneros académicos propios de las disciplinas.

En este sentido, esta investigación puede ser útil para distintas áreas: lingüística, lingüística aplicada, psicología, psicología social, psicología métrica, entre otras. Respecto a esto último, esta investigación cobra todavía más relevancia, pues al tratarse de un estudio del tipo cuantitativo descriptivo con enfoque exploratorio contrastivo realizado con base en textos reales producidos por estudiantes que asistieron a la Clínica de Escritura genera pistas en torno a que los estudiantes presentan vocabulario de su disciplina y que solo faltaría enseñarles a utilizarlo adecuadamente.

REFERENCIAS BIBLIOGRÁFICAS

- Álvarez Martín, L. (2019). Índices de riqueza léxica en estudiantes de 11 y 16 años de La Habana. *Universidad de La Habana*, 287, 267-286. <https://doi.org/10.4067/S0718-93032019000100341>
- Bargetto Fernández, M. Á. & Riffo Ocares, B. (2019). El reconocimiento de palabras y el acceso léxico: revisión de modelos y pruebas experimentales. *Boletín de filología*, 54(1), 341-361.
- Capsada Blanch, R. & Torruella Casañas, 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. (2013). Alfabetización académica diez años después. *Revista mexicana de investigación educativa*, 18(57), 355-381.
- Castillo Fadić, M. N. & Sologuren, E. (2020). Léxico frecuente, riqueza léxica y estereotipos sobre la lectura de profesores en formación. *Logos*, 30(1), 69-85. <https://doi.org/10.15443/RL3006>
- Coxhead, A. (2000). A new academic word list, *TESOL quarterly*, 34(2), 213-238.
- Cuba Vega, L. (2007). Riqueza léxica en escolares de primaria de Ciudad de La Habana. En M. Domínguez Hernández (ed.), *La lengua en Cuba. Estudios* (pp. 283- 308). Universidad de Santiago de Compostela.
- Gómez Vera, G., Sotomayor, C., Bedwell, P., Domínguez, A. M. & Jéldrez, E. (2014). Análisis de la calidad del léxico y su relación con la calidad de la escritura en niños chilenos de cuarto año de educación primaria. *Documento de Trabajo*, 16, 1-19.
- Haché de Yunén, A.M. (1991). Aportes de las pruebas de riqueza léxica a la enseñanza de la lengua materna. En H. López Morales (ed.), *La enseñanza del español como lengua materna* (pp. 47-60). Universidad de Puerto Rico.
- Honoré, A. (1979). Some Measures of Richness of Vocabulary. *ALLC Bulletin*, 7(2), 172-177.
- López Morales, H. (2011). Los índices de riqueza léxica y la enseñanza de lenguas. En F. J. de Santiago-Guervós, H. Bongaerts, J. J. Sánchez Iglesias & M. Seseña Gómez (coords.), *Del texto a la lengua: La aplicación de los textos a la enseñanza-aprendizaje del español L2-LE* (pp. 15-28). Asociación para la Enseñanza del Español como Lengua Extranjera-ASELE.
- Madrigal, M. & Vargas, E. (2016). Índice de riqueza léxica en redacciones escritas por estudiantes universitarios. *Kañina*, 40, 139-147. <https://doi.org/10.15517/rk.v40i3.29260>
- Martín Sánchez, T., Pascual Escagedo, C., & Puigdevall Balafuy, N. (2015). Nuevas metodologías para el aprendizaje de la conversación en ELE a través de las TIC.
- Molina-Natera, V. (2016). Los centros de escritura en Latinoamérica: consideraciones para su diseño e implementación. En G. Bañales, M. Castelló & N. Vega (eds.), *Enseñar a leer y escribir en la educación superior. Propuestas educativas basadas en la investigación* (pp. 339-362). Instituto de Investigación y Desarrollo Educativo.

- Mortensen, L., Smith-Lock, K., & Nickels, L. (2009). Text structure and patterns of cohesion in narrative texts written by adults with a history of language impairment. *Reading and Writing*, 22, 735-752.
- Narvaja de Arnoux, E. (2011). La escritura de tesis: apoyos institucionales y propuestas pedagógicas. En A. Vázquez, M. del C. Novo, I. Jakob & L. Pelliza (comps.), *Lectura, escritura y aprendizaje disciplinar* (pp. 25- 40). Subsede Universidad Nacional de Río Cuarto de la Cátedra unesco de Lectura y Escritura, Facultad de Ciencias Humanas, Centro de Investigación y Enseñanza de la Lectura y la Escritura.
- Parodi, G. (2008). Géneros académicos y géneros profesionales. En G. Parodi (ed.), *Accesos discursivos para saber y hacer* (pp. 75-96). Ediciones Universitarias de Valparaíso.
- Read, J. (2000). *Assessing Vocabulary*. Cambridge University Press. <https://doi.org/10.1017/CBO9780511732942>
- Rodríguez Velásquez, J., Prieto Bohórquez, S., Ortíz Salamanca, L., Bautista Charry, A., Bernal, P., & Avilán Vargas, N. (2006). Diagnóstico Matemático de la monitoria fetal aplicando la ley de Zipf-Mandelbrot. *Revista de la Facultad de Medicina*, 54(2), 96-107.
- Rojas Corona, M. G., Silva-Maceda, G. & Romero-Contreras, S. (2024). Relación del lenguaje académico y la comprensión lectora en la escritura académica en educación superior en pandemia. *RIDE Revista Iberoamericana para la Investigación y el Desarrollo Educativo*, 14(28). <https://doi.org/10.23913/ride.v14i28.1813>
- Romero-Contreras, S., Silva-Maceda, G. & Snow, C. E. (2021). Vocabulario académico y habilidades de lenguaje académico: predictores de la comprensión lectora de estudiante de primaria y secundaria en México. *Revista Pensamiento Educativo*, 58(2), 1-16. <https://doi.org/10.7764/PEL.58.2.2021.4>
- Sánchez-Avenidaño, C. (2005). Los problemas de redacción de los estudiantes costarricenses: una propuesta de revisión desde la lingüística del texto. *Revista de Filología y Lingüística*, 31(1), 267-295. <https://doi.org/10.15517/rfl.v31i1.4419>
- Snow, C. E. & Uccelli, P. (2009). The challenge of academic language. En N. Torrance & D. R. Olson (eds.), *The Cambridge Handbook of literacy* (pp. 112-133). Cambridge University Press. <https://doi.org/10.1017/CBO9780511609664.008>
- Starkie, E. I. G. & Sidotti, P. T. (2017). La evolución histórica de los Centros de Escritura Académica The Historic Evolution of Centres of Academic Writing. *Revista de educación*, 378, 9-29.
- Uccelli, P., Dobbs, C. L. & Scott, J. (2013). Mastering academic language: Organization and stance in the persuasive writing of high school students. *Written Communication*, 30(1), 36-62. <https://doi.org/10.1177/0741088312469013>
- Vanegas, D. A. M. (2020). La escritura académica y la universidad: Una aproximación desde el discurso ideológico. *Revista Educación Superior y Sociedad (ESS)*, 32(1), 115-140.

- Vine Jara, A., Fuentes Riffo, K., Neira Martínez, A. & Poza Molina, C. (2021). Riqueza léxica en textos escritos por aprendientes de ELE. una contribución a las nociones específicas del plan curricular del instituto cervantes. *Nueva Revista del Pacífico*, 75, 201-226. <https://doi.org/10.4067/S0719-51762021000200201>
- Wagner, C. (2000). La diversidad léxica en la comunicación. *Revista Documentos Lingüísticos y Literarios UACH*, 23, 67-73.
- Xodabande, I. & Xodabande, N. (2020). Academic Vocabulary in Psychology Research Articles: A Corpus-Based Study. *MEXTESOL Journal*, 44(3), 1-21.