

Vocabulario académico en los textos escolares  
chilenos: análisis de las palabras comunes a las 4  
disciplinas troncales de la escolaridad\* \*\*

Soledad Aravena

<https://orcid.org/0000-0001-8646-9755>

*Pontificia Universidad Católica de Chile*

maravena@uc.cl

Riva Quiroga

<https://orcid.org/0000-0002-1147-4135>

*Pontificia Universidad Católica de Chile*

riva.quiroga@uc.cl

Felipe Villegas

<https://orcid.org/0009-0006-6580-6912>

*Pontificia Universidad Católica de Chile*

fgvillegas@uc.cl

---

\* Estudio realizado en el marco del Proyecto 660299 “Complejidad y progresión en las lecturas escolares: análisis de lecturabilidad en distintos grupos de lectores y tres tipos de formato”, financiado por el Fondo del Libro y la Lectura, Ministerio de las Culturas, las Artes y el Patrimonio de Chile (MinCAP).

\*\* Soledad Aravena: investigadora responsable del estudio y encargada de la introducción, resultados y conclusiones del artículo; Riva Quiroga, coinvestigadora del estudio y encargada de la sección metodológica del artículo; Felipe Villegas, asistente del estudio y colaborador en la redacción final, búsqueda de ejemplos y revisión de bibliografía y referencias.



## RESUMEN

Este estudio se propuso explorar el vocabulario de los textos escolares chilenos de 4 disciplinas y 12 niveles educativos, analizando cómo se distribuyen las palabras por disciplina, nivel escolar, grado de polyvalencia y especialización disciplinar. De un corpus inicial de 2,6 millones de palabras, se extrajeron las palabras comunes a las 4 materias para conformar una lista de vocabulario académico. Los resultados indican que estas palabras aumentan su frecuencia en los niveles superiores, y se distribuyen de acuerdo con el carácter más humanista o más científico de las materias y su naturaleza más centrada en los contenidos o las habilidades. Se concluye la importancia de estudiar este vocabulario para promover su enseñanza a lo largo de toda la trayectoria escolar.

*Palabras clave:* vocabulario académico, textos escolares, disciplinas escolares

## Academic Vocabulary of Chilean Textbooks: Analysis of the Words Common to the 4 Core Disciplines of Schooling

### ABSTRACT

This study aimed to explore the vocabulary of Chilean textbooks from 4 disciplines and 12 educational levels, analyzing how words are distributed by discipline, school level, degree of polyvalence and disciplinary specialization. From an initial corpus of 2.6 million words, common words to the 4 subjects were extracted to form a list of academic vocabulary. The results indicate that these words increase in frequency at higher levels and are distributed according to the more humanistic or more scientific nature of the subjects and their nature more focused on content or skills. The study concludes by highlighting the importance of examining this vocabulary in order to promote its teaching throughout the entire school trajectory.

*Keywords:* academic vocabulary, textbooks, school disciplines

## 1. INTRODUCCIÓN

El vocabulario es uno de los aprendizajes fundamentales de la etapa escolar. Muchas de las palabras que se aprenden en la escuela no están disponibles en la vida cotidiana de los niños y resultan

fundamentales para su desarrollo lingüístico, conceptual y cognitivo. De hecho, muchos estudios han revelado la estrecha relación entre el conocimiento léxico y la competencia en el uso del lenguaje oral y escrito (Meneses, Uccelli, Santelices, Acevedo y Figueroa 2018). En particular, el vocabulario contribuye de forma decisiva a la comprensión lectora, una habilidad fundamental para el aprendizaje (Perfetti y Helder 2022).

Se estima que el vocabulario que enfrentarán los niños durante su escolaridad alcanza aproximadamente las 180 000 palabras (Graves 2016). Es una cantidad enorme, considerando que entran a la escuela con vocabularios de unas 5000 palabras, las cuales pueden ser menos por variables socioeconómicas o de contexto familiar, como muestran estudios previos de la región (Lohndorf, Vermeer, Cárcamo y Mesman 2018; Weisleder y Fernald 2013). Las palabras que aprenderán en la escuela son, por lo tanto, de la mayor relevancia, dado que les permitirán acceder a los textos escritos generales y disciplinares, mediante los cuales se construyen y transmiten los conocimientos.

### **1.1. Vocabulario académico y lenguaje académico**

El vocabulario académico incluye palabras transversales a varios campos de conocimiento y palabras más especializadas de un campo disciplinar específico, pasando por palabras propias de las tareas escolares o los libros de texto (Baumann y Graves 2010, Coxhead 2020). Su aprendizaje facilita el acceso a los saberes disciplinares y favorece el éxito escolar, incluyendo la participación en prácticas letradas (Fitzgerald, Relyea y Elmore 2022). Además, este vocabulario representa entre el 20 % y el 30 % de las palabras de los textos académicos (Chung y Nation 2003), proporción significativa si consideramos que, para comprender un texto, se requiere conocer cerca del 98 % de las palabras que contiene (Nation 2006).

Aunque muchos estudios sobre vocabulario académico provienen de contextos anglófonos, algunos trabajos en español confirman su relevancia en las habilidades de lenguaje académico (Meneses, Hugo, Montenegro, Valenzuela y Ruiz 2018; Romero,

Silva-Maceda y Snow 2021). Estos trabajos coinciden en la necesidad de incluir apoyos para el desarrollo de las habilidades del lenguaje académico en los planes de estudio de lengua española, debido a la evidencia de su impacto en la comprensión lectora. Por otra parte, se han desarrollado con resultados positivos instrumentos y experiencias de enseñanza explícita del lenguaje académico (Meneses, Hugo y otros 2018) que, en el contexto latinoamericano, resulta crucial para la comprensión y el aprendizaje disciplinar. Esta perspectiva, que es también la del presente estudio, dialoga con propuestas de alfabetización disciplinar en un marco más amplio (Shanahan y Shanahan 2008, Uccelli y Meneses 2015), que proponen favorecer el progreso desde habilidades generales a prácticas específicas de cada área, enseñando no solo más palabras a los estudiantes, sino promoviendo su uso activo en géneros y tareas propias de ciencias, historia o matemática.

## 1.2. VOCABULARIO ACADÉMICO GENERAL Y DISCIPLINAR

Se suelen distinguir dos grandes tipos de vocabulario académico, uno de carácter más específico y otro de carácter más general. El primero, también llamado de dominio o contenido específico (Hiebert y Lubliner 2008, Fisher y Frey 2008) es probablemente el tipo más común de vocabulario académico y es central en la construcción de los conocimientos disciplinares. Se presenta con alta frecuencia y poca dispersión en los textos de una disciplina en particular, pero con bajas frecuencias en los textos académicos de otras áreas. La importancia de estas palabras radica en su centralidad para comprender el contenido disciplinar, no solo para el uso receptivo de los términos, sino también para ampliar sus campos conceptuales, comprender las relaciones semánticas posibles entre los conceptos y alcanzar su uso productivo en el lenguaje oral y escrito (Hüttner y Dalton-Puffer 2024).

El segundo grupo incluye el llamado vocabulario académico general (Coxhead 2020) y de las actividades escolares, situado entre el lenguaje de la vida cotidiana y el léxico especializado (Beck,

McKeown, y Kucan 2013). Son palabras de alta frecuencia en los textos escolares y académicos y están relativamente ausentes de otros tipos de textos o situaciones (Coxhead y Nation 2001, Webb y Nation 2017). Por ello, no se aprenden fácilmente de manera incidental y resultan desconocidas para muchos estudiantes, particularmente para los estudiantes no nativos y aquellos de grupos desfavorecidos (Gardner y Davies 2014; Townsend, Filippini, Collins y Biancarosa 2012). Asimismo, son menos notorias que las palabras especializadas, pero se presentan de manera transversal y frecuente en los textos académicos (escolares o universitarios), por lo que no suelen ser objeto de enseñanza explícita por parte de los profesores, al considerarlas estos demasiado comunes y presumiblemente conocidas (Webb y Nation 2017). Esta misma transversalidad determina un tercer rasgo de este vocabulario, su alto grado de polivalencia. Es decir, a diferencia del vocabulario especializado, este vocabulario suele contener palabras polisémicas cuyos significados varían considerablemente de acuerdo con el campo disciplinar. Los rasgos descritos, es decir, su polivalencia, junto al hecho de no ser objeto habitual de aprendizaje incidental, ni de enseñanza explícita en el contexto disciplinar, hacen desafiante y necesario el estudio de este vocabulario.

### **1.3. Relevancia del texto escolar y su vocabulario**

Los textos escolares tienen un importante rol en la construcción de los conocimientos durante la educación formal (Bondi 2016). En Chile, esto se cumple a través de una política pública de elaboración y distribución gratuita de textos a todos los estudiantes de colegios públicos y subvencionados (más del 90 % de la matrícula nacional) para todos los niveles y materias escolares (Olivera 2017, Riveros 2021). Esta iniciativa ha cumplido históricamente una importante función en el sistema educativo al favorecer la apropiación del *currículum* y promover la igualdad de oportunidades en los aprendizajes y el acceso a la lectura (Olivera 2017). Al respecto, el texto escolar constituye un valioso —y cada vez más escaso— recurso de lectura impresa y relativamente extensa, que puede propiciar la

“lectura en profundidad” (Kovač y van der Weel 2020). Este tipo de actividad es cognitivamente exigente porque requiere construir nuevos conocimientos y promueve el desarrollo de formas sofisticadas de pensamiento, de lenguaje y, en particular, de vocabulario (Kovač y Kepic 2022, Støle 2020).

En efecto, una parte muy importante del vocabulario que aprenden los niños y adolescentes en su educación formal proviene de estos textos. Esto es más importante si recordamos que, en lengua materna, la mayor cantidad de palabras se aprende de manera incidental, particularmente a través de la lectura (Webb 2020). Es decir, aunque estemos concentrados en el contenido, aprendemos gradualmente las palabras incluidas en los textos al verlas una y otra vez en contexto. La cantidad de aprendizaje incidental de vocabulario depende de la cantidad de información; más información proporciona una mayor oportunidad de encontrar palabras repetidamente y, por lo tanto, una mayor probabilidad de que se aprendan las palabras (Webb y Chang 2015). Así, el estudio de este vocabulario tiene importantes implicancias desde la perspectiva del aprendizaje, pero también de la instrucción y de la elaboración de material escolar (Fitzgerald, Relyea y Elmore 2022). Conocer el volumen y otras características del vocabulario académico de estos textos informa sobre el léxico deseable de sus potenciales lectores. Por otro lado, también proporciona información importante para que los docentes puedan tomar mayor conciencia de las dificultades que enfrentan los estudiantes de distintas edades y puedan brindarles los apoyos necesarios para promover mejores aprendizajes. Esto, para favorecer tanto la instrucción directa de vocabulario, como su aprendizaje incidental a través de lecturas e interacciones orales en clases.

En este contexto, este estudio se propuso explorar el vocabulario de la escuela mediante la extracción y el análisis de todas las palabras comunes a los textos de las 4 disciplinas troncales de la formación escolar: Lenguaje, Matemáticas, Ciencias Sociales y Ciencias Naturales. Los objetivos específicos son: a) determinar cómo se distribuyen esas palabras en rangos de frecuencia

por disciplina y nivel escolar; b) proponer una clasificación de las palabras en función de su mayor o menor grado de polivalencia y especialización disciplinar; y c) analizar algunos ejemplos de cada grupo en relación con sus sentidos y contextos de uso a través de las disciplinas y los niveles escolares.

## 2. METODOLOGÍA

### 2.1. Compilación del corpus

Para el presente estudio se trabajó con los textos escolares de Enseñanza Básica (EB) y Media (EM) entregados por el Ministerio de Educación de Chile durante el año 2023<sup>1</sup> en los 12 niveles escolares (1° básico a IV medio) y 4 áreas de estudio: Lenguaje, Matemáticas, Ciencias Naturales (CN) y Ciencias Sociales (CS). En los dos últimos años de EM, en todas las disciplinas se utiliza un solo texto. Por esta razón, los resultados de ambos cursos se presentan reunidos como M34 (es decir, hay 12 niveles escolares pero 11 textos). Algo similar ocurre en Ciencias en I° y II° medio, pero en estos textos las unidades correspondientes a cada nivel están bien delimitadas, lo que hace posible reportar los resultados de forma separada. Todo lo anterior resulta en una muestra final de 101 textos escolares, que incluyen el texto principal (“Texto del Estudiante”) y los complementarios, con actividades de práctica y lecturas adicionales. Estos libros complementarios van desde 1° a 4° básico en Lenguaje y de 1° a 6° básico en Matemáticas, lo que explica en parte el mayor volumen del corpus en estos niveles y materias, como se aprecia más adelante (Tablas 1 y 2).

Se trabajó con los libros en formato PDF, que el Ministerio de Educación de Chile pone a disposición de la comunidad escolar a través de sus plataformas digitales. Para procesar los archivos, se utilizó el paquete *pdftools* (Ooms 2024) para el lenguaje de programación R (R Core Team 2024), que permitió extraer el texto de cada

---

<sup>1</sup> Algunos textos de 2023 siguen vigentes en 2025; otros se han modificado, aunque las editoriales mantienen una línea base de contenidos y lecturas.

archivo. Este mismo lenguaje de programación se utilizó para todas las siguientes etapas del procesamiento y análisis de datos. En una primera fase, se eliminaron las páginas correspondientes a la tabla de contenidos, créditos, portada, contraportada y solucionarios y glosarios, cuando había. En una segunda fase, se eliminaron los encabezados y pies de página utilizando expresiones regulares<sup>2</sup>, para conservar solo el texto que corresponde a las unidades de aprendizaje. Para los cálculos de frecuencia se agruparon los documentos de cada disciplina correspondientes a un mismo nivel. Así, el tamaño final del corpus es de unas 2.6 millones de palabras totales (*tokens*) como se muestra en la Tabla 1.

**Tabla 1.** Total de palabras (tokens) por disciplina y nivel

Nivel	Tokens				
	L	CS	CN	M	Total
<b>B1</b>	73778	15642	7781	11827	109.028
<b>B2</b>	68285	25499	11160	28652	133.596
<b>B3</b>	130055	33643	20360	58943	243.001
<b>B4</b>	127197	30078	25848	67486	250.609
<b>B5</b>	86429	51533	26983	111908	276.853
<b>B6</b>	68947	62196	27064	96312	254.519
<b>B7</b>	56616	72972	37466	67793	234.847
<b>B8</b>	68503	72452	38605	65281	244.841
<b>M1</b>	79177	86338	60740	73009	299.264
<b>M2</b>	82747	104725	58778	63249	309.499
<b>M34</b>	106527	84201	64776	70346	325.850
<b>Total</b>	948.261	639.279	379.561	714.806	2.681.907

<sup>2</sup> En el procesamiento computacional del lenguaje, una *expresión regular* es una notación formal que permite especificar reglas para localizar y extraer secuencias de texto que cumplen determinadas condiciones. En lingüística de corpus, las expresiones regulares se utilizan para tareas de preprocesamiento (como eliminar encabezados y pies de página, marcas tipográficas o metadatos), así como para identificar cadenas léxicas o patrones morfológicos y sintácticos recurrentes en grandes volúmenes de textos (Friedl 2006, Lüdeling y Kytö 2009).

## 2.2. Creación de listas de palabras distintas (*types*) y palabras comunes

Posteriormente, se calculó la frecuencia absoluta de cada tipo de palabra o *type* (la primera aparición de una palabra, sin considerar sus repeticiones) en los textos correspondientes a cada nivel. Esto permitió generar, para cada disciplina, 11 listas de frecuencias: desde el primer año de EB (B1) hasta el final de la EM en que los libros de III° y IV° están reunidos (M34). Se crearon, por lo tanto, 44 listas de frecuencias, cuyos resultados por disciplina y nivel se muestran a continuación:

**Tabla 2.** Palabras distintas (*types*) por disciplina y nivel

Nivel	Types				Total
	L	CS	CN	M	
B1	7471	3217	1354	1304	13.346
B2	7212	4473	1873	2556	16.114
B3	12151	4815	3039	3924	23.929
B4	12152	4748	3720	4606	25.226
B5	11718	7529	3802	5922	28.971
B6	10532	7858	3700	5494	27.584
B7	8741	9349	5082	5340	28.512
B8	10941	9730	5660	4699	31.030
M1	12255	10386	12694	4610	39.945
M2	12528	10760		5042	36.057
M34	15078	9790	7727	5476	30.344
<b>Total</b>	120.779	82.655	48.651	48.973	301.058

A partir de las listas de palabras distintas (*types*) se creó una nueva lista con las *palabras comunes a las 4 disciplinas*, objeto específico del presente estudio. Para ello, se seleccionaron aquellas palabras que aparecían al menos una vez en las 4 materias, lo que resultó en un

conjunto de 6540 palabras. Luego, se eliminaron automáticamente las palabras gramaticales (preposiciones, artículos, pronombres, cuantificadores, posesivos, etc.), todas las formas de los verbos *ser*, *estar* y *haber*, los nombres propios y las palabras con números o símbolos. Las palabras resultantes fueron etiquetadas de acuerdo con su frecuencia en el Corpus del Español del Siglo XXI (CORPES, Real Academia Española 2024), utilizado como corpus de referencia. Se establecieron 4 categorías: palabras que estaban dentro de las 1000, 3000 y 5000 formas más frecuentes del CORPES, y aquellas que estaban fuera de esos rangos. Este procedimiento se llevó a cabo para complementar los criterios de selección en fases posteriores del proceso de creación de las listas de frecuencias.

Posteriormente, en un proceso ciego, 4 investigadores etiquetaron manualmente en una planilla de cálculo las palabras que, a juicio de cada uno, eran palabras “de la escuela”. El criterio de inclusión, por lo tanto, fue determinar si eran o no palabras que los niños aprenden en la escuela a través de los textos escritos y que, a la vez, están relativamente ausentes de su vida cotidiana. Así, quedaron fuera de la lista términos como *escalera*, *llovía*, *grande*, *quiso*, etc., muchas de las cuales, además, estaban en el rango de las 1000 o las 3000 palabras más frecuentes del idioma de acuerdo con el CORPES. Esta comparación con el corpus de referencia sirvió para verificar que se trataba de palabras de alta frecuencia en la vida cotidiana; sin embargo, no se utilizó como criterio estricto. La razón es que algunas palabras muy frecuentes en el idioma como *comunicación* y *economía* (ambas en el rango de las 1000 más frecuentes del CORPES) son términos que los niños aprenden y utilizan principalmente en la escuela y por eso quedaron incluidas en la lista. En cambio, términos como *gente* (también entre las 1000 más frecuentes) o *cabritas* (fuera de las 10 000, es decir, muy infrecuente) fueron excluidas porque se consideraron palabras que se aprenden fuera de la escuela. Después de este etiquetado y de numerosas revisiones por parte del equipo para resolver los casos dudosos, se obtuvo una lista de 2792 palabras distintas o *types* comunes a las 4 materias de la escuela.

### 2.3. Lematización

En consonancia con estudios previos sobre vocabulario académico (Gardner y Davies 2014), se optó por trabajar con la unidad “lema”, es decir, la forma básica de la palabra que representa todas sus posibles variantes flexivas (Bosque y Demonte 1999). La categoría de “familia de palabras”, también adoptada por estudios sobre vocabulario académico en inglés (Coxhead 2000, 2020), resultaba demasiado “gruesa” y habría generado la pérdida de vocablos importantes al eliminar la distinción por categoría gramatical, relevante en una lengua de morfología rica como el español.

El proceso de lematización se hizo de forma manual, debido a la naturaleza de los datos recogidos en el corpus. La diagramación de los textos escolares incluye columnas, notas al margen, imágenes con pie de foto, ejercicios de completación, etc., diagramación que, durante la extracción del texto, se ve alterada. Es posible, entonces, que queden oraciones fragmentadas, lo que puede afectar la precisión de los resultados en un procedimiento de lematización automática, que requiere del contexto oracional para poder asignar correctamente las categorías gramaticales. Por ello, y dado que el volumen de formas ortográficas era abarcable para el equipo de investigación, se decidió lematizar de forma manual.

Se obtuvo así la lista definitiva de 1727 lemas comunes, identificada de ahora en adelante como la **Lista de Vocabulario Escolar (LVE)**<sup>3</sup>. Como se comentará, son palabras transversales que aparecen en todos o casi todos los niveles de escolaridad: solo 9 palabras aparecen en 3 niveles, mientras que 1144 lemas (85 %) aparecen en 9, 10, u 11 niveles.

---

<sup>3</sup> Por razones de espacio no es posible adjuntarla, pero puede consultarse y descargarse en el repositorio institucional de datos mediante el siguiente enlace: <<https://doi.org/10.60525/04teye511/5IXBFX>>

### 3. RESULTADOS

#### 3.1. Distribución de lemas comunes por disciplina y nivel

Las palabras que son objeto del estudio se presentan en todas las materias y de manera bastante frecuente en todos o en varios niveles de la trayectoria escolar, aunque con algunas particularidades, como se puede ver en la Tabla 3, con la *cantidad* de lemas de la LVE que aparecen en los textos por disciplina y nivel.

**Tabla 3.** Cantidad total de lemas comunes por disciplina y nivel

Nivel	Lenguaje	CS	CN	Matemáticas	Total	Total por ciclo
B1	612	423	216	183	1434	9.451
B2	658	629	304	344	1935	
B3	1123	707	557	565	2952	
B4	1017	750	706	657	3130	
B5	1101	1090	762	812	3765	15.295
B6	1041	1061	756	834	3692	
B7	960	1118	917	907	3902	
B8	1014	1126	955	841	3936	
M1	1119	1187	1179	860	4345	13.458
M2	1099	1212	1119	945	4375	
M34	1247	1222	1247	1022	4738	
<b>Total</b>	10991	10525	8718	7970	38.204	<b>38.204</b>

La cobertura de estas palabras en los textos, como se muestra, es mayor en los niveles superiores y mayor también en las materias humanistas (Lenguaje y CS). Al mirar los resultados por ciclo se observa que el segundo ciclo básico (B5 a B8) concentra la mayor cantidad de estas palabras (15 295). Al observar las diferencias por disciplina, por otro lado, vemos que las palabras se distribuyen de manera bastante desigual en los primeros niveles de EB: en B1 Lenguaje aporta 612 de estas palabras, mientras que Matemáticas

aporta solo 183 (menos de un tercio de las aportadas por Lenguaje). En los niveles terminales de la enseñanza, en cambio, los vocablos se distribuyen de una manera más homogénea: 1247 en Lenguaje (el 72 % del total de la lista) y 1022 en Matemáticas (el 59 % de las palabras de la lista), para el nivel M34.

Ello se explica porque, en las áreas de CN y Matemáticas, los libros incorporan más lecturas en estos niveles. Además, los contenidos, ejercicios y exposiciones hacen uso de un vocabulario más académico, como se ilustra en el Ejemplo 1b, a continuación, en que todas las palabras subrayadas son parte de la lista (LVE). En EB, en cambio, estas materias suelen utilizar un lenguaje más accesible para la formulación de problemas y situaciones, como se muestra en el Ejemplo 1a, junto con un vocabulario escaso pero especializado del área (en cursiva en el ejemplo), con palabras como *cráneo*, *fémur* o *vertebral* —que no forman parte de la LVE—, específicas de la disciplina:

**Ejemplo 1.** *Fragmentos de a) CN, 2º Básico y b) Biología 1º Medio*

a.

¿Qué pasaría si tus huesos se volvieran débiles y blandos? Descúbrelo con esta actividad. [texto multimodal sobre dejar un hueso en un vaso con vinagre]

Los huesos son firmes y rígidos. Están unidos entre sí mediante *articulaciones*. Las *articulaciones* facilitan el movimiento del cuerpo. La columna *vertebral* es el sostén del cuerpo. El *cráneo* protege el cerebro. Las *costillas* protegen el corazón y los *pulmones*. El *fémur* participa en el movimiento de las piernas.

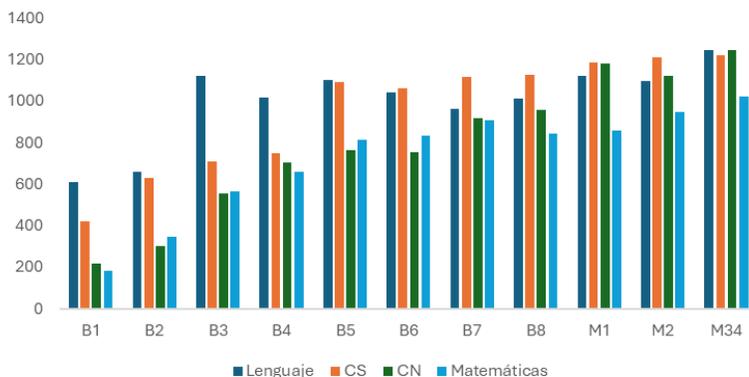
b.

Se trata de formular una interrogante a partir de una observación previa y de la relación de las variables involucradas. Una hipótesis es una respuesta tentativa a una pregunta de investigación. Esta debe ser lógica, comprobable y sustentada en

una teoría. Una inferencia es una interpretación o explicación de un hecho o fenómeno a partir de observaciones previas

El Gráfico 1 ilustra las cifras de la Tabla 3 y permite observar cómo Lenguaje, la materia que más aporta a este vocabulario hasta B5, desde B6 en adelante se empareja con CS, disciplina que supera a Lenguaje en los 4 niveles siguientes. Estas dos materias aportan, de hecho, una proporción cercana al 70 % de este vocabulario, entre B1 y B3, mientras que las CN y las Matemáticas aportan aproximadamente la mitad de esa proporción. En la etapa siguiente, desde B4 a B8, vemos que las materias humanistas todavía superan a las demás, aunque las diferencias se acortan; más adelante se observa cómo todas las disciplinas contribuyen de manera bastante similar.

**Figura 1.** Cantidad de lemas comunes por disciplina y nivel



Al observar el comportamiento general de las 4 disciplinas a lo largo de la trayectoria escolar, se constata que tienen perfiles diferentes: Lenguaje, la materia con más palabras en total, muestra un gran crecimiento en los primeros años (B1 a B3) y en los niveles posteriores se mantiene relativamente estable; las CS y CN presentan un aumento sostenido y constante durante toda la escolaridad, aunque más pronunciado en CS; finalmente, Matemáticas

muestra un aumento constante en casi toda la EB (de B1 a B7) y después tiende a mantenerse estable, con un leve crecimiento al final de la EM. Es relevante destacar el hecho de que muchas de estas palabras no son altamente especializadas, y que su conocimiento y comprensión desde los primeros años de escolaridad puede facilitar el acceso a los textos y al lenguaje de la escuela en general.

Ahora bien, si se considera no ya la cobertura de la lista sino la *frecuencia* de las palabras, es decir, todas sus ocurrencias en las 4 disciplinas y los 12 niveles, se observan algunas variaciones interesantes. Lenguaje sigue siendo una de las materias que más oportunidades ofrece para el encuentro con estas palabras en los niveles de EB, pero no ya con Ciencias Sociales sino con Matemáticas, tal como se ilustra en el Gráfico 2 (se han agrupado los seis primeros niveles en una primera etapa de “Básica” y los otros seis niveles en la etapa de “Media”).

Es decir que, aunque Matemáticas era la disciplina con menos palabras en general, debido a su alta repetición presenta las frecuencias más altas de palabras académicas en toda la trayectoria escolar, seguida de Lenguaje en Básica y muy cercana a CS en Media. Ello se debe en parte a que en Matemáticas se utilizan reiteradamente palabras abstractas alusivas a dimensiones temporales, espaciales o formales, tales como *vertical*, *curva* o *patrón*, y también relacionadas con actividades escolares y procesos mentales como *analizar*, *resolver* o *equivaler*. El Ejemplo 2, a continuación, reproduce toda la información verbal contenida en una sola página del libro de Matemáticas de B5 (la información restante son tablas, gráficas y otros tipos de material visual que acá se han omitido). El fragmento seleccionado tiene 103 palabras, de las cuales 60 (58 %) son de contenido léxico y 43 (42 %) son gramaticales (conjunciones, preposiciones, artículos, etc.). Entre las palabras de contenido léxico, 38 (subrayadas en el ejemplo) son parte de la LVE, es decir, el 37 % del texto, una proporción muy alta. Como puede apreciarse, entonces, es un texto con alta densidad léxica y con gran proporción de vocabulario académico general, aspectos relevantes a tener en cuenta para la comprensión de los estudiantes en los primeros años de escolaridad:

**Ejemplo 2.** *Fragmento del texto de matemáticas, B5*

Escribe los números representados en la tabla de valor posicional.

Señala si cada afirmación es verdadera o falsa. Justifica.

Resuelve los problemas.

Verifica las afirmaciones y explica quiénes están en lo correcto.

Compara y ordena estas cantidades de menor a mayor. Luego, ubica los números en una recta numérica. Apóyate en el recortable sugerido.

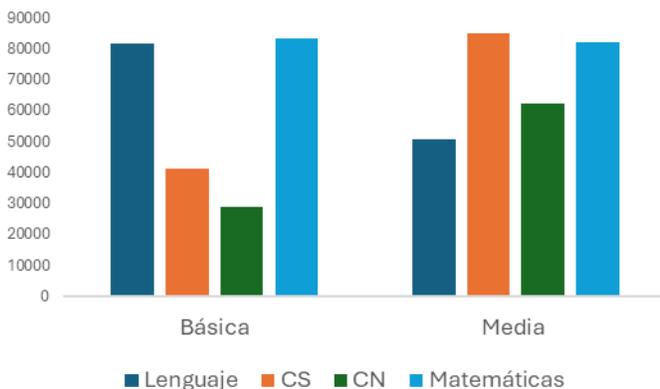
Revisa los ejemplos en los que se han representado distintos números de diferentes formas.

Representa de todas las formas que conozcas un número de 8 cifras.

Revisa los procedimientos asociados a la aproximación y comparación.

Aproxima a la centena de mil el número de 8 cifras representado.

Si bien hacen falta más estudios, es muy interesante tener en cuenta el potencial que ofrece esta disciplina para aprender y practicar este tipo de vocabulario con una perspectiva interdisciplinar durante toda la trayectoria escolar. En cuanto a Lenguaje, se reafirma su importancia no solo por su gran proporción de las palabras de la LVE, como se señaló en el apartado anterior, sino también por la alta repetición de estas palabras durante la enseñanza básica, posición que es menos prominente en EM y reemplazada por el aporte de Ciencias Sociales en esta etapa más avanzada.

**Figura 2.** Frecuencia de lemas comunes por disciplina y etapa escolar

### 3.2. Propuesta de clasificación

Como se ha explicado, este vocabulario es frecuente en los textos académicos y escaso en otros tipos de texto y en el discurso coloquial. Sus palabras son desafiantes porque suelen tener definiciones abstractas y significados diversos de acuerdo con las distintas áreas de contenido. Además, es un vocabulario muy variado, en un continuo que va desde términos frecuentes en todas las materias y niveles con gran transversalidad y polivalencia, hasta términos especializados frecuentes en una sola disciplina. En este corpus en particular, hemos observado algunas tendencias que permiten identificar ciertos grupos de palabras, lo que a su vez contribuye a comprender mejor sus rasgos. Estos grupos coinciden en parte con otras clasificaciones (Baumann y Graves 2010, Hyland y Tse 2007), lo que ayuda a verificar la naturaleza “académica” de este vocabulario. Esta propuesta es una primera aproximación a este corpus y no constituye una clasificación definitiva para este tipo de palabras. Además, como es natural tratándose de unidades semánticas, estos grupos no tienen fronteras claramente definidas y muchas palabras pueden ser consideradas como pertenecientes a más de un grupo.

Para su clasificación se hizo manualmente un etiquetado ciego entre 4 investigadores, teniendo en cuenta, junto con el juicio cualitativo, las frecuencias de las palabras, como se comentará. En este proceso se identificaron tres tipos de palabras en la lista: 1) Metalenguaje de textos y tareas escolares; 2) Palabras polivalentes; 3) Palabras disciplinares.

#### *A) Tipo 1. Metalenguaje de textos y tareas escolares*

Las palabras de este tipo pertenecen al dominio de los textos y las tareas escolares, así como de los actos verbales y cognitivos asociados a ellas, tal como se ha descrito en otras propuestas (Beck, McKeown y Kucan 2013). Son palabras de alta frecuencia en todos los niveles y materias, por lo que probablemente no imponen gran dificultad en la comprensión, porque se aprenden de manera incidental. Sin embargo, es importante tenerlas en cuenta en los primeros años de la escolaridad y también en el aprendizaje de estudiantes no nativos (Nation 2016; Dang, Webb y Coxhead 2022). En ambos casos, son parte importante de las instrucciones orales que se imparten en el aula y también de los textos instructivos en materiales y evaluaciones. A pesar de su gran frecuencia y accesibilidad, incluyen grados diversos de complejidad por su abstracción y polisemia, que resultan sin duda desafiantes para los más jóvenes. Así, por ejemplo, las palabras *cuaderno* o *responder* de la Tabla 4 se ubicarían en una zona de mayor accesibilidad, pero otras como *ejemplo*, *fuentes* o *relacionar* ofrecen más matices y deberían ser parte del lenguaje productivo de los niños desde los primeros años.

A continuación, presentamos, a modo de ejemplo, 10 palabras de cada categoría gramatical (omitiendo los adverbios), aunque enfatizamos que las categorías principales y más numerosas en este grupo de palabras son los sustantivos y los verbos. Los adjetivos son, naturalmente, palabras menos frecuentes en general y presentan un mayor grado de abstracción. Por lo tanto, pueden llegar a presentar mayores dificultades que las otras dos categorías.

**Tabla 4.** Ejemplos de lemas Tipo 1

Sustantivos	Adjetivos	Verbos
actividad	anterior	analizar
cuaderno	contrario	completar
ejemplo	correcto	describir
fuelle	equivalente	escribir
lección	general	explicar
palabra	informativo	leer
problema	principal	observar
respuesta	relevante	relacionar
texto	siguiente	resolver
unidad	total	responder

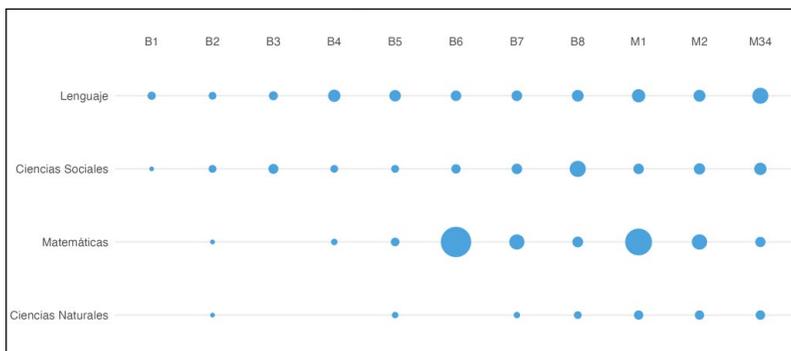
### *B) Tipo 2. Vocabulario académico general*

En este segundo grupo, las palabras son también muy frecuentes en las 4 disciplinas y la mayoría de ellas corresponden a las llamadas “palabras académicas generales” (Hiebert y Lubliner 2008). Se caracterizan por sus significados abstractos y diversos de acuerdo con la disciplina, por lo que representan un desafío mayor para la comprensión de los textos, particularmente en etapas tempranas de la escolaridad. Por su naturaleza, corresponden más bien a sustantivos abstractos y en algunos casos bastante polisémicos (como las palabras *columna* o *sello*); en otros casos, son palabras no tan polisémicas, pero cuyos usos y sentidos cambian mucho de acuerdo con el contexto disciplinar (como la palabra *residuo*). Además de sustantivos, hay otras categorías de palabras que también adquieren distintos significados según el área, como los adjetivos *real*, *grave* y *complejo*, o los verbos *justificar*, *registrar* y *tratar*. Sin embargo, en esta exposición nos centraremos en los sustantivos, algunos de cuyos ejemplos se presentan en la siguiente lista:

**Tabla 5.** Ejemplos de lemas Tipo 2

campo	factor	raíz
carácter	fuelle	razón
columna	fuerza	residuo
demanda	género	resistencia
dimensión	juicio	sello
dominio	materia	signo
eje	modelo	sistema
elemento	potencia	tendencia
escala	punto	valor
estado	radio	volumen

Dada su polivalencia y la imposibilidad de aprenderlas mediante definiciones unívocas, son palabras que requieren de más oportunidades de lectura y uso para ser aprendidas en sus diversos sentidos y posibilidades. Es decir, la exposición y la práctica reiteradas en contextos auténticos hacen posible que los estudiantes se apropien de ellas y las usen de manera flexible; de esta manera, sustentan diversos significados de ideas especializadas (Nagy y Townsend 2012). Desde el punto de vista del desarrollo léxico, el contacto temprano y reiterado de los niños con este tipo de vocabulario favorece la profundización de sus representaciones conceptuales con diversos sentidos y perspectivas, lo que estimula formas de pensamiento más flexibles y complejas (Tolchinsky 2004). Un ejemplo de esto puede verse en la Figura 3 a continuación, que ilustra la frecuencia y distribución disciplinar de la palabra *razón*:

**Figura 3.** Frecuencia y distribución del lema *razón*

Como muestra la figura anterior, la palabra *razón* tiene rasgos disciplinares más marcados asociados al área de las Matemáticas (círculos más grandes en la figura), particularmente a partir de sexto año básico, aunque está presente en todos los niveles de la escolaridad, de B1 a M34 y en todas las disciplinas. Es interesante destacar que los estudiantes están expuestos a esta palabra desde primero básico tanto en Lenguaje como en CS, antes de adquirir el sentido disciplinar propio de las Matemáticas, distinto y específico. Como puede verse en los casos del Ejemplo 3 a continuación, la palabra *razón* se presenta desde temprano con distintas funciones gramaticales (como sustantivo, como parte de un conector causal, de una perífrasis verbal, de una frase adverbial, etc.) y en distintos contextos discursivos y registros (en textos expositivos, instruccionales o literarios, por ejemplo). Como sustantivo, además, sus sentidos y significados son también diversos (con el sentido de proporción, argumento, motivo, forma de pensar, cualidad humana, valor de una época, o noción metalingüística, entre otros):

### Ejemplo 3. Algunos usos del lema *razón* en los textos

- No sabemos mucho acerca de estos peces, **parte de la razón por la que** los investigadores siguen fascinados por ellos (CS-B2).

- **Con sobrada razón** me ha sucedido esto, porque siendo yo cazador, no debí meterme a flautista (Lenguaje-B5).
- el cociente de dos números consecutivos de la serie se aproxima a la denominada «razón dorada», «sección áurea» o «divina proporción» (Lenguaje-B6).
- Y como le **diera razón** al Gran Khan de la gran riqueza de esta isla —Cublai, que entonces reinaba—, quiso apoderarse de ella (CS-B6).
- **Escribe la razón** entre la cantidad de cuadrados y la cantidad total de figuras (Matemáticas-B6).
- **Expresemos como razón** 30 ml de vinagre y 60 ml de soya (Matemáticas-B6).
- **Elevar la razón** como el modo de entender el mundo (predominio de la razón), y crear idealizados y estrictos principios de belleza (CS-B7).
- ...instrumento de conocimiento por excelencia es **la razón humana** (CS-B7).
- “¡Qué obra de arte es el hombre! ¡Qué noble **su razón!**” (CS-B7)
- Todo español o criollo blanqueado era **gente de razón**, pero solo quien entre estos tenía riqueza y posición era considerado gente decente... (CS-M1).
- **La voz de la razón** y de la verdad se oirán entre nosotros después del triste, e insufrible silencio de tres siglos (CS-M1).

### C) Tipo 3. Vocabulario académico disciplinar

Aunque las palabras de este tipo también tienen transversalidad, su presencia es mucho mayor en un área, es decir, son palabras más específicas de un campo disciplinar. Podrían corresponder a un vocabulario más técnico en la nomenclatura de Hiebert y Lubliner (2008) o Fisher y Frey (2008), pero en este corpus aparecen en todas las materias al menos una vez. Por ello las hemos considerado como parte de la LVE, aunque se ubican en una zona más próxima a las palabras especializadas. A modo de ejemplo, se presentan en la siguiente tabla las palabras más frecuentes en cada disciplina (sobre

el 70 % más frecuentes en cada una), pero infrecuentes en las demás. Se presentan por categoría gramatical, aunque son los sustantivos el tipo más abundante en un corpus como este (donde abundan exposiciones, instrucciones y argumentaciones), dada su capacidad de objetivar, conceptualizar y sintetizar los contenidos disciplinares (Biber y Gray 2016, Soto y Zenteno 2004). No obstante, en Lenguaje y también en Matemáticas, como se ha comentado, hay también muchos verbos bastante propios de la disciplina (como *resumir* o *multiplicar*), mientras que en áreas más tradicionalmente centradas en los contenidos conceptuales (CS, CN), los verbos no suelen ser tan disciplinares. De hecho, y dada la naturaleza de esta categoría gramatical, muchos de los verbos que aparecen con altas frecuencias en una disciplina son también frecuentes en otras (por ejemplo, *explorar* o *validar*), pero cada uno adquiere un matiz diferente según el área. Así por ejemplo, *describir* en Lenguaje no implica los mismos procesos y habilidades que *describir* en Geografía física o CN. Los adjetivos, en tanto, suelen ser de tipo *relacional*, más habituales en el lenguaje académico y en los textos expositivos. Es decir, no aportan cualidades a la entidad u objeto que determinan, sino que los relacionan con otras entidades de la misma clase al otorgarle un valor especificativo e incluso a veces, clasificatorio, como en (*zona*) *andina* o (*figura*) *geométrica* y los demás ejemplos de la Tabla 6, a continuación:

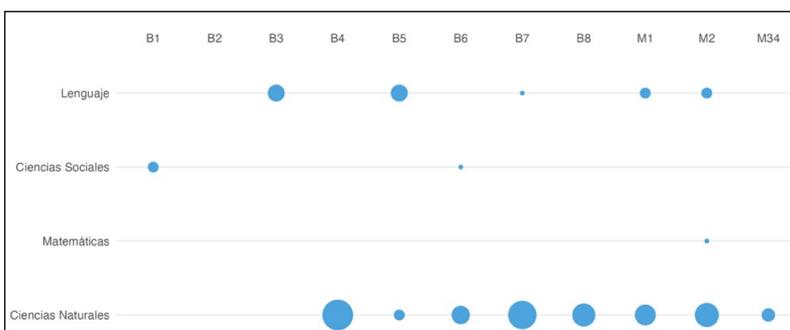
Tabla 6. Ejemplos de lemas de Tipo 3

	Lenguaje	CS	CN	Matemáticas
Sustantivos	cita	administración	atmósfera	ángulo
	ficción	civilización	carbono	azar
	fragmento	constitución	conducción	circunferencia
	leyenda	demanda	corteza	curva
	literatura	economía	fertilidad	dispersión
	mensaje	explotación	núcleo	ecuación
	metáfora	gobierno	observación	exponente
	personaje	mercado	radiación	intervalo
	significado	patrimonio	reproducción	tangente
	versión	transición	tallo	vértice
Adjetivos	explícito	andino	combustible	diagonal
	informativo	cívico	empírico	geométrico
	lógico	industrial	sólido	numérico
	preciso	laboral	térmico	perpendicular
	recurrente	público	tóxico	probable
Verbos	convencer	administrar	absorber	calcular
	corregir	ejercer	experimentar	despejar
	destacar	implementar	explorar	graficar
	inferir	manifestar	incrementar	multiplicar
	resumir	resguardar	prevenir	sumar
	significar	restablecer	validar	resolver

En la Tabla 6 se pueden constatar las particularidades del vocabulario más disciplinar de esta lista. Las palabras del área de Lenguaje, debido a su transversalidad, suelen ser menos especializadas que las de otras materias y a veces coinciden con el metalenguaje de los textos escolares (como en las palabras *leer* o *destacar*). En CS encontramos muchas palabras específicas del contenido disciplinar, pero dado su carácter humanista —y por ello su transversalidad en diversos aspectos de la vida de las personas—, son palabras que

también se encuentran con frecuencia en la vida cotidiana y en los textos periodísticos, como es el caso de *administración* y *mercado*, por ejemplo. En CN y Matemáticas, las palabras con altas frecuencias son probablemente las más específicas de sus áreas, por ejemplo, palabras como *atmósfera* y *vértice*. Un ejemplo de este tipo de vocabulario se muestra en la Figura 4, a continuación, con la trayectoria de la palabra *corteza*:

**Figura 4.** Frecuencia y distribución del lema *corteza*



Como se puede apreciar, la palabra *corteza* está fuertemente anclada al contexto disciplinar de las Ciencias Naturales desde B4 hasta el término de la escolaridad (M34), pero aparece también en las demás materias, sobre todo en Lenguaje (B3) y también en CS, incluso más tempranamente (B1). Como se muestra en el Ejemplo 5, sus significados y usos no solo son más acotados que los de las palabras de Tipo 2 como *razón*, sino que todos ellos están relacionados de alguna manera con su sentido original de “parte o capa superficial” de algo (una fruta, un árbol, un órgano del cuerpo humano, un cuerpo celeste, etc.). Sin embargo, esto no impide que sus usos sean variados y crecientemente complejos en los saberes, estilos y discursos disciplinares (textos expositivos, literarios, instruccionales, etc.).

**Ejemplo 5.** *Algunos usos del lema corteza*

- Las canoas eran de **corteza de coihue** (CS-B1)
- La **corteza terrestre** está fragmentada como un gran puzle (CS-B6)
- Lóbulo frontal: área de la **corteza cerebral** que regula funciones como la atención y la planificación (Lenguaje-B7)
- ...sabemos que la región posterior de la **corteza prefrontal** media del cerebro contribuye a este sesgo de confirmación en seres humanos (Lenguaje-M2)
- La **corteza terrestre mide** un ancho promedio de 25 km. ¿Cuál es su volumen? (Matemáticas-M2)
- Explica cómo varía la temperatura **desde la corteza hacia el núcleo**. (CN-B4)
- Cada riñón posee dos regiones: la **corteza renal** y la médula renal. (CN-B8)
- Las áreas de asociación contiguas interpretan estas señales y, asimismo, las relacionan con los recuerdos almacenados en la **corteza**. (CN-M2)

**4. CONCLUSIONES Y DISCUSIÓN**

Para explorar el vocabulario común de los textos escolares, este estudio se propuso, en primer lugar, determinar cómo se distribuyen las palabras en rangos de frecuencia por disciplina y nivel escolar. Respecto de los niveles, se encontró que estas palabras están presentes en todos los niveles escolares, pero son más abundantes en el segundo ciclo básico y en Enseñanza Media, es decir, hay una cobertura mayor de la lista desde B5 en adelante. Al segmentar la escolaridad en tres ciclos, la mayor cantidad de estas palabras se concentran en el segundo ciclo básico (B5 a B8). Es natural que la cantidad de vocabulario académico vaya en aumento durante la escolaridad. Sin embargo, es interesante destacar, como se ha comentado, que estas palabras contribuyen desde los primeros años a la construcción de los conocimientos generales y disciplinares de la escuela.

En cuanto a las disciplinas, los resultados muestran distribuciones desiguales al inicio y una tendencia de homogeneización creciente: en los primeros años de enseñanza básica, las materias humanistas de Lenguaje y CS concentran una gran cantidad de estos vocablos y las CN y Matemáticas aportan una cantidad escasa, en proporción. En los años intermedios (B5 a B8), hay una distribución más pareja, aunque se mantiene la mayor proporción en las materias humanistas. En los años terminales (M1 a M34) todas las materias aportan de manera más equilibrada. Si se considera la aparición de estas palabras en términos de frecuencia (es decir, considerando todas sus apariciones en los textos), se verifica que las materias de Lenguaje (en básica) y CS (en media) siguen siendo las que más oportunidades ofrecen para el encuentro con estas palabras, junto con Matemáticas. Esta materia, a pesar de que tiene pocas palabras en general, hace un uso abundante de las palabras comunes que se repiten bastante en estos textos, por lo que ofrece así un gran potencial para el trabajo interdisciplinar con el léxico abstracto propio de la escuela.

Al considerar el comportamiento general de las 4 materias, se observan, en suma, dos factores importantes: por una parte, la identidad más humanista o más científica del área, y por la otra, su naturaleza más centrada en las habilidades o en los contenidos. Así, las materias humanistas contribuyen con una gran cantidad de este vocabulario en las distintas etapas de la escolaridad, particularmente en básica para el caso de Lenguaje y en enseñanza media para el caso de CS. Este resultado confirma el papel fundamental que cumple la lectura de fuentes externas (de textos literarios, periodísticos, históricos, etc.) en la conformación del conocimiento léxico durante la formación escolar.

Por otra parte, los resultados muestran que las materias más centradas en el desarrollo de habilidades (de razonamiento, de comprensión, etc.), es decir, Lenguaje y Matemáticas, utilizan una gran cantidad de palabras académicas a lo largo de toda la trayectoria (Lenguaje con mayor cantidad de tipos de palabras, y Matemáticas con menor cantidad de tipos, pero con mayor repetición). Ambas

contribuyen de manera significativa desde los primeros años de enseñanza básica y mantienen un aporte relativamente estable hasta el fin de la escolaridad. En cambio, las áreas más centradas en los contenidos, como CS y CN, muestran un perfil de crecimiento distinto: aportan una cantidad menor de estas palabras en los primeros años de enseñanza básica y las van incorporando de manera sostenida y creciente hasta el fin de la enseñanza media.

Los siguientes objetivos específicos eran proponer una clasificación de las palabras de esta lista y analizarlas en función de su mayor o menor grado de polivalencia y especialización disciplinar. Al respecto, los resultados replican lo señalado en otros estudios de vocabulario escolar y académico (Baumann y Graves 2010, Hyland y Tse 2007), en cuanto a la presencia de tres grandes tipos en un continuo gradual de especialización: palabras propias de textos y tareas escolares, vocabulario académico general y vocabulario académico especializado. Es una propuesta tentativa de categorías flexibles, cuyo interés es comprender mejor algunos de sus rasgos para profundizar en sus posibilidades de enseñanza. Así, el primer grupo reúne palabras muy frecuentes en los textos (y habituales incluso en la vida cotidiana) propias de las actividades escolares, tales como *figura* o *recurso*, que no suponen dificultad en la lectura, ya que se aprenden tempranamente debido a su alta repetición en las actividades de instrucción. Sin embargo, muchas de ellas son abstractas y polisémicas, como las palabras *fuentes* o *unidad*, y por lo tanto desafiantes para los menores, particularmente en su uso productivo, así como en su comprensión profunda. También, naturalmente, representan un desafío importante para los estudiantes que se escolarizan en una segunda lengua (Coxhead 2017, Nation 2016).

El segundo grupo reúne las palabras habitualmente señaladas como de vocabulario académico general, con alto grado de abstracción, transversalidad y polivalencia, tales como *género* o *valor*. Representan un desafío mayor que las del grupo anterior porque son menos frecuentes y más abstractas, pero, al igual que aquellas, también ofrecen un enorme potencial de trabajo léxico sistemático

y transversal. En efecto, un abordaje planificado e interdisciplinar con las palabras de ambos grupos desde el comienzo de la escolaridad contribuiría a ampliar el vocabulario de los estudiantes y, por lo tanto, a mejorar sus habilidades de comprensión y aprendizaje en general (Meneses, Uccelli y otros 2018). Además, contribuiría también a sus habilidades de pensamiento abstracto y razonamiento metalingüístico, las cuales movilizan y favorecen nuevos aprendizajes durante la trayectoria escolar.

El tercer tipo agrupa las palabras que, si bien aparecen en todas las materias, son mucho más frecuentes en una de ellas y escasas en las demás. Es el caso de vocablos como *huésped*, *intervalo* o *civilización*, con significados fuertemente anclados en una disciplina específica, pero no exentos de usos con sentidos diversos, incluso metafóricos. Por lo tanto, estas palabras ofrecen un desafío diferente. Se aprenden de manera menos incidental, mediante instrucción directa o definiciones en el contexto disciplinar y por eso contribuyen a la creación de nuevos dominios conceptuales de naturaleza especializada (Hüttner y Dalton-Puffer 2024). Sin embargo, también en este caso un abordaje sistemático e interdisciplinar ofrece la posibilidad de profundizar y ampliar esos dominios a la vez que ejercitar los usos figurados de estos términos tanto en la comprensión como en la producción oral y escrita.

El presente estudio y la lista de vocabulario que compartimos, aunque preliminares aún, pueden contribuir a ese trabajo sistemático e interdisciplinar con el vocabulario de los distintos niveles durante la escolaridad básica y media. En concreto, se podría, por ejemplo, seleccionar anualmente unos 30-40 lemas de alta polivalencia (por ejemplo, *razón*, *evidencia*, *modelo*, *variable*) y de metalenguaje escolar (*explicar*, *justificar*, *comparar*) para ejercicios tales como definir en contexto, dar ejemplos y contraejemplos, reformular y explicar el contraste disciplinar de sus usos. También se podrían trazar rutas interdisciplinarias por ciclo (B5 a B8 y M1 a M34) que exploten la alta repetición en Matemática y el aporte de Lenguaje/CS, mediante breves lecciones que incluyan actividades de “antes, durante y después” de la lectura y tareas de escritura

por género (explicación científica, comentario histórico, problema resuelto) que exijan el uso productivo del léxico. Además, se podría trabajar la polisemia y las familias morfológicas a través de mapas semánticos, nominalizaciones y adjetivos relacionales, o elaborar un cuaderno léxico disciplinar (con entrada y definición más ejemplo en dos asignaturas más error común). Muchas de estas actividades, como se ha explicado, contribuyen al desarrollo del pensamiento abstracto y a la reflexión gramatical y metalingüística, reflexión que, aunque ha estado ausente de los programas educativos de las últimas décadas (Rubio, Moya, Castellón, Harms y Navia 2023), constituye uno de los aspectos enfatizados en la última propuesta de ajuste curricular en Chile.

En cuanto a sus limitaciones, este trabajo se propuso analizar las palabras comunes de las 4 disciplinas, lo que favorece la transversalidad, pero subestima el vocabulario relevante específico de cada área. Esto último sería interesante de indagar en una nueva exploración. Además, tal como se expuso en la metodología, se optó por hacer una lematización manual, lo que reduce los errores de etiquetado automático, pero introduce criterios humanos y posibles sesgos de consistencia. En esta misma línea, la construcción de la lista implicó decisiones de inclusión y exclusión respecto de las “palabras de la escuela” tomadas por 4 investigadores, proceso que agrega subjetividad al procedimiento. Por último, sería muy interesante incluir, además del componente verbal, el vocabulario contenido también en los materiales visuales y multimodales, que forman parte importante de los textos escolares.

La enseñanza y el aprendizaje del vocabulario académico general y disciplinar constituyen una tarea relevante en el desarrollo educativo de los estudiantes a lo largo de toda su escolaridad. Como se ha expuesto, este vocabulario no solo es esencial para mejorar la comprensión lectora, sino también para acceder y construir los conocimientos en las diversas áreas del saber. Abordar su enseñanza sistemática permite integrar diferentes contextos y enfoques, y así enriquecer las experiencias de aprendizaje y facilitar una comprensión más profunda de los contenidos. Por lo tanto, fortalecer estas

competencias es necesario para garantizar una educación integral y favorecer un adecuado desarrollo no solo lingüístico, sino también conceptual y cognitivo.

## REFERENCIAS BIBLIOGRÁFICAS

- BAUMANN, James F.; y GRAVES, Michael F.  
2010 “What is academic vocabulary?”. *Journal of Adolescent & Adult Literacy*. 54, 1, 4.
- BECK, Isabel; MCKEOWN, Margaret; y KUCAN, Linda  
2013 *Bringing words to life: Robust vocabulary instruction*. 2ª ed. Nueva York: Guilford Press.
- BIBER, Douglas; y GRAY, Bethany  
2016 *Grammatical Complexity in Academic English: Linguistic Change in Writing*. Cambridge: Cambridge University Press.
- BONDI, Marina  
2016 “Textbooks”. En *The Routledge handbook of English for academic purposes*. Eds., Ken Hyland y Philip Shaw. Londres: Routledge, 323-334.
- BOSQUE, Ignacio; y DEMONTE, Violeta. (Eds.)  
1999 *Gramática descriptiva de la lengua española*. Madrid: Espasa.
- CHUNG, Teresa Mihwa; y NATION, Paul  
2003 “Technical vocabulary in specialised texts”. *Reading in a Foreign Language*. 15, 2, 103-116. <http://hdl.handle.net/10125/66770>
- COXHEAD, Averil  
2000 “A new academic word list”. *TESOL Quarterly*. 34, 2, 213-238. <https://doi.org/10.2307/3587951>
- COXHEAD, Averil  
2017 “Academic vocabulary in teacher talk: Challenges and opportunities for pedagogy”. *Oslo Studies in Language*. 9, 3. <https://doi.org/10.5617/osla.5845>
- COXHEAD, Averil  
2020 “Academic Vocabulary”. En *The Routledge Handbook of Vocabulary Studies*. Ed., Stuart Webb. Londres: Routledge, 97-110.

COXHEAD, Averil; y NATION, Paul

2001 “The specialised vocabulary of English for academic purposes”. En *Research perspectives on English for academic purposes*. Eds., John Flowerdew y Matthew Peacock. Cambridge: Cambridge University Press, 252-267.

DANG, Thi Ngoc Yen; WEBB, Stuart; y COXHEAD, Averil

2022 “Evaluating lists of high-frequency words: Teachers’ and learners’ perspectives”. *Language Teaching Research*. 26, 4, 617-641. <https://doi.org/10.1177/1362168820911189>

FISHER, Douglas; y FREY, Nancy

2008 *Word wise and content rich: Five essential steps to teaching academic vocabulary*. Portsmouth: Heinemann.

FITZGERALD, Jill; RELYEA, John E.; y ELMORE, John

2022 “Academic vocabulary volume in elementary grades disciplinary textbooks”. *Journal of Educational Psychology*. 114, 6, 1257. <https://doi.org/10.1037/edu0000735>

FRIEDL, Jeffrey

2006 *Mastering regular expressions*. O’Reilly Media, Inc

GARDNER, Dee; y DAVIES, Mark

2014 “A new academic vocabulary list”. *Applied Linguistics*. 35, 3, 305-327. <https://doi.org/10.1093/applin/amt015>

GRAVES, Michael

2016 *The Vocabulary Book: Learning and Instruction*. 2ª ed. Nueva York: Teachers College Press.

HIEBERT, Elfrieda H.; y LUBLINER, Shira

2008 “The nature, learning, and instruction of general academic vocabulary”. En *What research has to say about vocabulary instruction*. Eds., Alan E. Farstrup y S. Jay Samuels. Newark: International Reading Association, 106-129.

HÜTTNER, Julia; y DALTON-PUFFER, Christiane (eds.)

2024 *Building Disciplinary Literacies in Content and Language Integrated Learning*. Londres: Taylor & Francis.

HYLAND, Ken; y TSE, Polly

2007 “Is there an ‘Academic Vocabulary?’”. *TESOL Quarterly*. 41, 2, 235-253. <https://doi.org/10.1002/j.1545-7249.2007.tb00058.x>

- KOVAČ, Miha; y KEPIC, Alenka  
 2022 “The changing role of textbooks in primary education in the digital era: what can we learn from reading research?”. *CEPS Journal*. 12, 2, 11-27. <https://doi.org/10.25656/01:25210>
- KOVAČ, Miha; y VAN DER WEEL, Adriaan  
 2020 “La lectura en una era postextual”. En *Lectura en papel vs. lectura en pantalla*. Eds., Miha Kovač y Adriaan van der Weel. Bogotá: CERLALC, 11-30.
- LOHNDORF, Regina; VERMEER, Harriet; CÁRCAMO, Rodrigo; y MESMAN, Judi  
 2018 “Preschoolers’ vocabulary acquisition in Chile: the roles of socioeconomic status and quality of home environment”. *Journal of Child Language*. 45, 3, 559-580. <https://doi.org/10.1017/S0305000917000332>
- LÜDELING, Anke; y KYTÖ, Merja  
 2009 *Corpus Linguistics: An International Handbook*. Vol. 2. Berlín/Nueva York: Mouton de Gruyter.
- MENESES, Alejandra; HUGO, Evelyn; MONTENEGRO, Maximiliano; VALENZUELA, Andrea; y RUIZ, Marcela  
 2018 “Explicaciones científicas: propuestas para la enseñanza del lenguaje académico”. *Boletín de Lingüística*. 30, 49-50, 134-157.
- MENESES, Alejandra; UCCELLI, Paola; SANTELICES, Verónica; ACEVEDO, Daniel; y FIGUEROA, Javiera  
 2018 “Academic language as a predictor of reading comprehension in monolingual Spanish-speaking readers: Evidence from Chilean early adolescents”. *Reading Research Quarterly*. 53, 2, 223-247. <https://doi.org/10.1002/rrq.192>
- NAGY, William; y TOWNSEND, Dianna  
 2012 “Words as tools: Learning academic vocabulary as language acquisition”. *Reading Research Quarterly*. 47, 1, 91-108. <https://doi.org/10.1002/RRQ.011>
- NATION, I. S. Paul  
 2006 “How large a vocabulary is needed for reading and listening?”. *Canadian Modern Language Review*. 63, 1, 59-82. <https://doi.org/10.3138/cmlr.63.1.59>

NATION, I. S. Paul

2016 *Making and Using Word Lists for Language Learning and Testing*. Ámsterdam: John Benjamins.

OLIVERA, María Paz

2017 “Estudio de uso y valoración de textos escolares”. *Biblioteca Digital del Gobierno de Chile*. <https://bibliotecadigital.mineduc.cl/handle/20.500.12365/18822>

OOMS, Jeroen

2024 “Package ‘pdftools’: Text Extraction, Rendering and Converting of PDF Documents”. Versión 3.4.0. *The Comprehensive R Archive Network*. <<https://cran.r-project.org/web/packages/pdftools/pdftools.pdf>>. Consultado: 10 de diciembre de 2025.

PERFETTI, Charles y HELDER, Anne

2022 “Progress in reading science: Word identification, comprehension, and universal perspectives”. En *The Science of Reading: A Handbook*. Eds., Margaret J. Snowling y Charles Hulme. 2ª ed. Hoboken: Wiley-Blackwell, 5-35.

R CORE TEAM

2024 *R: A Language and Environment for Statistical Computing*. Viena: R Foundation for Statistical Computing. <<https://cran.r-project.org/doc/manuals/r-release/fullrefman.pdf>>. Consultado: 10 de diciembre de 2025.

REAL ACADEMIA ESPAÑOLA

2024 *Corpus del Español del Siglo XXI (CORPES)*. <<https://www.rae.es/corpes/>>. Consultado: 10 de diciembre de 2025.

RIVEROS, Francisco

2021 “La política de textos escolares en Chile, un balance al 2018”. En *Investigaciones sobre libros de texto y medios de enseñanza*. Eds., Graciela María Carbone, Jesús Rodríguez, Nilson Días y Tânia Braga García. Curitiba: Universidad Federal de Paraná, 117-146.

ROMERO-CONTRERAS, Silvia; SILVA-MACEDA, Gabriela; y SNOW, Catherine

2021 “Vocabulario académico y habilidades de lenguaje académico: predictores de la comprensión lectora de estudiantes de primaria y secundaria en México”. *Pensamiento Educativo*.

*Revista de Investigación Educativa Latinoamericana (PEL)*.  
58, 2, e4. <https://doi.org/10.7764/PEL.58.2.2021.4>

- RUBIO, María; MOYA, Patricio; CASTELLÓN, María; HARMS, Valeria; y NAVIA, Javiera  
2023 “Enseñanza de la gramática: Una revisión bibliográfica (2010-2021)”. *Nueva Revista del Pacífico*. 78, 259-280. <http://dx.doi.org/10.4067/S0719-51762023000100259>
- SHANAHAN, Timothy; y SHANAHAN, Cynthia  
2008 “Teaching Disciplinary Literacy to Adolescents: Rethinking Content-Area Literacy”. *Harvard Educational Review*. 78, 1, 40-59. <https://doi.org/10.17763/haer.78.1.v62444321p602101>
- SOTO VERGARA, Guillermo; y ZENTENO BUSTAMANTE, Carlos  
2004 “Los sintagmas nominales en textos científicos escritos en español”. *ELUA: Estudios de Lingüística. Universidad de Alicante*. 18, 275-292. <http://dx.doi.org/10.14198/ELUA2004.18.14>
- STØLE, Hildegunn  
2020 “El mito del nativo digital: ¿por qué necesitan libros?”. En *Dossier lectura digital en la primera infancia*. Ed., CERLALC. Bogotá: CERLALC, 49-69.
- TOLCHINSKY, Liliana  
2004 “The nature and scope of later language development”. En *Language development across childhood and adolescence*. Ed., Ruth A. Berman. Amsterdam: John Benjamins, 233-248.
- TOWNSEND, Dianna; FILIPPINI, Alessandra; COLLINS, Penny; y BIANCAROSA, Gina  
2012 “Evidence for the importance of academic word knowledge for the academic achievement of diverse middle school students”. *The Elementary School Journal*. 112, 3, 497-518. <https://doi.org/10.1086/663301>
- UCCELLI, Paola; y MENESES, Alejandra  
2015 “Habilidades de lenguaje académico y su aplicación con la comprensión de la lectura en la escuela primaria y media: un nuevo constructo operacional”. *Miríada Hispánica*. 10, 179-206.

WEBB, Stuart

2020 "Incidental vocabulary learning". En *The Routledge Handbook of Vocabulary Studies*. Ed., Stuart Webb. Londres: Routledge, 225-239.

WEBB, Stuart; y CHANG, Anna C. S.

2015 "How does prior word knowledge affect vocabulary learning progress in an extensive reading program?". *Studies in Second Language Acquisition*. 37, 4, 651-675. <https://doi.org/10.1017/S0272263114000606>

WEBB, Stuart; y NATION, Paul

2017 *How Vocabulary is Learned*. Oxford: Oxford University Press.

WEISLEDER, Adriana; y FERNALD, Anne

2013 "Talking to children matters: early language experience strengthens processing and builds vocabulary". *Psychological science*. 24, 11, 2143-2152. <https://doi.org/10.1177/0956797613488145>

Recepción: 04/03/2025

Aceptación: 11/10/2025