

Documento de Trabajo N° 16, Octubre 2014

Análisis de la calidad del léxico y su relación con la calidad de la escritura en niños chilenos de cuarto año de educación primaria

Gabriela Gómez Vera, Carmen Sotomayor, Percy Bedwell, Ana María Domínguez, Elvira Jéldrez

Análisis de la calidad del léxico y su relación con la calidad de la escritura en niños chilenos de cuarto año de educación primaria

Gabriela Gómez Vera¹, Carmen Sotomayor¹, Percy Bedwell², Ana María Domínguez², Elvira Jéldrez¹

INTRODUCCIÓN

¿Por qué el léxico?

La escritura es una competencia básica para el desarrollo del aprendizaje, la comunicación y la integración en la sociedad. A pesar de ello, el sistema nacional de evaluación chileno ha focalizado sus programas exclusivamente en la lectura. Solo en el año 2008 se desarrolló un primer estudio piloto para medir la calidad de la escritura de los escolares (4to año básico). Los resultados fueron preocupantes: 38% de los evaluados daba cuenta de un dominio inicial de la escritura (Unidad de Currículum y Evaluación, 2009). A pesar de este diagnóstico, pocas investigaciones han abordado las posibles causas de estos resultados. En este contexto, el estudio de la calidad del léxico es un aspecto fundamental, dada la etapa de desarrollo en que se encuentran los niños evaluados. Tal como subrayan Smith-Lock, Nickels & Mortensen (2009), la escritura supone la aplicación de múltiples habilidades en forma simultánea. Entre otras, concurren en la escritura la capacidad de usar el vocabulario, la ortografía y la organización discursiva; cada tarea supone un desafío para el procesamiento cognitivo y tiene ciertas dificultades que le son específicas. Es por ello que resulta pertinente un estudio en profundidad del léxico, ya que es un componente que se encuentra en pleno proceso de aprendizaje por parte de los niños de cuarto año básico. En esta etapa escolar, ya se ha cerrado la etapa del aprendizaje guiado del código (que en Chile se desarrolla en los dos primeros años de la educación primaria) y ahora se comienza a escribir y leer como forma de comunicación y aprendizaje (Chall, 1996). En relación con la lectura, investigaciones anteriores han demostrado que el dominio del vocabulario es un potente predictor de la comprensión (Aarnoutse et al., 2005; Verhoeven et al., 2011) que repercute particularmente en esta etapa (Senechal et al., 2006), ya que la tarea no se centra más en la simple decodificación, sino que en la comprensión, ante lo cual un escaso vocabulario es un impedimento directo.

¹ Centro de Investigación Avanzada en Educación

² Fundación Arauco

Respecto de la escritura, en la medida en que un aprendiz adquiere un mayor dominio de los procesos lingüísticos involucrados en la tarea, entre ellos el vocabulario, disminuye la carga de la memoria de trabajo durante la producción del texto. Esto implica un aumento de la disponibilidad de recursos que pueden destinarse a la memoria de largo plazo (McCutchen, 2011). En la literatura especializada actual, la calidad, también llamada "riqueza" léxica (Ávila, 1991), se define como un objeto multidimensional, que no puede ser evaluado mediante un indicador único, sino por un conjunto diverso de indicadores, distintos aunque interrelacionados (Malvern et al., 2004; Read, 2000). Concretamente, los dos grandes conceptos implicados en la evaluación de la calidad del léxico se refieren a la cantidad del léxico usado en la producción de un texto y a las características propias de este (McNamara et al., 2010; Perfetti and Hart, 2001).

Respecto de la cantidad de vocabulario, una mayor diversidad léxica, se relaciona con el número de palabras que domina el productor del texto (McNamara et al., 2010). Tal como subrayan estos autores, desde el punto de vista del vocabulario, un texto de mayor calidad podría entenderse como un texto más fácil de comprender para el lector: con más palabras familiares y con una estructura más simple. Sin embargo, desde el punto de vista del escritor, lo propio de un texto de mayor calidad es el uso de un vocabulario más diverso y el uso de una sintaxis y un vocabulario más complejos. Así es como, respecto de las características, la literatura destaca los conceptos de sofisticación léxica y densidad léxica. La primera se refiere al uso de vocabulario poco frecuente y de mayor complejidad y al uso de palabras más largas y con estructuras silábicas más elaboradas. Este tipo de palabras permitiría a los escritores "to express their meanings in a precise and sophisticated manner" (Read, 2000, p. 200). Respecto de la densidad léxica, esta se refiere a la proporción de unidades léxicas con contenido presentes en un texto. El vocabulario con contenido es aquel que tiene relevancia semántica, tanto aisladamente como en el contexto de una oración y este se identifica habitualmente con sustantivos, adjetivos, verbos (no auxiliares) y adverbios. Las otras partes de la oración (artículos, pronombres, conjunciones, auxiliares, etc.) se catalogan como palabras funcionales, puesto que cumplen un rol más vinculado con la gramática y con la producción formal del texto (Read, 2000). Respecto de las propiedades tanto cuantitativas como cualitativas del léxico, hay estudios recientes que vinculan positivamente los indicadores que las representan con la calidad de la escritura (McNamara et al., 2010; Olinghouse and Leaird, 2009; Olinghouse and Wilson, 2013).

1. Objetivos del estudio

El objetivo específico de este estudio es medir, utilizando un conjunto amplio de indicadores, la calidad del léxico producido por escolares chilenos de cuarto año de educación básica. Tal como señalan Porres (2005) y Ávila (1991), a pesar de que es un componente esencial, hay pocos estudios que se preocupen de este en el marco del desarrollo de la competencia escrita en español. Tampoco hay antecedentes precisos para guiar su enseñanza escolar, menos aún, en el marco del sistema educativo chileno. Un

análisis descriptivo de las características del vocabulario utilizado por los escolares constituye un punto de partida fundamental para mejorar su enseñanza. En este contexto, un aporte importante del presente estudio es la distinción entre géneros, puesto que, como se demostrará en seguida, hay aspectos del desarrollo del vocabulario en que hay diferencias de un género a otro, mientras que, otros aspectos son independientes del texto producido. Tal como Olinghouse & Wilson (2013) mostraron en su estudio de producción escrita en inglés, es esperable que existan interacciones significativas entre el tipo de texto y el vocabulario seleccionado. Estos autores constataron diferencias respecto de la variedad (mayor en el texto narrativo, menor en el informativo) y la densidad (mayor en los textos informativos). En relación con esto mismo, los autores afirmaron que los predictores léxicos de la calidad de la escritura varían según el género textual, lo que se traducirá en una interacción distinta de los indicadores respecto de una medida referencial de la calidad.

2. METODOLOGÍA

Muestra y procesamiento de la información

Este estudio analiza las propiedades del léxico utilizado por niños de cuarto año de educación básica en la producción de textos en el contexto de una evaluación externa desarrollada por el Ministerio de Educación a través del Sistema de Medición de la Calidad de la Educación: SIMCE (Unidad de Currículum y Evaluación, 2009). Esta prueba fue aplicada a una muestra representativa de estudiantes de cuarto año básico a nivel nacional. En total, fueron evaluados 22,854 estudiantes distribuidos en 706 establecimientos, lo que equivale al 9% de la matrícula de ese año escolar. De este total, se seleccionaron 687 niños considerando un error muestral de 3,5% a nivel total y conservando la distribución original en la población de acuerdo al género de los niños y niñas, la ubicación geográfica de las escuelas (norte, centro, sur y Región Metropolitana), la dependencia administrativa de estas (municipal, particular subvencionado o particular pagado) y su nivel socioeconómico (bajo-medio bajo, medio o medio alto-alto). Cada niño o niña redactó tres textos (una narración, una carta de solicitud y una noticia), por lo que el corpus de análisis consta de 2056 escritos, uno de cada tipo, salvo que no se encontrase el texto en la muestra SIMCE. Para transcribir los escritos, una lingüista digitalizó los textos de los niños en base a las siguientes instrucciones:

- Transcribir corrigiendo errores ortográficos de carácter literal y acentual.
- No transcribir puntuación.
- Corregir hipo e hipersegmentaciones.
- Calificar con NN cuando un texto no pueda ser transcrito por ilegible.
- Marcar palabras ilegibles con NN.

- Transcribir todos los elementos escritos en la hoja de respuesta, incluidos los elementos externos del texto (por ejemplo cuando se escribe "firma" al pie de la carta)

Ante casos particulares, se realizaron nuevas correcciones (ver detalle en anexo A), teniendo siempre como objetivo generar una base de datos de nivel palabra, totalmente despejada de errores vinculados a la ortografía, al débil dominio de la estructura oracional o a los problemas de caligrafía propios de esta etapa del aprendizaje. La base resultante llegó a 134.168 palabras compuestas por 568.046 caracteres.

Indicadores

Para representar las propiedades cuantitativas y cualitativas del léxico, se implementaron un conjunto de indicadores validados en la literatura especializada sobre el tema:

Diversidad léxica. También llamado Variación léxica, este factor se refiere a qué tan variado o diverso es el vocabulario utilizado en un texto. La medida clásica de la diversidad léxica es dividir el número de palabras variadas por el número total de palabras presentes en el texto, lo que se conoce como Type/Token Ratio (McCarthy and Jarvis, 2010). Sin embargo, es sabido que esta medida se ve afectada por la extensión del texto. En la medida que hay más palabras, estas tenderán a ser menos variadas (Malvern et al., 2004). Para controlar este factor, se calculó el índice CTTR, Corrected Type/Token Ratio. Esta es una corrección de la medida original que, en vez de dividir las palabras variadas por el total de palabras, lo divide por la raíz cuadrada de este total, asumiendo que el impacto de la extensión del texto es proporcional a esta (Malven et al., 2004, p. 26). El autor recién citado afirma que esta medida corregida sigue siendo sensible a la extensión del texto (Olinghouse and Leaird, 2009; Olinghouse and Wilson, 2013), sin embargo, el impacto de la extensión solo sería significativo para textos mayores a varios cientos de palabras (Malven et al., 2004, p. 29), lo que no es el caso de los textos que componen la muestra de nuestro estudio (ver tabla 1)³.

Sofisticación Léxica. Como se mencionó previamente, este concepto hace alusión al uso de un vocabulario menos frecuente y más complejo que permite al escritor expresarse con mayor precisión (Read, 2000). Para cuantificar la sofisticación en los textos de la muestra, se utilizaron como indicadores la longitud de las palabras, medida por el número de caracteres de las palabras usadas por los niños en cada texto y la proporción de palabras polisílabas. A partir del conteo de caracteres, se definieron las palabras polisilábicas como aquellas con más de 6 caracteres, lo que es igual a la suma de la media y la desviación estándar de la longitud de las palabras en la muestra. En seguida se calculó el porcentaje de polisílabas presentes en cada texto, que se empleó

³ Existen índices más recientes que no estarían afectados por la extensión del texto, como el MTLD: *Measure of textual lexical diversity* (McCarthy and Jarvis, 2010; Olinghouse and Wilson, 2013) o el índice D (Malven et al., 2004), sin embargo son propuestas aún en desarrollo y no probadas con textos en español. Además, según McNamara et al. (2010) el MTLD requiere textos de al menos 100 palabras, lo que es más alto de la extensión promedio de los escritos aquí estudiados.

como indicador en los análisis. Se tomó esta medida como referencia, dado que la literatura confirma que el uso de palabras más largas sirve como representación de la sofisticación léxica (Crossley 2011).

Densidad léxica. Para caracterizar la densidad, fue necesario clasificar cada ítem léxico en palabra de contenido o funcional, para lo cual, previamente se identificaron las partes de la oración a la que cada una correspondía: sustantivo, adjetivo, verbo y adverbio (vocablos con contenido) o pronombre, artículo, conjunción y preposición (vocablos funcionales). Esta clasificación se llevó a cabo por medio del software TreeTagger (Schmid, 1994). El programa permitió separar del conjunto de verbos, los que son auxiliares y que se consideran vocabulario funcional e identificar los lemas a los cuales se asociaba cada palabra. El lema es la palabra guía o referencia de diccionario (headword) que en el texto se puede manifestar por alguna de sus inflexiones (López-Mezquita Molina, 2005). Por ejemplo, el verbo ser, en infinitivo, es un lema, mientras que todas sus conjugaciones son las variaciones por las cuales el lema se puede representar en el texto. En relación con conocimiento de vocabulario por parte del productor del texto, la referencia fundamental son los lemas, más que las palabras efectivamente escritas (Read, 2000). Utilizando ambos criterios: lemas y categorías léxicas, fue posible diferenciar la frecuencia, ya no de palabras totales, sino de lemas, de forma semejante al trabajo de Porras, 2005). Así es como se reportan los sustantivos, verbos y adjetivos más utilizados por los escolares para cada tipo de texto. Además, utilizando la metodología de cálculo del índice CTTR, se concibió una medida análoga pero que solo considera los lemas con contenido (LC): LC-CTTR, lo que constituye una estimación de la densidad léxica de cada texto.

3. RESULTADOS

Caracterización de la muestra

Diversidad léxica.

Tal como se indica en la Tabla nº 1, en total, los niños escribieron textos de una extensión de 65 palabras en promedio; de estas palabras, aproximadamente 43 fueron palabras variadas, mientras que el resto fueron repeticiones de las mismas. Las medias por tipo de texto dan cuenta de una marcada diferencia entre los textos narrativos y el resto de los escritos. Concretamente, la extensión de los textos narrativos duplica la extensión de las cartas y noticias, respecto del total de palabras.

Tabla nº 1 Palabras totales, variadas e Índice CTTR, media y desviación estándar por tipo de texto y en total

Textos	N estudiantes	Media p. totales	D.E. p. totales	Media p. variadas	D.E. p. variadas	% de palabras variadas respecto de la media total
Cuento	681	99,965	33.832	58,253	16,277	60,17%
Carta	686	49,515	22.278	37,210	13,289	78,56%
Noticia	686	46,796	23.368	33,456	13,580	75,38%
total	2053	65,338	36,382	42,936	18,088	71,4%

Respecto de las palabras variadas, a pesar de que su frecuencia es más alta en los cuentos que en las cartas y noticias, el porcentaje que representan respecto del total de la muestra es menor. Ello da cuenta del efecto de la cantidad de palabras escritas sobre la variedad. Esto es, los textos narrativos, mucho más extensos, dan lugar a que se reduzca la variedad a medida que aumenta la cantidad de palabras escritas. El índice CTTR permite comparar la diversidad entre los tres tipos de texto, controlando por el efecto de la extensión de los escritos, lo cual da como resultado una apreciación más precisa de la variedad.

Tabla nº 2 Índices de la calidad del léxico y diferencias significativas entre tipos de textos (Repeated Measures Anova)

Índices	Medias (error estándar)			Diferencias*
	Cuento	Carta	Noticia	
Diversidad (CTTR) (Variedad)	4.11 (0.02)	3.72 (0.02)	3.44 (0.02)	Cuentos > Cartas > Noticias
Sofisticación (Polisílabas)	13.62 (0.19)	21.93 (0.26)	26.11 (0.27)	Noticias > Cartas > Cuentos
Densidad (LC-CTTR) (Lemas con contenido)	3.20 (0.03)	2.44 (0.02)	2.36 (0.02)	Cuentos > Cartas > Noticias

*Test de comparaciones múltiples con corrección Šidàk, $\alpha=0.001$. N=680

La Tabla 2 presenta una comparación de las medias por género del índice CTTR y los otros dos índices derivados. La equivalencia de estas medias se testeó por medio de un análisis de la varianza para medidas repetidas (tomando en consideración que se trata de textos producidos por los mismos sujetos). Sobre la base de estas estimaciones se aplicó el test de comparaciones múltiples con corrección Šidàk (Abdi, 2007) con un alfa de 0.001. Respecto de la diversidad, los resultados presentados en la Tabla 2 muestran que hay diferencias significativas entre las medias según tipo de texto. Los cuentos son los escritos que presentan mayor diversidad, seguidos por las cartas y, en último lugar, las noticias.

Sofisticación Léxica.

Como señalamos previamente, la sofisticación se manifiesta en el uso de palabras más complejas, tales como las polisílabas. Los datos nos muestran que las palabras usadas por los niños tienden a ser cortas, tanto en el total de textos, como según los tipos específicos. Las palabras fueron medidas en función de la cantidad de caracteres que las componen, lo que arrojó una media de cuatro caracteres por palabra. La Tabla 3 muestra que este promedio solo es más alto en el caso de la noticia, sin embargo, es altamente probable que ello se deba al estímulo utilizado. En este, se pedía a los niños redactar una noticia sobre el descubrimiento de huesos de dinosaurio: "Ayer se enteraron de que un investigador descubrió restos de dinosaurio cerca de tu escuela. Escribe una noticia sobre este descubrimiento". La gran mayoría de los niños incorporó las palabras escuela, investigador y dinosaurio a su escrito, algunos, yendo más allá, incluyeron nombres de profesiones relacionadas como arqueólogo o científico y nombres de dinosaurios, tales como alosaurio, espinosaurio o terodáctilo. El caso contrario a este es el texto narrativo, en que el estímulo es simplemente una fotografía de un niño y un delfín en el agua, acompañada de la instrucción "Escribe un cuento a partir de esta foto". A pesar de ellos, la gran mayoría de los niños utilizó la palabra delfín o alguna de sus variaciones, así como la palabra niño. El estímulo del segundo texto (carta de petición) también puede haber aumentado la tasa de palabras de mayor extensión, ya que se pide a los escolares que escriban "una carta al director de tu escuela para convencerlo de que les dé permiso para ir de paseo" (encabezado por: Tu curso juntó plata para ir de paseo. Te eligieron para convencer al director de la escuela de que les dé permiso para ir). En consecuencia, gran parte de los niños y niñas incorporó la palabra "director" en su escrito. A pesar de esta posible incidencia, la proporción de palabras polisílabas (de más de seis caracteres) es menor, como lo presenta la Tabla nº 3, el promedio fue de 12 polisílabas por texto.

Tabla nº 3 Longitud media de caracteres y promedio de palabras polisílabas por tipo de texto y en total

Textos	Media de caracteres por palabra	D.E. de caracteres por palabra	Media palabras polisílabas	D.E. de palabras polisílabas
Cuento	3,892	2,222	13,543	6,330
Carta	4,366	2,431	10,671	5,164
Noticia	4,824	3,010	11,845	5,787
Total	4,235	2,513	12,016	5,895

El índice de sofisticación, presentado por género en la Tabla 2, confirma que hay diferencias significativas en el uso de polisílabas de acuerdo al tipo de texto. Tal como se presumió, hay más sofisticación en las noticias y menos en los cuentos. No es posible descartar que este resultado esté afectado por los estímulos presentados en la evaluación.

Densidad léxica.

En directa relación con el tema anterior, la observación del tipo de palabras utilizadas por los escolares también apunta en la dirección de un dominio inicial, poco especializado, del vocabulario. Como se mencionó previamente, las palabras fueron clasificadas de acuerdo a la parte de la oración que cada una de ellas presenta y, a partir de esta clasificación, se distinguió entre vocabulario con contenido y vocabulario funcional. En todos los casos, cerca de la mitad de las palabras corresponden a vocabulario con contenido, mientras que la otra parte es vocabulario funcional.

Tabla nº4 Partes de la oración, porcentajes promedio y diferencias significativas entre tipos de textos

Partes de la oración	Media del porcentaje por texto (error estándar)			Diferencias*
	Cuento	Carta	Noticia	
Sustantivos	20,61 % (0,19)	25,45 % (0,28)	27,21 % (0,30)	Noticias > Cartas > Cuentos
Verbos	22,27 % (0,15)	22,78 % (0,22)	19,49 % (0,23)	(Cuentos = Cartas) > Noticias
Adjetivos	6,14 % (0,12)	5,15 % (0,15)	6,65 % (0,17)	(Cuentos = Noticias) > Cartas
Adverbios	4,52 % (0,08)	3,13 % (0,11)	4,96 % (0,15)	(Cuentos = Noticias) > Cartas
Pronombres	9,56 % (0,14)	12,02 % (0,18)	5,95 % (0,17)	Cartas > Cuentos > Noticias
Artículos	14,38 % (0,15)	6,27 % (0,16)	12,45 % (0,21)	Cuentos > Noticias > Cartas
Conjunciones	11,13 % (0,14)	11,15 % (0,17)	7,61 % (0,19)	(Cuentos = Cartas) > Noticias
Preposiciones	8,33 % (0,11)	10,38 % (0,17)	13,03 % (0,19)	Noticias > Cartas > Cuentos
Otros	3,05 % (0,08)	3,64 % (0,12)	2,62 % (0,11)	Cartas > (Cuentos > Noticias ^a)

*Test de comparaciones múltiples con corrección Šidàk, $\alpha=0.001$

^a Significant at 0,003

Para ahondar sobre los tipos de palabras y cómo estas se distribuyen según tipo de texto, se utilizó la misma metodología de análisis de medidas repetidas para testear diferencias significativas para cada categoría. La Tabla nº 4 presenta estas comparaciones, allí se observa que las dos partes de la oración más presentes son los sustantivos y los verbos. Más allá de lo general, se observan diferencias significativas respecto de los porcentajes representados por cada parte de la oración, que dependen del tipo de texto y de la categoría sintáctica. Respecto del vocabulario con contenido con mayor presencia en cada género, en los cuentos el primer lugar lo ocupan los verbos, en segundo lugar los sustantivos y en tercer lugar los adjetivos. Esto es coherente con el sentido de la

narración, en la medida en que las acciones, que se representan textualmente por medio de verbos, ocupan el lugar principal en estos escritos, los sujetos de estas acciones, que se representan por los sustantivos, también ocupan un lugar importante. En el caso de las cartas de petición, la parte de la oración más utilizada son los sustantivos, seguidos por los verbos y, en tercer lugar, los pronombres. Estos últimos, son más usados en las cartas de petición que en los otros géneros. Esta proporción resulta también coherente con las propiedades del género, ya que se trata de textos apelativos, que interactúan directamente con el receptor, invitándolo a responder por medio de una acción. Finalmente, las noticias, tienen una mayor proporción de sustantivos, seguida por los verbos y, en tercer lugar, por las preposiciones. Esta distribución también puede verse como pertinente en función del propósito de la noticia (exponer un hecho) que el estímulo solicitaba desarrollar.

Dado que sustantivos y verbos, son los dos tipos de palabras más utilizados, resulta pertinente profundizar más en su descripción. Como se afirmó previamente, se constató en los escritos de los niños una fuerte influencia de los estímulos respecto del vocabulario utilizado. Ello es una respuesta esperable y positiva respecto de la tarea encomendada, ya que gran parte de los niños se ajusta al tema y al género inducido en el estímulo. Por otra parte, la influencia de los estímulos también incide en que las palabras más frecuentes no fueron utilizadas espontáneamente por los escolares, sino como respuesta directa al estímulo. Los Gráficos 1 y 2 presentan los sustantivos y verbos más utilizados, descontando aquellos referidos en los estímulos (en Anexo 2 se presenta una lista de estos y su frecuencia). Cabe señalar que los representados en los gráficos son los lemas, no directamente las palabras, de modo que, la palabra "profesor", por ejemplo, incluye en su frecuencia la suma de todas sus derivaciones: profesor (38), profesora (90), profesoras (6), profesores (23) y profe (9).

Respecto de los sustantivos, aun habiendo apartado las palabras que directamente provenían de los estímulos, se encuentra una influencia del contexto y los temas a los que estos se referían. Por las fechas de las cartas de los niños, es posible saber que esta prueba fue desarrollada en un día jueves del mes de noviembre, razón por la cual ambas palabras se encuentran entre las más comunes. Por otro lado, palabras relativas al mar, la playa y ciertos animales marinos también son muy comunes y son mayoritariamente mencionadas en los textos narrativos. Otros dos ejes temáticos los constituyen la familia y la escuela. Respecto del primero, se encuentran conceptos como mamá, papá, padre, hijo, tío y familia. Respecto del segundo, se encuentran conceptos como colegio, alumno, profesor y compañero.

En la frecuencia de verbos se incluyeron los copulativos y auxiliares, lo que permite observar que los verbos ser, haber y estar, son los más utilizados por los escolares. Llama la atención el frecuente uso del verbo encontrar, que no estaba en los estímulos, pero que es cercano al concepto de descubrir, que sí lo estaba. Es posible que los niños recurriesen al primer verbo como una alternativa al segundo, revelando una estrategia de correferencialidad.

Gráfico 1: Los 50 sustantivos más utilizados.

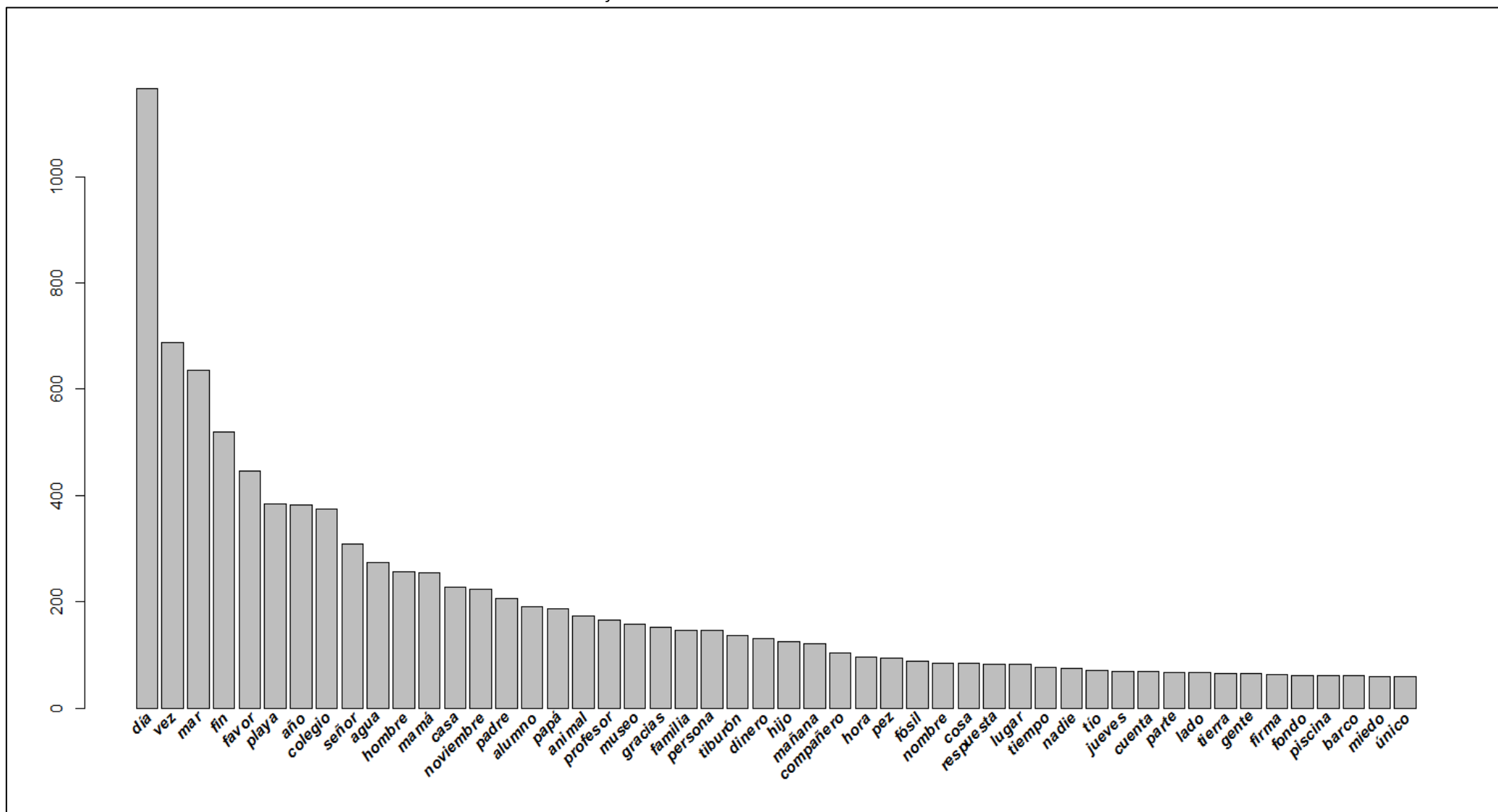
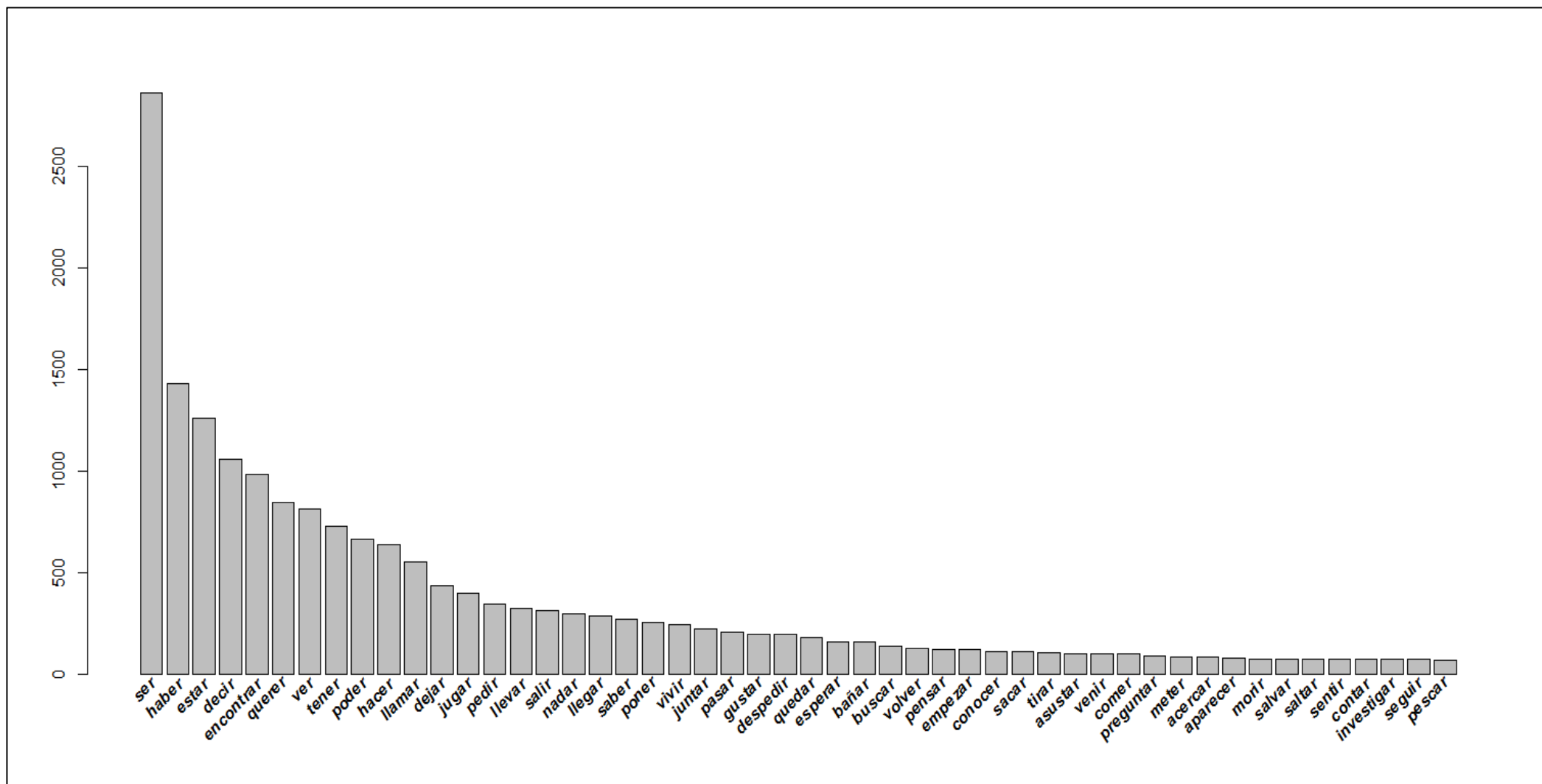


Gráfico 2: Los 50 verbos más utilizados.



"Análisis de la calidad del léxico y su relación con la calidad de la escritura en niños chilenos de cuarto año de educación primaria". Gabriela Gómez Vera, Carmen Sotomayor, Percy Bedwell, Ana María Domínguez, Elvira Jéldrez

La distribución de las frecuencias es semejante en sustantivos y verbos, y refleja la tendencia general respecto de la selección del vocabulario: hay un número reducido de palabras que son frecuentemente utilizadas en los textos y un gran número que solo son usadas muy pocas veces, o solo una vez. Más aún, incluso aquellos lemas que se encuentran al final de la lista de los más utilizados tienen ya frecuencias de aparición bastante bajas (sustantivos miedo y barco utilizados en 60 ocasiones, verbo seguir=73, pescar=68).

Un último aspecto de la densidad léxica, es el índice LC-CTTR, cuyos valores se presentaron en la Tabla nº2. Allí, el índice de densidad léxica da cuenta de diferencias significativas según el tipo de texto, hay una mayor densidad en los cuentos, seguidos por las cartas y finalmente las noticias.

Análisis interpretativo

Análisis de Correlaciones.

Es posible analizar cómo estos indicadores de la calidad del léxico se relacionan entre sí utilizando la técnica de la correlación, tomando como referencia los índices desarrollados.

Tabla nº 6 Correlación entre indicadores de la calidad del léxico según tipo de texto

	Cuento		Carta		Noticia	
	Densidad	Sofisticación	Densidad	Sofisticación	Densidad	Sofisticación
Diversidad	0,75*** 681	0,21*** 681	0,75*** 686	0,08* 686	0,77*** 686	-0,15*** 686
Densidad	---	0,11** 681	---	0,07 686	---	0,00 686

* $p < .05$; ** $p < .01$; *** $p < .001$

La Tabla nº6 presenta los resultados de los tres indicadores construidos para cada tipo de texto. En los tres casos, la correlación más importante se da entre diversidad y densidad, la correlación es alta y constante para los tres tipos de texto. Por otra parte, la sofisticación se relaciona de manera diferente con los otros dos indicadores y depende del tipo de texto. La relación entre sofisticación y diversidad es significativa y positiva en el caso de los cuentos, mientras que es negativa en el caso de las noticias. Ello supone que, en el marco de este tipo de texto, el uso de un vocabulario más variado se corresponde con un menor uso de polisílabas. Finalmente, la sofisticación solo tiene una correlación significativa con la densidad (positiva) en el caso de los cuentos. En el caso de cartas y noticias, una mayor o menor presencia de polisílabas no se corresponde con el uso de un vocabulario de mayor o menor densidad. El principal resultado de este análisis es que permite constatar que las relaciones entre los factores varían de un tipo de texto a otro. Es por ello que resulta pertinente desarrollar un análisis que permita cuantificar el rol de cada uno de estos determinantes considerándolos en su conjunto.

Relación entre los estímulos y la imagen de género.

El género tiene otra forma de incidir en los textos producidos por los escolares que es necesario describir. Si bien, dada la estrategia de análisis, este estudio se ha concentrado en aspectos formales del vocabulario, hay un elemento semántico que es clave y que es necesario mencionar antes de cerrar esta revisión de resultados. En las secciones anteriores describimos los estímulos en base a los cuales los escolares desarrollaron sus textos. En todos ellos las referencias a sujetos y personajes son masculinas: en la foto que sirve de estímulo al cuento es un niño o joven quien acompaña al delfín; en la carta de solicitud es el "director" del colegio quien es sujeto de esta petición; finalmente, es un "investigador" quien descubre los restos de dinosaurio cerca de la escuela. Si bien podría argumentarse que los estímulos son solo puntos de partida para la escritura de los niños, una revisión del uso del género en los textos de los niños, da cuenta de cuán limitante resultó la falta de imágenes femeninas (ver detalle en Anexo D). La palabra niño fue utilizada 1441 veces, mientras que la palabra niña solo figura 32 veces en los textos (aun cuando en la muestra un 49,5% son escolares de género femenino). La palabra investigador fue utilizada 266 veces en comparación con solo 2 usos de la palabra investigadora. Este último caso resulta paradigmático debido a la asociación del género masculino con las profesiones científicas, lo cual es predominante en la sociedad chilena. La palabra director fue utilizada 842 veces, mientras que la palabra directora solo fue utilizada 56 veces. Ello probablemente sea incluso contradictorio con la realidad del sistema educativo chileno en que la tasa de directivos escolares de género femenino es más alta. Prueba de ello es el contraste que se da con el uso del lema profesor que, como ya se mencionó fue más utilizado en sus formas femeninas: profesora (90) y profesoras (6), que en sus formas masculinas: profesor (38) y profesores (23). Ello demuestra que, cuando no hay restricción impuesta por los estímulos, el uso del género puede cambiar, lo cual justifica la necesidad de tomar en consideración este aspecto en el diseño de instrumentos de evaluación de la escritura.

4. DISCUSION

En base a la magnitud de este estudio y a la amplitud de aspectos del léxico que se han abordado, los resultados permiten describir algunos aspectos de las competencias escritas de los escolares chilenos. Un primer hallazgo importante es la escasez de vocabulario común entre los escolares y la simpleza de este. Como se pudo constatar los verbos y sustantivos de uso más frecuente son poco numerosos y se reducen a los más simples. Si se descuenta el léxico inducido por los estímulos, las palabras que más frecuentemente utilizan los niños están muy relacionadas con su experiencia cotidiana, son simples y breves. No se dispone de referencias que nos permitan comparar este hallazgo con el vocabulario de niños en otras etapas de la adquisición de léxico en el caso del español de Chile, por lo cual no es posible afirmar que se trata de una condición permanente o si es, simplemente, un momento muy temprano del desarrollo léxico que podrá ser compensado en un desarrollo posterior. De todos modos, el hallazgo en sí mismo, supone un resultado relevante desde el punto de vista

"Análisis de la calidad del léxico y su relación con la calidad de la escritura en niños chilenos de cuarto año de educación primaria". Gabriela Gómez Vera, Carmen Sotomayor, Percy Bedwell, Ana María Domínguez, Elvira Jéldrez

pedagógico, ya que implica la necesidad de tomar en cuenta este factor en el diseño de programas de enseñanza de la escritura, tanto a nivel de currículo, como en el trabajo cotidiano de los profesores. El desarrollo de un vocabulario más variado y complejo supone que ello sea trabajado explícitamente durante el aprendizaje. Los modelos estadísticos, confirman la relevancia de esta tarea, puesto que la variedad, el uso de vocabulario con contenido y la sofisticación del léxico son factores determinantes de la calidad de la escritura.

Un resultado importante ha sido el constatar que distintos indicadores de la calidad del léxico tienen un rol determinante respecto de la escritura. Adicionalmente, se ha demostrado que la influencia de los indicadores de la calidad del léxico en relación con la escritura varía de un tipo de texto a otro. Estos resultados, concuerdan con los obtenidos por Olinghouse & Wilson (2013) quienes también constataron que la capacidad predictiva de los aspectos del léxico respecto de la escritura varía por género textual. Estos autores concluyeron, además, que la diversidad era el principal determinante léxico de la calidad de la escritura. En nuestro caso, ese rol lo comparten los indicadores de diversidad y densidad y varía según el tipo textual. Por otra parte, Olinghouse & Wilson (2013), encontraron que la diversidad era semejante entre textos narrativos y argumentativos, lo que en este estudio no se constató.

Un tercer resultado a destacar es el rol que los estímulos juegan en el desarrollo de los textos. El vocabulario presente en los estímulos es altamente determinante del vocabulario utilizado por los escolares. Desde el punto de vista del aprendizaje de habilidades más generales, como la creatividad o el desarrollo de las ideas, este apego al estímulo puede entenderse como una limitación. Sin embargo, desde el punto de vista del desarrollo léxico, es interesante destacar la oportunidad pedagógica que esto representa. Estímulos que incentiven el uso de un vocabulario más diverso y sofisticado pueden ser un adecuado medio para mejorar la calidad del léxico, aumentando el vocabulario que los niños tienen disponible y familiarizándolos con el uso de un vocabulario más amplio y complejo.

Finalmente, en relación con los factores contextuales, cabe destacar que su relevancia es menor si se la compara con los factores vinculados al léxico. La escasa incidencia en los resultados del factor género (masculino o femenino) solo relevante respecto de textos narrativos, marca una diferencia con la literatura precedente. A partir de la comparación con la lectura que sí da cuenta de mayores diferencias (Ma, 2008; Logan and Johnston, 2010), se suele dar por sentado que estas diferencias son genéricas respecto del dominio lingüístico. En el área de lectura, también sería necesario estudiar más de cerca el fenómeno, indagando el rol del género respecto de tipos de texto específicos. Esto permitiría una comprensión en mayor profundidad de este fenómeno y el desarrollo de propuestas pedagógicas mejor fundamentadas respecto de las necesidades de estudiantes de distinto género. Aun reconociendo la importancia del nivel socioeconómico y del género, queda mucho espacio para el desarrollo de la competencia escrita en los escolares chilenos.

5. BIBLIOGRAFÍA

- Aarnoutse et al., 2005Aarnoutse, C., van Leeuwe, J., and Verhoeven, L. (2005). Early literacy from a longitudinal perspective. *Educational Research and Evaluation*, 11(3):253–275.
- Abdi, H. (2007). The Bonferonni and Šidák corrections for multiple comparisons. In Salkind, N., editor, *Encyclopedia of Measurement and Statistics*, pages 1–9. Sage, Thousand Oaks (CA).
- Ávila, 1991Ávila, R. (1991). Densidad léxica y adquisición del vocabulario: niños y adultos. In *El español de América: actas del III congreso internacional del español en América: Valladolid, 3 a 9 de julio de 1989*, pages 621–630. Consejería de Cultura y Turismo.
- Bressoux, P. (2008). *Modélisation statistique appliquée aux sciences sociales*. De Boeck, Bruxelles.
- Carroll, 1964Carroll, J. B. (1964). *Language and thought*. Prentice Hall, Englewood Cliffs, N.J.
- Chall, 1996Chall, J. S. (1996). *Stages of reading development*. Harcourt Brace College Publishers, Fort Worth, TX.
- Contreras, D., Sepúlveda, P., and Bustos, S. (2010). When schools are the ones that choose: The effects of screening in Chile. *Social Science Quarterly*, 91(5):1349–1368.
- Crossley, S. A., Weston, J. L., McLain Sullivan, S. T., and McNamara, D. S. (2011). The development of writing proficiency as a function of grade level: A linguistic analysis. *Written Communication*, 28(3):282–311.
- Foster, J. J., Barkus, E., and Yavorsky, C. (2006). *Understanding and Using Advanced Statistics: A practical guide for students*. Sage, London.
- AUTHORS1Gómez et al., 2012AUTHORAUT (2012).
- Graham, S., Gillespie, A., and McKeown, D. (2013). Writing: importance, development, and instruction. *Reading and Writing*, 26(1):1–15.
- Logan and Johnston, 2010Logan, S. and Johnston, R. (2010). Investigating gender differences in reading. *Educational Review*, 62(2):175–187.
- López-Mezquita Molina, 2005López-Mezquita Molina, M. T. (2005). *La evaluación de la competencia léxica: tests de vocabulario. Su fiabilidad y validez*. PhD thesis,

"Análisis de la calidad del léxico y su relación con la calidad de la escritura en niños chilenos de cuarto año de educación primaria". Gabriela Gómez Vera, Carmen Sotomayor, Percy Bedwell, Ana María Domínguez, Elvira Jéldrez

Universidad de Granada, Facultad de Filosofía y Letras, Departamento de Filología Inglesa.

Ma, 2008Ma, X. (2008). Within-school gender gaps in reading, mathematics, and science literacy. *Comparative Education Review*, 52(3):437–460.

Malvern et al., 2004Malvern, D., Richards, B., Chipere, N., and Durán, P. (2004). *Lexical diversity and language development: Quantification and assessment*. Palgrave Macmillan, Basingstoke, UK.

McCarthy and Jarvis, 2010McCarthy, P. and Jarvis, S. (2010). Mtd, vocd-d, and hd-d: A validation study of sophisticated approaches to lexical diversity assessment. *Behavior Research Methods*, 42(2):381–392.

McCutchen, 2011McCutchen, D. (2011). From novice to expert: Implications of language skills and writing-relevant knowledge for memory during the development of writing skill. *Journal of Writing Research*, 3(1):51–68.

McNamara et al., 2010McNamara, D. S., Crossley, S. A., and McCarthy, P. M. (2010). Linguistic features of writing quality. *Written Communication*, 27(1):57–86.

Olinghouse and Leaird, 2009Olinghouse, N. and Leaird, J. T. (2009). The relationship between measures of vocabulary and narrative writing quality in second- and fourth-grade students. *Reading and Writing*, 22(5):545–565.

Olinghouse and Wilson, 2013Olinghouse, N. and Wilson, J. (2013). The relationship between vocabulary and writing quality in three genres. *Reading and Writing*, 26(1):45–65.

Perfetti and Hart, 2001Perfetti, C. A. and Hart, L. (2001). The lexical quality hypothesis. In Verhoeven, L., Elbro, C., and Reitsma, P., editors, *Precursors of functional literacy*, page 189–214. John Benjamins, Amsterdam, the Netherlands.

Porras, 2005Porras, M. R. (2005). Una responsabilidad escolar olvidada: el desarrollo del componente léxico. *Revista Educación*, 29(1):31–44.

Read, 2000Read, J. (2000). *Assessing vocabulary*. Cambridge University Press, Cambridge, UK.

Schmid, 1994Schmid, H. (1994). Probabilistic part-of-speech tagging using decision trees. In *Proceedings of International Conference on New Methods in Language Processing*, pages 44–49. Manchester, UK.

Senechal et al., 2006Senechal, M., Ouellette, G., and Rodney, D. (2006). The misunderstood giant: On the predictive role of early vocabulary to future reading. In Dickinson, D. K. and Neuman, S. B., editors, *Handbook of early literacy research Vol. 2.*, page 173–182. Guilford, New York, NY.

"Análisis de la calidad del léxico y su relación con la calidad de la escritura en niños chilenos de cuarto año de educación primaria". Gabriela Gómez Vera, Carmen Sotomayor, Percy Bedwell, Ana María Domínguez, Elvira Jéldrez

Smith-Lock et al., 2009 Smith-Lock, K. M., Nickels, L., and Mortensen, L. (2009). Story writing skills of adults with a history language-impairment. *Reading and Writing*, 22(6):713–734.

AUTHORS2 et al. 2015

Troia, G., Harbaugh, A., Shankland, R., Wolbers, K., and Lawrence, A. (2013). Relationships between writing motivation, writing activity, and writing performance: effects of grade, sex, and ability. *Reading and Writing*, 26(1):17–44.

Unidad de Currículum y Evaluación, 2009 Unidad de Currículum y Evaluación (2009). Informe de Resultados de Escritura SIMCE 2008. Ministerio de Educación, Santiago de Chile.

Verhoeven et al., 2011 Verhoeven, L., van Leeuwe, J., and Vermeer, A. (2011). Vocabulary growth and reading development across the elementary school years. *Scientific Studies of Reading*, 15(1):8–25.

Weisberg, S. (2005). *Applied linear regression*. John Wiley & Sons, New Jersey.

6. ANEXOS

1. Anexo A: Resoluciones del proceso de transcripción:

1. Las duplicaciones que no muestran intención comunicativa sino que parecen ser reflejo de equivocación o de distracción del niño fueron corregidas, de esta forma, se elimina una de las palabras repetidas. Las duplicaciones con intención comunicativa ("auxilio, auxilio"; "extra, extra") sí se conservaron.
2. Los nombres inventados se escribieron tal cual aparece en los textos originales. No obstante los nombres conocidos fueron corregidos en caso de presentar alguna falta de ortografía. Por ejemplo: el niño/a escribe "Cristobal" y se transcribe la forma correcta con tilde "Cristóbal".
3. Los extranjerismos se conservan y, en caso de estar mal escritos, han sido corregidos. Por ejemplo el niño/a escribe "chor" por "short" y, siendo esta una forma aceptada por la RAE, es la que se conserva. Otro caso es "Janter" que fue transcrito como "hámster".
4. Las interjecciones han sido corregidas en caso de estar escritas de forma errada. Por ejemplo "eeee" se transcribe como "eh".
5. Cuando aparecen formas numéricas (1, 2, 3, etc.) se han conservado de esa manera en las transcripciones y no se ha cambiado por "uno, dos, tres".
6. Las abreviaciones se transcriben en su forma completa. Por ejemplo "etc." se transcribe como "etcétera", "Hrs" se transcribe como "horas", "Ud." se transcribe como "usted".
7. Las siglas que son conservadas en la transcripción y señaladas con mayúscula, por ejemplo "SIMCE", "P.D." "P.M.", "PSU", etc.

"Análisis de la calidad del léxico y su relación con la calidad de la escritura en niños chilenos de cuarto año de educación primaria". Gabriela Gómez Vera, Carmen Sotomayor, Percy Bedwell, Ana María Domínguez, Elvira Jéldrez

8. Todos los signos que no corresponden a letras, excepto los tildes, han sido eliminados en la transcripción a fin de obtener un texto lo más plano posible. De esta forma se han eliminado los paréntesis, signos de interrogación y exclamación, puntos, comas, guiones, etc.
9. Los errores de sintaxis se mantuvieron en la transcripción de forma que se conservan de los textos originales debido a que el estudio es de léxico y no de estructura sintáctica, estos errores son irrelevantes en este contexto. De esta forma encontramos casos como "me le olvidaba"; "mande una la respuesta"; "salió en la televisor".

1. Anexo B: Frecuencias de lemas inducidos por los estímulos

Sustantivos	Frecuencia	Verbos	Frecuencia
Delfín	2505	ir	2188
Niño	1691	dar	703
dinosaurio	1291	descubrir	255
Paseo	962	escribir	76
Hueso	959		
director	842		
permiso	588		
Curso	519		
Resto	428		
escuela	411		
investigador	266		
Plata	188		
Carta	185		
descubrimiento	112		

2. Anexo C: Variaciones por género en palabras inducidas por los estímulos

Palabra	Frecuencia	Palabra	Frecuencia	Palabra	Frecuencia
niño	1441	investigador	266	Director	842
niños	180	investigadores	51	Directores	1
niña	32	investigadora	2	Directora	56
niñito	25	investigadoras	1	Directoras	0
niñas	11				
niñita	1				