

**RIQUEZA LÉXICA EN TEXTOS ESCRITOS POR APRENDIENTES DE ELE.
UNA CONTRIBUCIÓN A LAS NOCIONES ESPECÍFICAS DEL PLAN
CURRICULAR DEL INSTITUTO CERVANTES**

**LEXICAL RICHNESS ON WRITTEN TEXTS BY FOREIGN SPANISH
LANGUAGE LEARNERS: A CONTRIBUTION TO THE SPECIFIC NOTIONS
OF THE CURRICULAR PLAN OF THE CERVANTES INSTITUTE**

Ana Vine Jara
Universidad de Concepción. Chile
avine@udec.cl

Karina Fuentes Riffo
Universidad de Concepción. Chile
kafuente@udec.cl

Angie Neira Martínez
Universidad de Concepción. Chile
angieneira@udec.cl

Carlos Poza Molina
Universidad Católica de la Santísima Concepción. Chile
cpoza@ucsc.cl

Resumen:

Los objetivos del presente estudio son determinar la diversidad y la densidad léxicas en ensayos argumentativos producidos por estudiantes de español como lengua extranjera (ELE), analizar la influencia de la lengua materna y el nivel de español en dichas medidas y determinar el léxico nuevo usado por los participantes en sus textos según las nociones específicas del *Plan Curricular del Instituto Cervantes*. El estudio es de tipo descriptivo cuantitativo con un diseño transeccional. La muestra está compuesta por 46 textos escritos por estudiantes de ELE en dos universidades tradicionales chilenas, con niveles de proficiencia A2, B1 y B2. Los resultados indican que el nivel de español incide en la diversidad y en la densidad léxicas de los textos escritos, no así la lengua materna. Además, se observa la presencia de una serie de ítems léxicos usados por los estudiantes que no se encuentran inventariados en las nociones específicas del Plan Curricular del Instituto Cervantes, por lo que se sugiere su incorporación en dicho plan.

Palabras clave: densidad léxica, diversidad léxica, escritura, enseñanza, español como lengua extranjera.

Abstract:

The aims of this study are to determine lexical diversity and density in argumentative essays produced by students of Spanish as a foreign language (SFL), to analyze the influence of first language and Spanish level on such measures, and to determine the new lexicon used by the participants in their texts according to the specific notions of the Instituto Cervantes Curricular Plan. The study is a quantitative descriptive study with a

cross-sectional design. The sample is composed of 46 texts written by students SFL in two traditional Chilean universities, with proficiency levels A2, B1 and B2. The results indicate that the level of Spanish has an impact on the lexical diversity and density; the first language does not influence the measures. The presence of a series of lexical items used by students that are not found in the specific notions of the Cervantes Institute's Curricular Plan is observed, so it is suggested that they be incorporated into this plan.

Key words: lexical density, lexical diversity, written, teaching, spanish as a foreign language

Recibido: 1 de octubre de 2021

Aceptado: 22 de noviembre de 2021

Introducción

En la actualidad, existe un creciente interés por aprender español como un segundo idioma, pues su conocimiento permite un acercamiento hacia los ámbitos económico, académico, cultural y laboral en países de habla hispana. Según el Anuario del Instituto Cervantes (Instituto Cervantes *El español 5*), cerca de 22 millones de personas estudian el español como lengua extranjera (ELE) o segunda lengua (SL¹), situándolo como la segunda lengua de comunicación internacional.

Si bien, el léxico ha ocupado, tradicionalmente, un lugar secundario en relación con la atención que se le ha prestado a la gramática y a las nociones y funciones en la enseñanza de una lengua extranjera, se ha ido revalorizando la importancia del aprendizaje del léxico por el rol que cumple en el desarrollo de la interlengua (Albrechtsen et al. 67).

El Marco Común Europeo de Referencia para las Lenguas, MCER, (Consejo de Europa 106) incorpora en las competencias comunicativas de la lengua a la competencia léxica, definida como el conocimiento del vocabulario de una lengua por parte de un aprendiente y su capacidad para utilizarlo. Así, la competencia léxica se compone de: a) Expresiones hechas, compuestas por más de una palabra y que se aprenden como un todo,

¹ Se denomina segunda lengua a aquella que se aprende después de la lengua materna. Si bien existen diferencias entre los conceptos de lengua extranjera y segunda lengua en relación con el contexto de aprendizaje, en este estudio se utilizarán indistintamente ambos términos (Pastor 66).

entre ellas, fórmulas fijas, modismos, frases hechas, etc.; y de b) Polisemia, aludiendo a palabras que pueden tener más de un significado.

Como se puede apreciar, la referencia del MCER (Consejo de Europa 108) sobre la competencia léxica es incompleta, pues solo incluye las expresiones hechas en sus distintas modalidades y las palabras polisémicas, pero no incluye las palabras con significado referencial del tipo 1 forma= 1 ítem léxico, su estructura y su significado (Baralo 29). La palabra se define como “una unidad lingüística, reconocida por los hablantes, que consta de una forma fonológica (el significante, imagen visual o imagen auditiva) asociada a una representación mental, a un concepto (significado)” (Baralo 29).

En este contexto, los objetivos del presente estudio son: 1) Determinar la diversidad y la densidad léxicas en ensayos argumentativos producidos por estudiantes universitarios de niveles A2, B1 y B2 en español como lengua extranjera, 2) Analizar la influencia de la lengua materna y el nivel de español en la diversidad y en la densidad léxicas en ensayos argumentativos escritos por aprendientes universitarios de niveles A2, B1 y B2 de español como lengua extranjera y 3) Determinar el léxico nuevo usado por los participantes en sus textos según las nociones específicas del *Plan Curricular del Instituto Cervantes* para los niveles A2, B1 y B2 de español como lengua extranjera.

Importancia del léxico en la adquisición y en el aprendizaje de una Segunda Lengua

El aprendizaje del léxico es fundamental en la adquisición de una segunda lengua, puesto que las palabras son los primeros elementos lingüísticos que se adquieren y dan contenido para el desarrollo sintáctico, fonológico y morfológico del estudiante (Agustín 17). En el conocimiento del vocabulario, pueden reconocerse tres dimensiones: el tamaño del léxico (*lexical size*), esto es, la cantidad de palabras que el estudiante conoce; la profundidad léxica (*lexical depth*), es decir, cuán bien el alumno sabe las palabras; y la fluidez, referida a la capacidad del aprendiente de recuperar la forma y significado de la palabra desde su memoria y usarla en una situación determinada (Ha 6).

El aprendizaje del vocabulario es un proceso complejo que requiere la adquisición tanto de la forma de las palabras como de sus significados (Ha 6). En el desarrollo de la competencia léxica en una lengua extranjera, se produce una influencia interlingüística, en

especial, con la lengua materna, “que permite facilitar el aprendizaje y proporciona un andamiaje temporal para aprender y estructurar el nuevo conocimiento de la LE” (Agustín 21). De esta forma, las similitudes entre la lengua materna del estudiante y la lengua extranjera proveerán fuentes de andamiaje.

Con respecto a las actividades que propician un mayor aprendizaje del léxico, San Mateo y Andión (224) afirman que tanto las tareas de escritura como las actividades que precisan mayor nivel de participación o esfuerzo cognitivo por parte del aprendiz son las más eficaces para incorporar nuevas palabras al lexicón mental del estudiante de L2.

Enseñanza y evaluación del español como lengua extranjera

El MCER organiza el aprendizaje y la evaluación de una segunda lengua en seis niveles: Básico (A1, A2); Independiente (B1, B2); Competente (C1 y C2). Esta división de niveles proporciona información relevante respecto de la actuación que se espera de un aprendiente de una segunda lengua según su nivel en las distintas destrezas y componentes lingüísticos. En cuanto a la competencia léxica, el MCER provee dos escalas ilustrativas que permiten visualizar la gradación en el conocimiento del vocabulario y en la capacidad para dominar ese conocimiento (ver figura 1).

Nivel	Riqueza de vocabulario	Dominio del vocabulario
C2	Tiene un buen dominio de un repertorio léxico muy amplio, que incluye expresiones idiomáticas y coloquiales; muestra que es capaz de apreciar los niveles connotativos del significado.	Utiliza con consistencia un vocabulario correcto y apropiado.
C1	Tiene un buen dominio de un amplio repertorio léxico que le permite superar con soltura sus deficiencias mediante circunloquios; apenas se le nota que busca expresiones o que utiliza estrategias de evitación. Buen dominio de expresiones idiomáticas y coloquiales.	Pequeños y esporádicos deslices, pero sin errores importantes de vocabulario.
B2	Dispone de un amplio vocabulario sobre asuntos relativos a su especialidad y sobre	Su precisión léxica es generalmente alta, aunque tenga alguna confusión o cometa

	temas más generales. Varía la formulación para evitar la frecuente repetición, pero las deficiencias léxicas todavía pueden provocar vacilación y circunloquios.	alguna incorrección al seleccionar las palabras, sin que ello obstaculice la comunicación.
B1	Tiene suficiente vocabulario para expresarse con algún circunloquio sobre la mayoría de los temas pertinentes para su vida diaria, como, por ejemplo, familia, aficiones e intereses, trabajo, viajes y hechos de actualidad.	Manifiesta un buen dominio del vocabulario elemental, pero todavía comete errores importantes cuando expresa pensamientos más complejos, o cuando aborda temas y situaciones poco frecuentes.
A2	Tiene suficiente vocabulario para desenvolverse en actividades habituales y en transacciones cotidianas que comprenden situaciones y temas conocidos. Tiene suficiente vocabulario para expresar necesidades comunicativas básicas. Tiene suficiente vocabulario para satisfacer necesidades sencillas de supervivencia.	Domina un limitado repertorio relativo a necesidades concretas y cotidianas.
A1	Tiene un repertorio básico de palabras y frases aisladas relativas a situaciones concretas.	No hay descriptor disponible.

Figura 1. Riqueza y dominio de vocabulario según niveles de proficiencia del MCER (Consejo de Europa 109).

Por su parte, el *Plan Curricular del Instituto Cervantes. Niveles de referencia para el español*, conocido también por la sigla PCIC (Instituto Cervantes 2006), es un texto de referencia obligada para los profesores, especialistas e investigadores del ámbito del ELE o SL, porque proporciona un inventario amplio y detallado de diversas especificaciones del español agrupadas en distintos componentes (gramática, pronunciación, ortografía, funciones, tácticas y estrategias pragmáticas, géneros discursivos y textuales, nociones generales, nociones específicas, referentes culturales, saberes y comportamientos socioculturales, habilidades y actitudes interculturales, procedimientos de aprendizaje). El tratamiento del léxico en el PCIC se aborda específicamente en los componentes de nociones generales y de nociones específicas. Dichas nociones se sustentan en el enfoque nociofuncional y en la competencia comunicativa y se presentan agrupadas en series según criterios semánticos.

En el PCIC, las nociones generales se organizan a través de categorías que aumentan su grado de complejidad a medida que el aprendiente avanza en su nivel de

dominio lingüístico del español, por ejemplo, nociones existenciales (no haber ~ nadie/nada, Nivel A2; perder, encontrar, Nivel B1) , cuantitativas (números cardinales, Nivel A1; número cardinal + de + SN, Nivel B2), espaciales (en/por ~ ninguna parte, Nivel B1; estar/hallarse/encontrarse ~ en, Nivel C1), etc. cada una con sus correspondientes subcategorías. Estas nociones generales no se circunscriben a un ámbito específico, sino que corresponden a palabras o estructuras de uso común. Desde la perspectiva de los estudios de frecuencia léxica, a este tipo de nociones se les conoce como *palabras atemáticas*, porque aparecen frecuentemente, independiente del tema del texto. Por el contrario, las *palabras temáticas* son aquellas que aparecen en determinadas áreas según el contenido del texto y en el Plan Curricular se las denomina nociones específicas. Concretamente, las nociones específicas propuestas por el Plan Curricular se clasifican en veinte temáticas: 1. Individuo: dimensión física, 2. Individuo: dimensión perceptiva y anímica, 3. Identidad personal, 4. Relaciones personales, 5. Alimentación, 6. Educación, 7. Trabajo, 8. Ocio, 9. Información y medios de comunicación, 10. Vivienda, 11. Servicios, 12. Compras, tiendas y establecimientos, 13. Salud e higiene, 14. Viajes, alojamiento y transporte, 15. Economía e industria, 16. Ciencia y tecnología, 17. Gobierno, política y sociedad, 18. Actividades artísticas, 19. Religión y filosofía, 20. Geografía y naturaleza. Se debe aclarar que el listado de temas sugerido por el Plan Curricular tiene un inventario abierto, esto significa que se pueden ampliar y agregar nuevos temas según los fines de uso, lo cual permitirá actualizar este inventario a través del tiempo. Por lo tanto, los distintos componentes y nociones del Plan representan una propuesta de orientación y en ningún caso constituyen un inventario cerrado.

Diversidad y densidad léxicas

Una línea tradicional en los estudios del léxico ha sido la distinción entre el tamaño o cantidad de palabras conocidas y la profundidad de conocimiento sobre esas palabras. Otra línea en la que se han desarrollado numerosos estudios ha sido la disponibilidad léxica. Según López (246), el interés en la disponibilidad léxica surge a partir de la crítica que generaba el uso de corpus de frecuencia en los análisis lingüísticos, debido a que estos

no incorporan palabras de uso común. Para ello, se propusieron estímulos denominados “centros de interés” en relación con los cuales los participantes deberían producir listas de vocabulario disponibles (López 246). Sin embargo, esta recopilación de información léxica no se condice con los enfoques comunicativos en la enseñanza de lenguas, pues la producción lingüística no se sustenta ni se potencia mediante listas de palabras. Por lo tanto, una propuesta más auténtica puede consistir en la redacción de un texto sobre un determinado centro de interés, donde no solo se analiza el uso del léxico, sino también otros aspectos como la coherencia, la cohesión, el propósito comunicativo, etc. Si bien, en el ámbito del español predominan los estudios sobre disponibilidad léxica, a nivel internacional uno de los constructos que más se ha revisado es el de riqueza léxica (Bertón 26).

Según Read (55), la riqueza léxica es un constructo amplio que incluye los aspectos de diversidad léxica, sofisticación léxica y densidad léxica. A los anteriores, Ha (8) agrega la fluidez léxica. Para efectos del presente estudio, solo definiremos los conceptos de diversidad y densidad.

La diversidad léxica se concibe como la capacidad de un hablante para utilizar palabras distintas en una muestra de lengua (sea L1, lengua materna, o L2, segunda lengua o lengua extranjera). El índice de diversidad léxica se calcula considerando las palabras de contenido semántico o palabras temáticas y no considera las palabras funcionales o atemáticas (preposiciones, conjunciones, artículos, etc.). A las palabras distintas que presenta un texto se les llama *types* y al total de palabras (incluidas palabras de contenido semántico y funcionales) que presenta el escrito se le denomina *tokens*. Por lo tanto, el índice de diversidad léxica consiste en la división del número de *types* por el número de *tokens* (total de palabras de un texto), lo que se conoce como la fórmula TTR (Type/Token Ratio). Sin embargo, la mayor dificultad que conlleva la TTR se refiere a la extensión del texto, pues mientras más extenso sea el texto la probabilidad de que aparezcan palabras nuevas es más escasa (Durán et al. 224). Esto se debe, por un lado, al uso de palabras gramaticales que pertenecen a clases cerradas (pronombres, preposiciones, artículos, etc.), por lo que se repiten a lo largo del texto y, por otro lado, a la temática desarrollada en el escrito, pues al tratarse de un tópico específico, a lo largo del texto se continúan utilizando los mismos vocablos. Frente a esta dificultad, se han propuesto distintas fórmulas que

buscan expresar de manera confiable la diversidad léxica de un texto, independiente de su extensión. En este estudio, nos referiremos a uno de estos programas que analiza la diversidad léxica basado en el ajuste de curvas (Durán et al. 224, McCarthy y Jarvis 473), específicamente, el *VOCD-D*. El *VOCD* se basa en distintas series de muestreos aleatorios del texto, por lo que comienza tomando de un texto de 100 palabras, muestras aleatorias de 35 tokens y calcula la TTR para cada una de estas muestras y ese promedio se va guardando y se continúa repitiendo el mismo procedimiento hasta completar el texto. Este cálculo produce una curva que describe “de la mejor manera posible los cambios de la TTR” (Bertón 31). Debido a la complejidad para calcular *VOCD*, existen programas computacionales como *Text Inspector* que lo calculan automáticamente. Los valores que proporciona *VOCD* oscilan entre 10 y 100, no obstante, si el valor es más alto que 100 esto indica una mayor diversidad léxica.

Para comprender la interpretación de los valores de diversidad léxica proporcionados por *VOCD*, es necesario explicar brevemente la escala de rangos de Durán et al. (238) en la figura 2.

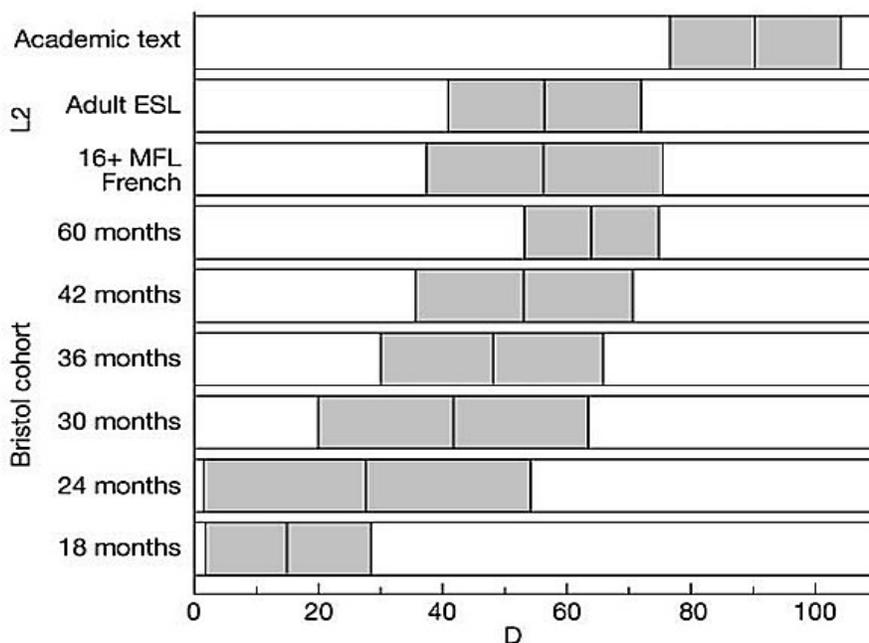


Figura 2. Promedios y subrangos de D para varias cohortes. (Durán et al. 238)

La figura 2 muestra sub-rangos de diversidad léxica que presentan diferentes grupos según edades y tipo de lengua (materna y segunda lengua). El primer grupo de la escala se sustenta en los resultados de los estudios del corpus de Bristol (Wells 281), que muestra el desarrollo de la diversidad léxica de niños que aprenden su lengua materna entre los 18 y los 60 meses de edad (5 años) con un valor mínimo de VOCD de 1,48 (18 meses) y un máximo de 64,02 (60 meses); el segundo grupo (16+MFL French) lo componen 27 británicos aprendientes de francés como segunda lengua, de 16 años de edad, con valores VOCD de 29,64 como mínimo y 87,35 como máximo; el tercer grupo (Adult ESL) corresponde a 32 adultos aprendientes de inglés como segundo idioma con L1 variadas, con un nivel pre-intermedio y con edades entre los 18 y 30 años con valores mínimos y máximos de VOCD de 35,78 y 91,99 respectivamente; y el cuarto grupo (Academic text) se compone de 23 muestras de textos académicos producidos por adultos hablantes nativos, que evidencian un rango de diversidad léxica (VOCD) entre 69,74 y 119,20 con un promedio de 90,59.

La densidad léxica, por su parte, corresponde a la relación entre el total de las palabras con contenido semántico comparado con las palabras funcionales o gramaticales (Fuentes et al. 505, Read 55). Según Fuentes et al. (505), la densidad léxica es un “indicador de la cualidad informativa del texto: un alto índice de densidad léxica significa que tiene a su haber más palabras de contenido, por lo tanto, entrega mayor información”. Se considera, entonces, que un texto es denso cuando incluye un alto número de ítems léxicos en relación con el número total de palabras (Ha 8).

Metodología

El estudio que se presenta es de tipo descriptivo cuantitativo con un diseño transeccional (Hernández et al. 92), puesto que los datos se recopilan en un solo momento con la finalidad de describir la diversidad y densidad léxicas en una muestra de textos producidos por aprendientes de español como SL o LE.

Muestra

La muestra está compuesta por 46 textos escritos por estudiantes de español como lengua extranjera en dos universidades tradicionales chilenas. Los niveles de proficiencia o

dominio de los aprendientes corresponden al A2 (4), B1 (30) y B2 (12) según los resultados obtenidos en la prueba de diagnóstico de español para fines académicos (Vine 172, Vine y Sáez 120). La extensión de los textos fluctúa entre 97 y 554 palabras, con un promedio de 275. En cuanto a la lengua materna, el 32,6% (15) de los estudiantes habla francés; el 23,9% (11) chino; el 19,6% (9) alemán; el 10,9% (5) inglés; el 4,3% (2) portugués; 2,2% (1) italiano; 2,2% (1) ruso; 2,2% sueco (1) y 2,2% checo (1).

Descripción de las tareas de escritura

El diseño de las tareas de escritura se sustenta en los principios metodológicos del enfoque basado en tareas, esto es, autenticidad de la tarea (relación de la tarea con una actividad del mundo real), propósito comunicativo y relevancia del proceso de desarrollo que conduce hacia la tarea final (Estaire 11). Los textos fueron escritos en el contexto de un curso de español para extranjeros en dos universidades tradicionales de Chile. La tarea consistió en la producción de un ensayo argumentativo breve, cuya extensión fuera alrededor de 250 palabras, con una estructura textual que incluyera título, introducción, establecimiento de una postura frente al tema, planteamiento de ventajas y desventajas del tema tratado y conclusión. Se propusieron tres temas, entre los cuales los alumnos debían escoger uno: educación a distancia, crisis del agua y uso del teléfono celular.

Procedimiento

Antes de procesar los textos, se realizó una revisión exhaustiva del corpus que consistió en la corrección de los errores ortográficos y de tipeo, de este modo se evitaron errores de reconocimiento de palabras en el análisis de los textos. Luego, se identificaron fragmentos copiados y pegados desde páginas de Internet a través de un detector de plagio, los que fueron eliminados del corpus.

Para responder a los objetivos 1 y 2 del estudio, determinar la diversidad y densidad léxicas, se elaboró un listado de *stopwords* con el fin de filtrar las palabras funcionales o gramaticales del corpus (preposiciones, conjunciones, pronombres, etc.). Con este paso, se

asegura el cálculo de la densidad léxica, al considerar solo *types* de contenido semántico o palabras temáticas.

Una vez depurado el corpus se procedió a calcular la densidad y diversidad léxicas con el software en línea *Text Inspector* (www.textinspector.com) que entrega diversos índices de análisis lingüísticos. En particular, para establecer la diversidad léxica del corpus, se utilizaron las medidas TTR (type/token ratio) y *VOCD*. Para asegurar el cálculo de la densidad léxica se aplicó una fórmula de TTR ajustado, la que solo considera *types* temáticas y su relación con la extensión de cada texto (TTR ajustado). Además, para efectos de visualizar la relación existente entre las palabras de contenido semántico y el total de palabras con contenido semántico en cada texto, se estableció la medida TTRDens, que realiza el cálculo una vez eliminadas todas las palabras gramaticales o atemáticas incluidas en el listado de *stopwords*.

Medida	Fórmula	Objetivo
TTR	$ttr = \frac{\text{type}}{\text{token}}$	Calcula la relación entre las palabras únicas y la extensión total del texto.
TTR ajustado	$ttrajustado = \frac{\text{type semántico}}{\text{token}}$	Calcula la relación entre las palabras únicas de contenido semántico y la extensión total del texto. Ofrece una mirada que relaciona la densidad léxica con la diversidad de cada texto.
TTRDens	$ttrdens = \frac{\text{type semántico}}{\text{token semántico}}$	Calcula la relación entre las palabras únicas de contenido semántico y el total de palabras de contenido semántico de cada texto. Ofrece una mirada de la densidad léxica del texto al relacionar solo el contenido semántico.

Figura 3. Fórmulas adaptadas para medir diversidad y densidad léxica. Elaboración propia.

Para llevar a cabo el objetivo 3 del estudio se seleccionaron las nociones o palabras específicas del PCIC (Instituto Cervantes 1003) según los tópicos de las tareas de escritura (educación a distancia, crisis del agua o uso del teléfono celular), esto es, (6). Educación, (9). Información y medios de comunicación y (20). Geografía y naturaleza, con sus correspondientes subtemas con la finalidad de comparar aquellas palabras del listado de

frecuencias que se encontraban en el Plan Curricular y aquellas que no figuraban en dicho documento.

Descripción del corpus

El corpus está conformado por 46 textos, que en su conjunto reúnen un total de 12784 palabras o *tokens* y 6292 palabras únicas o *types*. Por otra parte, al filtrar el corpus de las *stopwords*, se obtiene el total de palabras con contenido semántico, que alcanzan a 5892 *tokens* y 4161 *types* o palabras únicas de contenido semántico.

Los aprendientes de español con nivel A2 escribieron textos con una extensión promedio (*tokens*) de 243 palabras, el grupo de aprendientes B1 promedió 280.5 palabras y el grupo con proficiencia B2 escribió en promedio 263 palabras totales. En cuanto a palabras únicas con contenido semántico, el grupo A2 reportó en promedio 75.25 *types* semánticas, el grupo B1 90.6 y el grupo B2 95.2 *types* semánticas. Como se observa en la tabla 1, el grupo A2 produjo 4 textos; el B1, 30 textos; y el grupo B2, elaboró un total de 12 textos. Cabe indicar que cada sujeto elaboró un texto, es decir, el total de textos corresponde al *n* de participantes en el estudio. En la tabla 1 se observa el desglose del reporte léxico de acuerdo con el nivel de español de los sujetos de la muestra.

Tabla 1. Desglose del reporte léxico de acuerdo con el nivel de español

	Token			Type			Token semántico			Type semántico		
	A2	B1	B2	A2	B1	B2	A2	B1	B2	A2	B1	B2
Texto	4	30	12	4	30	12	4	30	12	4	30	12
Mean	243	280.5	263.1	118	136.1	144.7	110.75	131.23	126	75.25	90.6	95.2
Total	972	8415	3157	472	4084	1736	443	3937	1512	301	2718	1142

Se desarrollaron tres temas en los textos analizados: educación a distancia, crisis del agua y el teléfono celular en la vida actual. El primero fue abordado por 18 estudiantes, 15 con nivel de español B1 y 3 con nivel A2. El segundo fue expuesto por 11 alumnos, 10 con nivel de español B1 y 1 con nivel A2. El tercero fue planteado por 17 estudiantes, 12 de ellos con nivel de español B2 y 5 con nivel B1.

Resultados y discusión

Los resultados obtenidos con *Text Inspector* fueron analizados con el software de análisis estadístico JASP 0.14. La tabla 2 presenta los estadísticos descriptivos organizados según el nivel de proficiencia de los tres grupos (A2, B1 y B2).

En términos generales, se observan valores promedio similares en las medidas correspondientes de TTR, TTR dens y TTR ajustado por cada nivel de español. Por ejemplo, en la medida TTR los tres niveles presentan valores iguales o mayores a 0,5. No ocurre lo mismo con los valores de VOCD que varían por cada nivel de manera gradual.

El grupo de nivel A2 obtuvo un promedio de 0.5 en TTR; 0.698 a nivel de TTRDens; 0.32 en TTR ajustado y 67.118 en VOCD. Es importante recordar que los datos de este grupo pueden verse sobrestimados al realizar el análisis de sus promedios debido a que solo cuenta con cuatro textos asociados. Al revisar sus valores mínimos y máximos, es interesante observar que, si bien en el valor de TTR no puntuó con el puntaje mínimo menor entre los tres grupos, sí fue el valor máximo más bajo, situación que se replica en la TTR ajustado (0.38). Se aprecia algo distinto con el valor promedio VOCD (67.118) que sitúa el desempeño en diversidad léxica de este grupo por debajo de los niveles B1 y B2. Cabe destacar que para estos valores se debe ponderar el peso de la muestra con relación al n de los otros dos grupos.

El grupo B1 obtuvo en promedio un TTR de 0.50, un TTRDens de 0.70, un TTR ajustado de 0.33 y un VOCD de 68.889. Llama la atención que, pese a no ser el nivel más bajo de proficiencia, generó el texto con el TTR de menor puntuación entre los tres grupos, con un mínimo de 0.37. Así también en este nivel de proficiencia se generó el texto con el valor TTR más alto (0.64). Esta situación se replica al analizar los máximos de TTRDens y el texto con el valor mínimo de TTR ajustado (0.25). Sin embargo, estos valores extremos se ven sublimados al realizar el análisis de los promedios por grupo, quedando B1 ubicado entre los valores promedio de A2 y B2.

Por su parte, el grupo B2 presenta en promedio un TTR de 0.55, un TTRDens de 0.76, un TTR ajustado de 0.36; así como un VOCD de 93,85. En cuanto a sus valores

mínimos y máximos, todos puntúan sobre los otros dos grupos, quedando en evidencia que la proficiencia a nivel léxico de B2 es superior dentro de los tres grupos.

Tabla 2. Estadísticos descriptivos según nivel de español

	Type/token ratio			TTRDens			TTR ajustado			(VOCD)		
	A2	B1	B2	A2	B1	B2	A2	B1	B2	A2	B1	B2
Valid	4	30	12	4	30	12	4	30	12	4	30	12
Mean	0.500	0.508	0.552	0.698	0.706	0.756	0.320	0.333	0.364	67.118	68.889	93.850
Std. Deviation	0.080	0.068	0.037	0.104	0.075	0.044	0.054	0.043	0.041	6.440	12.032	15.213
Minimum	0.440	0.370	0.490	0.490	0.560	0.690	0.270	0.250	0.310	58.680	38.950	70.210
Maximum	0.610	0.640	0.630	0.840	0.850	0.830	0.380	0.410	0.450	74.260	95.870	119.970

Los valores mínimos y máximos de diversidad léxica obtenidos en la muestra mediante VOCD se sitúan entre 58,6 y 74,2 respectivamente, para el nivel A2; entre 38,9 y 95,8 para B1 y entre 70,2 y 119,9 para B2. De acuerdo con la escala de sub-rangos que proponen Durán et al. (238), el desempeño del nivel B1 sería comparable con el que presenta el grupo *Adult ESL* con un nivel pre-intermedio de inglés, mientras el grupo B2 se asemejaría más al comportamiento del grupo *Academic text*, es decir, el dominio léxico que presentan los hablantes nativos en un contexto académico. En este sentido, coincidimos con McCarthy y Jarvis (473) y Durán et al. (238) en cuanto a la confiabilidad en el análisis de curvas (VOCD) por sobre la TTR. Como se explicó en el apartado teórico, una de las limitaciones de la TTR corresponde a la extensión del texto, si el texto es más largo disminuye la diversidad léxica debido a la repetición de palabras según el tópico de la tarea; por lo tanto, podemos señalar que el índice que calcula VOCD, a diferencia de la TTR, caracteriza de mejor forma a la muestra. Se observa en la tabla 2 que el único índice que arroja promedios diferentes para los distintos niveles de español es el VOCD, mientras las demás fórmulas pueden llevar erróneamente a interpretar que los tres niveles de la muestra tienen una diversidad léxica similar y, por lo tanto, no habría una distinción gradual acorde con el dominio del español (A2, B1 y B2), tal como lo confirman los valores promedio de VOCD que muestran una curva ascendente en relación a lo que se espera para los diferentes niveles de competencia lingüística según los descriptores del MCER (26).

Las medidas TTRDens y TTR ajustado aportan datos sobre la densidad léxica que los sujetos han presentado en sus textos. El cálculo de TTRDens relaciona únicamente palabras con contenido semántico, de ahí que la relación parezca tan alta. En el caso de A2, el valor TTRDens de 0.68 implica que por cada 10 palabras de contenido semántico casi 7 son distintas. Esta situación ocurre también con B1 y B2 cuyos valores de TTRDens son mayores a medida que aumenta la proficiencia, aunque no varían de manera estadísticamente significativa. Sin embargo, estos datos optimistas no deben ser generalizados hacia textos con extensiones mayores, los datos obtenidos en este estudio fueron realizados sobre textos cuyas extensiones promedios fluctuaban entre los 240 y los 280 *tokens*. En textos más extensos la aparición de palabras nuevas -o diferentes- baja a medida que aumentan los *tokens*, esto significa que la densidad disminuye.

Para el caso de TTR ajustado, el objetivo fue identificar el valor informativo del texto. Al relacionar las palabras únicas de contenido semántico con la extensión total del texto se puede obtener la carga informativa que cada sujeto aportó en su tarea. La premisa es que una relación más cercana a 1 entre las palabras distintas con contenido semántico y la extensión del texto, resultaría un texto con más información, debido a que el texto tiene más palabras de contenido semántico. Los resultados, indicaron que el promedio de TTR ajustado en A2 fue de 0.32, en B1 de 0.33 y en B2 corresponde a 0.36, todas relaciones muy por debajo de lo observado en la diversidad léxica (TTR) y en la relación entre las palabras de contenido semántico (TTR Dens), estos valores se pueden explicar a partir de la frecuente repetición de las mismas palabras con contenido semántico a lo largo del texto.

La tabla 3 expone los estadísticos descriptivos organizados según la lengua materna de los sujetos. Tal como se indica en el apartado *muestra*, las lenguas maternas de los sujetos corresponden a francés (15), chino (11), alemán (9); inglés (5); portugués (2), italiano (1), ruso (1), sueco (1) y checo (1). Se dejaron fuera del análisis los textos cuyas lenguas maternas tuvieran un *n* igual o menor a 2, y así evitar posibles distorsiones o falsas interpretaciones.

En términos de promedio de TTR, la lengua materna con mejor puntaje fue el chino (0.54), le sigue en promedio el francés (0.51), luego el alemán (0.5) y finalmente el inglés (0.47). En la medida de densidad TTRDens el promedio más alto también correspondió al chino (0.73), seguido del francés (0.73), en tercer lugar el alemán (0.71) y, por último, el

inglés (0.64). La situación se repite en la medida de TTR ajustado, donde el valor de la lengua china obtuvo el promedio más alto con 0.35, seguido por el francés (0.346), luego el alemán (0.32) y el inglés con un promedio de 0.3. La diversidad léxica bajo la medida VOCD mostró una tendencia similar, aunque el puntaje mayor fue para el francés (82.08), luego el chino (80.61), en tercer lugar el alemán con un puntaje de 65.61 y, finalmente, el inglés con 62.77. Estos resultados evidencian que la lengua materna del aprendiente no incide en la riqueza léxica (diversidad y densidad) de sus escritos, puesto que los estudiantes chinos y franceses obtuvieron mejores puntajes debido a su nivel de español y no, precisamente, por su L1, sobre todo, considerando las distancias lingüísticas entre estas dos lenguas (Escofet et al. 117, Bega 229). Los 15 participantes de lengua materna francesa presentan un nivel intermedio e intermedio avanzado en español, 9 alumnos tienen B1 y 6 B2; sucede lo mismo con los alumnos chinos, 6 con nivel B2 de español y 5 con nivel B1; situación distinta para el inglés y el alemán, donde se sitúan los aprendientes con niveles más bajos de español (alemán, 2 nivel A2 y 7 nivel B1; inglés, 1 nivel A2 y 4 nivel B1).

Tabla 3. Estadísticos descriptivos según lengua materna del aprendiente

	Type/token ratio				TTRDens				TTR ajustado				(VOCD)			
	alemán	chino	francés	inglés	alemán	chino	francés	inglés	alemán	chino	francés	inglés	alemán	chino	francés	inglés
Valid	9	11	15	5	9	11	15	5	9	11	15	5	9	11	15	5
Mean	0.506	0.538	0.514	0.466	0.707	0.727	0.729	0.644	0.323	0.352	0.346	0.302	65.61	80.61	82.08	62.77
Std. Dev.	0.086	0.046	0.066	0.040	0.091	0.057	0.070	0.059	0.047	0.042	0.047	0.040	9.581	18.57	17.9	16.28
Min.	0.410	0.470	0.370	0.430	0.610	0.650	0.580	0.560	0.260	0.310	0.250	0.260	53.60	54.68	57.63	38.95
Max.	0.640	0.630	0.600	0.510	0.850	0.830	0.840	0.720	0.410	0.450	0.430	0.360	79.93	120	108.3	83.38

Si bien el análisis descriptivo indica que los estudiantes de lengua materna china presentan un mejor desempeño en cuanto a la diversidad y a la densidad léxica de sus textos, esta variable no explicaría en sí misma una relación directa con los mejores puntajes de riqueza léxica. De hecho, al aplicar un análisis Anova factorial, los resultados no muestran diferencias significativas que expliquen este fenómeno a nivel de lengua materna. Sin embargo, el factor nivel de español es discriminante en cuanto a la diversidad léxica bajo el índice de VOCD (ver tabla 4). Es decir, el análisis de factores entre las variables lengua materna y nivel de español demuestra que la variable nivel de español resulta estadísticamente significativa ($p < 0.001$), esto explica que los participantes chinos y

franceses presenten una diversidad léxica mayor en sus escritos que el resto de las lenguas del estudio, debido a que en estas lenguas se concentran los estudiantes con mejores niveles de español de la muestra (B1 y B2).

Tabla 4. Anova Nivel de español / Lengua Materna en VOCD

Cases	Sum of Squares	df	Mean Square	F	p
Nivel español	2894.692	1	2894.692	22.481	< .001*
Lmaterna	58.943	1	58.943	0.458	0.506
Nivel*	80.203	1	80.203	0.623	0.439
Lmaterna					
Residuals	2703.951	21	128.760		

Note. Type III Sum of Squares

Los resultados obtenidos en este estudio coinciden con los datos de Bertón (196) en cuanto a la incidencia de la variable nivel de español en la diversidad léxica del alumnado extranjero aprendiente de español, esto significa que los sujetos con mayor proficiencia en español lograron producir textos más ricos en cuanto al uso de distintas palabras. En relación con la influencia de la lengua materna, Galdames et al. (147) concluyen que la L1 del aprendiente de ELE no influye en su disponibilidad léxica.

La información obtenida permite concluir que el ajuste de curvas mediante la fórmula VOCD describió de mejor manera los cambios de la TTR en los textos, independiente de su extensión, por lo tanto, para mayor confiabilidad en la medición de la diversidad léxica, se sugiere utilizar este índice u otros similares disponibles (como *measure of textual lexical diversity*, MTLD) en programas computacionales especializados en estas mediciones (Durán et al. 238).

Léxico no inventariado en tres ámbitos temáticos de las nociones específicas del PCIC

De acuerdo con lo señalado en el apartado de nociones generales y de nociones específicas, los inventarios temáticos del Plan Curricular están abiertos y requieren de actualización acorde con las transformaciones sociales, tecnológicas, culturales, etc. En este

estudio, se compararon las palabras obtenidas por los participantes en la tarea de escritura según su nivel de español con el listado léxico propuesto por el Plan para ese nivel en los tres ámbitos seleccionados: (6) educación, (9) información y medios de comunicación, (20) geografía y naturaleza y sus correspondientes subtemas. Para ello, se utilizó una marcación que permitiera identificar aquellas palabras del listado de frecuencias que se encontraban en el Plan Curricular y aquellas que no figuraban en dicho documento. Las tablas 5, 6 y 7 evidencian este nuevo léxico temático que produjo la muestra en los niveles A2, B1 y B2 y que no está inventariado en el Plan Curricular.

En la tabla 5, se observa que entre las palabras más frecuentes que produjo la muestra del nivel A2 se encuentran: *cursos* (29), educación (16), distancia (14) y agua (10). No obstante, se aprecia un listado léxico más extenso y que es relevante, porque propone nuevas palabras que deberían incorporarse en los ámbitos correspondientes al nivel A2 del Plan Curricular. De manera específica, proponemos incluir en el inventario del PCIC en la noción de educación los términos *cursos, educación, distancia, online, posibilidad, problema, conocimiento, desarrollo, enseñanza, idioma, método, presencial, química, relaciones, responsabilidad, respuesta, situación y social*; en la noción de información y medios de comunicación, los vocablos *gente, problema, trabajo, personas, presencia, conocimiento, filtros, ritmo, situación, social y verdad*; y en el ámbito de geografía y naturaleza, los conceptos *agua, gente, trabajo, regiones, personas, aspecto, beneficios, desarrollo, forma, química, ropa, situación, social, tradicional, vida y tóxicos*.

Tabla 5. Léxico no inventariado producido por estudiantes de nivel A2

Educación, medios de comunicación y geografía					
Nivel A2	Frec.	Nivel A2	Frec.	Nivel A2	Frec.
cursos	29	tóxicos	2	química	2
educación	16	aspecto	2	relaciones	2
distancia	14	beneficios	2	responsabilidad	2
agua	10	conocimiento	2	respuesta	2
online	6	desarrollo	2	ritmo	2
gente	5	enseñanza	2	ropa	2
posibilidad	5	filtros	2	situación	2
problema	5	forma	2	social	2
trabajo	5	idioma	2	tradicional	2
regiones	4	método	2	verdad	2
personas	3	presencial	2	vida	2
presencia	3				

En la tabla 6, se refleja la frecuencia léxica del nivel B1, los términos más recurrentes son: agua (126), personas (55), formación (49), tiempo (37), gente (36), consumo (33), crisis (32), móvil (29), vida (29), problema (27), celular (26), profesor (26), ejemplo (25), países (25), entre otras. En la noción de geografía y naturaleza se pueden considerar los vocablos *agua, personas, consumo, crisis, vida, problema, países, ventajas, mundo, distancia, uso, maneras, ropa, forma, mundial, día, lugar, desventajas, empresas, industria, posibilidad, salud, trabajo, cantidad, casa, posible, productos, diferentes, difícil, sociales, sociedad, etc.* Mientras que en la noción de educación, los términos *personas, formación, tiempo, gente, vida, problema, profesor, ejemplo, importante, académica, nivel, educación, alumnos, trabajo, difícil, escuela, etc.* Por último, en la noción de información y medios de comunicación se debería agregar las palabras *personas, tiempo, consumo, móvil, vida, celular, mundo, distancia, mundial, población, tecnología, futuro, etc.*

Tabla 6. Léxico no inventariado producido por estudiantes de nivel B1

Educación, medios de comunicación y geografía									
Nivel B1	Frec.	Nivel B1	Frec.	Nivel B1	Frec.	Nivel B1	Frec.	Nivel B1	Frec.
agua	126	tradicional	9	grave	4	ambiente	2	mano	2
personas	55	bienes	8	hecho	4	argumentos	2	mantener	2
formación	49	calidad	8	idea	4	ausencia	2	material	2
tiempo	37	aprendizaje	7	largo	4	cabeza	2	mayor	2
gente	36	efecto	7	libre	4	ciencia	2	mercado	2
consumo	33	escasez	7	litros	4	científicos	2	monto	2
crisis	32	parte	7	momento	4	compuestos	2	motivación	2
móvil	29	poder	7	tratamiento	4	consecuencia	2	mujeres	2
vida	29	realidad	7	vez	4	consumidores	2	método	2
problema	27	situación	7	vista	4	conveniencia	2	nacimiento	2
celular	26	oportunidades	7	precios	4	conveniente	2	negativa	2
importante	26	acceso	6	preguntas	4	costo	2	niveles	2
profesor	26	amigos	6	actitudes	3	cuidado	2	notificación	2
ejemplo	25	condiciones	6	acuerdo	3	cultura	2	obligación	2
países	25	contacto	6	adecuada	3	daños	2	ordenador	2
ventajas	25	desarrollo	6	apoyo	3	demanda	2	palabras	2
mundo	24	dinero	6	caballero	3	demasiadas	2	pantalla	2
distancia	22	general	6	cantidades	3	desigualdad	2	papel	2
maneras	22	humano	6	capacidad	3	discusiones	2	pasado	2
lado	21	población	6	caso	3	escala	2	pasión	2
académica	20	suficiente	6	claro	3	esfuerzo	2	peligroso	2
uso	20	tecnología	6	cotidiana	3	estudiante	2	permitido	2
ropa	19	distintos	6	crecimiento	3	evidentes	2	personal	2
forma	18	factores	6	dificultades	3	evolución	2	polución	2
mundial	18	alto	5	directo	3	familias	2	presencia	2
empresas	18	atención	5	económico	3	fiesta	2	productores	2
día	17	causa	5	enseñanza	3	final	2	propuestos	2
lugar	17	futuro	5	espacio	3	gafas	2	pueblo	2
posibilidad	17	fábricas	5	existencia	3	gestión	2	razón	2
desventajas	16	hijos	5	fundamental	3	globalización	2	relacionado	2
nivel	16	institución	5	ganas	3	gratis	2	respuesta	2
educación	15	métodos	5	industrial	3	graves	2	ríos	2
industria	15	nueva	5	jóvenes	3	hombre	2	sala	2
salud	13	opción	5	millones	3	hombro	2	sector	2
alumnos	12	padres	5	moderna	3	hora	2	sentido	2
escuela	12	potable	5	nuevo	3	horario	2	sequías	2
punto	12	productos	5	online	3	idioma	2	señal	2
trabajo	11	tipo	5	pedagogía	3	imprescindible	2	sistema	2

universidad	11	unesco	5	procesos	3	inadecuado	2	sobreconsumo	2
cantidad	10	varios	5	producción	3	indirecto	2	sobreexplotación	2
casa	10	funciones	5	razones	3	influenciada	2	solos	2
posible	10	aspecto	4	real	3	informe	2	suelo	2
productos	10	bueno	4	recurso	3	innecesario	2	tipos	2
conocimientos	10	computador	4	redes	3	inteligente	2	trabajos	2
relaciones	10	conclusión	4	trabajadores	3	intercambio	2	trajes	2
diferentes	9	conflictos	4	prácticas	3	interesantes	2	tóxicos	2
difícil	9	cosa	4	abuso	2	juegos	2	virtual	2
necesario	9	fin	4	actual	2	leyes	2	vital	2
sociales	9	global	4	afectada	2	limpia	2	volumen	2
sociedad	9	grande	4	alimentos	2	malas	2	único	2

Por último, en la tabla 7, se observa la frecuencia de palabras producidas por estudiantes de nivel B2. En este nivel destacan las palabras celular (59), móvil (36), personas (21), mundo (15), vida (14), amigos (13), sociales (13), uso (11), cosas (11), lado (11) y salud (10). En la noción de información y medios de comunicación, se tendría que incorporar al nivel B2 los términos *celular, móvil, personas, amigos, sociales, salud, relaciones, gente, ventajas, desventajas, familiares, tecnología, etc.* En el ámbito de educación, los vocablos *celular, móvil, personas, mundo, amigos, sociales, uso, jóvenes, cerebro, investigaciones, etc.* Para el ámbito de geografía y naturaleza, las palabras *personas, mundo, vida, sociales, uso, relaciones, gente, ventajas, desventajas, lugar, casa, perjudiciales, trabajo, acciones, indispensable, negativos, ondas, precaución, revolución, sociedad, etc.*

Tabla 7. Léxico no inventariado producido por estudiantes de nivel B2

Educación , medios de comunicación y geografía							
Nivel B2	Frec.	Nivel B2	Frec.	Nivel B2	Frec.	Nivel B2	Frec.
celular	59	jóvenes	5	objeto	3	fotografías	2
móvil	36	manera	5	país	3	fotos	2
personas	21	momento	5	punto	3	francia	2
mundo	15	perjudiciales	5	tecnológica	3	fácil	2
vida	14	trabajo	5	teléfonos	3	herramienta	2
amigos	13	acciones	4	utilización	3	idea	2
importante	13	adicción	4	vidas	3	llena	2
sociales	13	indispensable	4	vista	3	llevado	2
uso	13	negativos	4	época	3	música	2
lado	11	ondas	4	acosadas	2	niños	2

cosas	11	precaución	4	caras	2	notificaciones	2
salud	10	revolución	4	causa	2	opinión	2
relaciones	9	situación	4	chile	2	particularidad	2
gente	8	sociedad	4	constante	2	posible	2
ventajas	8	cerebro	3	correos	2	positivos	2
desventajas	7	cuidado	3	demostrado	2	problema	2
familiares	7	diferentes	3	dependencia	2	segundos	2
lugar	7	familia	3	electrónicos	2	simple	2
tecnología	6	investigaciones	3	excepción	2	universidad	2
aspectos	5	juegos	3	excesivo	2	usada	2
casa	5			forma	2	veces	2
ejemplo	5					útil	2

La propuesta de actualización de las nociones específicas del PCIC es una acción necesaria, puesto que el léxico posee un inventario abierto, por lo cual es pertinente que constantemente se actualicen los vocablos según los avances de la tecnología, de la sociedad, el contacto con otras lenguas, la creación de nuevas palabras, etc. Por ejemplo, en las nociones inventariadas del PCIC aparecen términos como *tarjeta telefónica*, *teléfono analógico*, *enviar una postal*, *fax*, etc. En efecto, se requiere una revisión frecuente de estos listados, pues las modificaciones que se incorporen en el documento de referencia del PCIC impactan en la elaboración de materiales para enseñar la lengua, en los syllabus de los cursos y en la evaluación y certificación del idioma; sobre todo, considerando la influencia que ejerce el nivel de español en la riqueza léxica de los textos. Asimismo, sería interesante que los textos que se usen como *input* en las clases o evaluaciones de ELE puedan ser sometidos previamente a mediciones de índices de riqueza léxica, para así utilizar materiales confiables y adecuados a los niveles de competencia de los destinatarios. Esto se ve reafirmado por el estudio de Galdames et al. (148), quienes concluyen que es necesario reforzar el léxico en los niveles intermedios y avanzados para afianzar los términos, tanto los que aparecen en los inventarios del PCIC como aquellos que no han sido inventariados y que surgen a partir de las temáticas objeto de estudio.

Conclusiones

El conocimiento léxico es fundamental en la comunicación e interacción en una segunda lengua y, además, favorece la comprensión lectora, la escritura y la organización discursiva. De acuerdo con Agustín (17), un error gramatical tiene poca relevancia en la comunicación, pese a ello, se puede mantener la interacción; no así cuando se comete un error léxico, porque puede conllevar a una falta de comprensión o entendimiento. Galdames et al. (136) manifiestan que los aprendientes de español presentan dificultades léxicas para expresarse adecuadamente, lo cual puede causarles frustración e incidir negativamente en el aprendizaje del idioma. En particular, el carácter formal de la producción escrita exige que el estudiante realice una elección cuidadosa de los términos que pretende utilizar (Laufer 1).

Los hallazgos de este estudio contribuyen de manera significativa en cuanto a la consideración del léxico desde la perspectiva de la enseñanza y de la investigación en el contexto de la Lingüística Aplicada y su relación con los niveles de dominio lingüístico del aprendiente. Sobre esto, destacamos la importancia que tiene el conocimiento léxico en la escritura y cómo la riqueza léxica de los textos evidencia el nivel de proficiencia del estudiante, por lo tanto, no solo se debe dedicar tiempo y esfuerzo a enseñar y retroalimentar los errores gramaticales, sino que es necesario considerar el fortalecimiento de la riqueza léxica y la corrección de este tipo de errores en todos los niveles, principalmente en los iniciales donde el vocabulario tiende a presentar mayor influencia de la L1 mediante la transferencia lingüística (Quiñones 2).

En la enseñanza y el aprendizaje de una lengua, la argumentación se considera una de las habilidades más complejas, porque debe conjugar el propósito de esta práctica discursiva, que es convencer o persuadir al receptor sobre un tema que genera polémica con el dominio del esquema o estructura del texto que debe producir. A nivel léxico, en la argumentación se emplea un vocabulario propio de los razonamientos, palabras que indican refutación de otras opiniones, verbos que explicitan la acción de argumentar, adjetivación y calificación de los referentes (Marín 183, Álvarez 203). Se trata de un tipo de texto complejo de construir, pues el escritor debe establecer relaciones entre los hechos y su punto de vista, lo que implica una mayor abstracción que se refleja en una selección léxica

que transmite calificación y valoración. En este sentido, consideramos que el tipo de tarea descrita en este estudio (escritura de un ensayo breve) pudo incidir en los resultados observados en los estudiantes con menor competencia en español, sobre todo, teniendo en cuenta que el MCER (Consejo de Europa 26) incorpora la producción de textos argumentativos a partir del nivel B2. Sería interesante que nuevos estudios en este ámbito pudieran explorar la incidencia de la variable tipo y complejidad de la tarea en la diversidad y densidad léxicas de textos escritos.

El tamaño o extensión del vocabulario de los aprendientes de una segunda lengua y la profundidad de este conocimiento son dos aspectos fundamentales al tratar la competencia léxica (López-Mezquita 179). Este trabajo aporta metodológicamente a comprender y evaluar ambos aspectos (amplitud y profundidad) desde una perspectiva cuantitativa, y, especialmente, desde la lengua en uso. Se contribuye a la medición de la riqueza léxica al proporcionar una aproximación para el análisis de la competencia léxica de los aprendientes de español. Incluye, por una parte, aspectos relacionados con el acceso a la amplitud del vocabulario, al abordar la diversidad léxica (TTR) con fórmulas ajustadas que consideran la variable de contenido semántico y, por otra parte, analiza la densidad léxica de los textos, como un intento de acceder al grado de profundidad que el aprendiente tiene sobre un léxico temático. El objeto de dicho diseño, es lograr un indicador de calidad de escritura y dominio de la lengua del aprendiente. Además, el filtrado de las palabras sin contenido semántico en el corpus evidencia el léxico que utilizan los estudiantes en un contexto comunicativo determinado. Como resultado de esta metodología, no solo se obtienen los indicadores de competencia léxica en torno a la diversidad y a la densidad, sino que además permite realizar una propuesta para incorporar el caudal no inventariado que aportan los aprendientes de español para enriquecer y actualizar el conocimiento y dominio léxico en contextos generales y académicos en español como L2.

REFERENCIAS

- Agustín, María. “Aprendizaje de vocabulario en la Lengua Extranjera Influencia interlingüística, modo de aprendizaje y otras variables moldeadoras”. *Palabras Vocabulario Léxico. La lexicología aplicada a la didáctica y a la diacronía*, edición de Florencio del Barrio de la Rosa. Editorial Ca Foscari, 2017, pp. 35-48.
- Albrechtsen, Dorte, Kirsten Haastrup y Birgit Henriksen. *Vocabulary and writing in a first and second language: Processes and development*. Editorial Palgrave Macmillan, 2008.
- Álvarez, Gerardo. *Textos y discursos*. Editorial Universidad de Concepción, 2001.
- Baralo, Marta. “La competencia léxica en el Marco común europeo de referencia”. *Revista Carabela*, 58, 2005, pp. 27-48.
- Bega, Michel. “Dificultades metodológicas de los estudiantes sinohablantes de español como lengua extranjera”. *Onomázein*, núm. 32, 2015, pp. 227 – 238.
- Berton, Marco. *Riqueza léxica y expresión escrita en aprendices suecos de ELE: Proficiencia general, competencia léxica pasiva, tipo y complejidad de la tarea*. Stockholm University, 2020, <http://urn.kb.se/resolve?urn=urn:nbn:se:su:diva-180993>. Consultado el 20 de marzo de 2021.
- Consejo de Europa. *Marco Común Europeo de referencia para las lenguas: aprendizaje, enseñanza y evaluación*. Ministerio de Educación, Cultura y Deporte de España, 2002.
- Durán, Pilar, David Malvern, Brian Richards y Ngoni Chipere. (2004). “Developmental Trends in Lexical Diversity”. *Applied Linguistics*, vol. 25, núm. 2, 2004, pp. 220-242.
- Escofet, María Isabel, Mar Colón y Pedro Medina. “Desarrollo de un test de nivel de comprensión pragmática para estudiantes sinohablantes de español como lengua extranjera”. *Onomázein*, núm. 49, 2020, pp.115-139.
- Estaire, Sheila. “La programación de unidades didácticas a través de tareas”. *RedELE: Revista electrónica de didáctica español lengua extranjera*, núm. 1, 2004, pp. 1-20.
- Fuentes, Karina, Sergio Hernández y Pedro Salcedo. “Descripción de la diversidad y densidad léxicas en noticias escritas por estudiantes de periodismo”. *Revista Brasileira de Lingüística Aplicada*, vol. 19, núm. 3, 2019, pp. 499-528.
- Galdames, Angela, Silvana Guerrero y Gloria Toledo. “Disponibilidad léxica de aprendientes de español como segunda lengua en Santiago de Chile: una plataforma para la enseñanza del léxico”. *Logos: Revista de Lingüística, Filosofía y Literatura*, vol. 28, núm. 1, 2018, pp. 135-150.
- Ha, Hye Seung. “Lexical Richness in EFL Undergraduate Students’ Academic Writing”. *English Teaching*, vol. 74, núm. 3, 2019, pp. 3 - 38.
- Hernández, Roberto, Carlos Fernández y Pilar Baptista. *Metodología de la investigación. Sexta edición*. McGraw Hill Education, 2014.
- Instituto Cervantes. *El español en el mundo. Anuario del Instituto Cervantes 2019*. Editorial Instituto Cervantes, 2019. https://cvc.cervantes.es/lengua/anuario/anuario_19/. Consultado el 10 de diciembre de 2020.
- . *Plan curricular del Instituto Cervantes. Niveles de referencia para el español*, 2006, https://cvc.cervantes.es/ensenanza/biblioteca_ele/plan_curricular/default.htm. Consultado el 11 de diciembre de 2020.

- JASP Team. *JASP (Version 0.14.1)*, Computer software, 2020, <https://jasp-stats.org>. Consultado en enero-mayo de 2021.
- Laufer, Batia. “Vocabulary and writing”. *The Encyclopedia of Applied Linguistics*, edición de Carol Chapelle. Wiley-Blackwell, 2013, pp. 1-5.
- López, Humberto. “Los estudios de disponibilidad léxica: pasado y presente”. *Boletín de Filología*, núm. 35, 1993, pp. 245 - 229.
- López-Mezquita, María Teresa. *La evaluación de la competencia léxica: Test de vocabulario. Su viabilidad y validez*. Universidad de Granada, 2007.
- McCarthy, Philip y Scott Jarvis. “Vocd: a theoretical and empirical evaluation”. *Language Testing*, núm. 24, 2007, pp. 459-488.
- Marín, Marta. *Lingüística y enseñanza de la lengua*. Aique, 2008.
- Pastor, Susana. *Aprendizaje de segundas lenguas: Lingüística aplicada a la enseñanza de idiomas*. Editorial Universidad de Alicante, 2004.
- Quiñones, Alba. “El análisis de errores en el campo del Español como Lengua Extranjera: Algunas cuestiones metodológicas”. *Revista Nebrija de Lingüística Aplicada*, vol. 5, núm. 3, 2009, pp. 1-16.
- Read, John. *Assesing Vocabulary*. Cambridge University Press, 2000.
- San-Mateo, Alicia y María Antonieta Andión. “Eficacia de tres actividades de aprendizaje de vocabulario en español LE/L2: la selección de definiciones, la selección de ejemplos y la escritura de oraciones”. *Verba Hispánica*, núm. 27, 2019, pp. 209-229.
- Vine, Ana y Katia Sáez. “Evaluación del Español como Lengua Extranjera con Fines Académicos: Una necesidad actual en el ámbito universitario”. *Onomázein*, núm. 42, 2018, pp. 102-124.
- . “Validez de un prototipo de prueba de Español como Lengua Extranjera con Fines Académicos”. *Revista Nebrija de Lingüística Aplicada*, vol. 12, núm. 24, 2018, pp. 112-153.
- Wells, Gordon. *Language development in the pre-school years*. Cambridge University Press, 1985.