

LOS GÉNEROS DEL CONOCIMIENTO EN EL TEXTO ESCOLAR DE LENGUAJE Y COMUNICACIÓN: UN ANÁLISIS SEMIAUTOMÁTICO DE SU LECTURABILIDAD¹

KNOWLEDGE GENRES IN THE SCHOOL TEXTBOOK OF
LANGUAGE AND COMMUNICATION: A SEMI AUTOMATIC
ANALYSIS OF ITS READABILITY

DANIEL ROJAS VILLARROEL
Pontificia Universidad Católica de Valparaíso
daniel.rojas.v02@mail.pucv.cl

ROMUALDO IBÁÑEZ ORELLANA
Pontificia Universidad Católica de Valparaíso
romualdo.ibanez@pucv.cl

FERNANDO MONCADA NAHUELQUÍN
Universidad del Bío-Bío
fmoncada@ubiobio.cl

ANDREA SANTANA COVARRUBIAS
Universidad Técnica Federico Santa María
andrea.santana@usm.cl

RESUMEN

El objetivo del presente estudio fue caracterizar, de acuerdo con indicadores de lecturabilidad, dos de los Géneros de Conocimiento (GCO) más frecuentes en el Texto Escolar (TE) de Lenguaje y Comunicación de Octavo Año Básico en Chile: Exposición de Contenido (EC) y Nota (NOT). Se consideraron indicadores pertenecientes a la dimensión de la palabra, oracional y discursiva, para analizar semiautomáticamente un corpus constituido por 60 textos pertenecientes a los CGO seleccionados. Los resultados revelaron que, en la dimensión de la palabra, los textos presentaron mayormente palabras de breve extensión y escasas nominalizaciones; mientras que, en la dimensión oracional, se identificó una elevada presencia de construcciones extensas y estructuras subordinadas; por su parte, en la dimensión discursiva, se observó una alta presencia de marcadores y

¹ El trabajo del segundo y tercer autor se enmarca respectivamente en los proyectos Fondecyt Regular 1201440 y Postdoctorado Fondecyt 3180779, ambos financiados por la Agencia Nacional de Investigación y Desarrollo.

relaciones de coherencia aditivas. Se concluyó que los textos analizados no se adecuan plenamente al nivel de comprensión lectora que, en principio, poseen los estudiantes.

Palabras clave: lecturabilidad, marcadores discursivos, relaciones de coherencia, texto escolar, géneros del conocimiento.

ABSTRACT

The objective of the present study was to characterize, according to readability indicators, two of the most frequent Knowledge Genres in the Textbook of Language and Communication of the Eighth Grade in Chile: Content Exposition and Note. Indicators belonging to word, sentence and discursive dimensions, were considered to semi automatically analyze a corpus consisting of 60 texts belonging to the selected Knowledge Genres. The results revealed that, in the word dimension, the texts presented mostly short words and few nominalizations; while, in the sentence dimension, a high presence of extensive constructions and subordinate structures were identified; on the other hand, in the discursive dimension, a high presence of markers and additive coherence relations were observed. It was concluded that the texts analyzed do not fully conform to the level of reading comprehension that, in principle, the students have.

Keywords: readability, discourse markers, coherence relations, textbook, knowledge genres.

Recibido: 05/05/2020. *Aceptado:* 30/10/2020.

1. INTRODUCCIÓN

La lecturabilidad corresponde a la complejidad o facilidad con que un texto puede ser leído y comprendido (Alliende, 2002; DuBay, 2007; McNamara et al., 2014). Se ha utilizado como una forma de clasificar textos en diversas áreas, por ejemplo, en la salud, donde interesa precisar si los documentos médicos son asequibles para los pacientes (Skierkowski et al., 2019); también en el derecho, donde resulta imprescindible determinar si las normas de distinta naturaleza son comprensibles para los ciudadanos (Long & Christensen, 2011; Dyson & Schellenberg, 2017); y, por supuesto, en la educación, ámbito en el cual es fundamental estipular criterios objetivos para que la elección del material didáctico resulte adecuada en función del nivel educativo (Dufty et al., 2006; McNamara et al, 2014; Kasule, 2011; Crossley et al., 2019).

Tradicionalmente, la lecturabilidad se ha calculado de acuerdo con fórmulas sustentadas en indicadores circunscritos de forma exclusiva a la dimensión de la palabra y la oración (DuBay, 2007). Estos indicadores se emplean bajo el supuesto de que un texto que posee palabras y oraciones más extensas es más difícil de procesar y comprender que otro conformado por palabras y oraciones más cortas

(Kleijn, 2018). El inconveniente de estos indicadores es que solo permiten calcular la lecturabilidad de un texto asociándola al nivel más superficial de la comprensión, es decir, al Código de Superficie (Dufty et al., 2006; McNamara et al., 2014).

Con el propósito de calcular la lecturabilidad de los textos de forma más precisa, se han propuesto actualmente fórmulas de medición que complementan los indicadores tradicionales con indicadores discursivos (McNamara et al., 2014; Crossley et al., 2014; Crossley et al., 2017; Kleijn, 2018). Específicamente, dado su rol central en la construcción del Modelo de Situación, las relaciones de coherencia –así como los marcadores discursivos que las explicitan– son sugeridas como indicadores que permitirían dar cuenta de la lecturabilidad de los textos, considerando niveles de representación más profundos, tales como la Base Textual y el Modelo de Situación (Hall et al., 2014; McNamara et al., 2014; Crossley et al., 2014, 2016, 2017; Kleijn, 2018). Ahora bien, y aunque en lenguas como el inglés (McNamara et al., 2014) y el holandés (Pander Maat et al., 2014) la integración de la dimensión discursiva ha resultado en fórmulas más precisas y eficientes desde hace más de una década (Dufty et al., 2006), en español la lecturabilidad se continúa calculando principalmente por medio de fórmulas que contemplan solo indicadores asociados a la dimensión de la palabra y la oración.

En el contexto anteriormente descrito y considerando el rol central que se le asigna al Texto Escolar en los procesos de enseñanza aprendizaje (Meneses et al., 2014; Ibáñez et al., 2018; Soaje de Elías, 2018) y, como consecuencia, los recursos invertidos por el Estado en la implementación de políticas públicas que permitan proveer de textos escolares a la mayor cantidad de estudiantes del país (MINE-DUC, 2018), resulta particularmente necesario saber si el nivel de dificultad de tales textos es adecuado a las habilidades de comprensión de los estudiantes. Por tal razón, la presente investigación se propone caracterizar, de acuerdo con las dimensiones de la palabra, la oración y el discurso, la lecturabilidad de los textos pertenecientes a dos de los Géneros de Conocimiento (GCO) más frecuentes en el Texto Escolar (TE) de Lenguaje y Comunicación de Octavo Año de Enseñanza Básica en Chile: Exposición de Contenidos (EC) y Nota (NOT) (Ibáñez et al., 2017). Para alcanzar el objetivo, se llevó a cabo un análisis automático con las herramientas computacionales Inflesz (Barrio-Cantalejo, 2007; Barrio-Cantalejo et al., 2008) y NooJ (Silberztein, 2016), ambos *softwares* gratuitos y libres. En la dimensión de la palabra, se evaluaron los indicadores extensión de la palabra y nominalización; en la dimensión oracional, se consideró la extensión de la oración y la subordinación; finalmente, en la dimensión discursiva, se identificaron marcadores del discurso y tipos de relaciones de coherencia. Cabe precisar que para los indicadores nominalización, subordinación, marcadores y tipo de relación de coherencia no existe en español un índice con el cual asociar un grado de complejidad –p. ej. fácil, difícil– o un lector ideal según su edad. Por tal razón, este estudio está orientado a caracterizar la lecturabilidad en términos de rasgos lingüísticos y

discursivos, y no a determinar grados de dificultad de los textos analizados.

A continuación, se revisan los conceptos principales en que se basa este estudio. Luego, se abordan los aspectos metodológicos involucrados en la investigación. Posteriormente, se revisan y discuten los resultados. Por último, se presentan las principales conclusiones.

2. GÉNEROS DE CONOCIMIENTO EN EL TEXTO ESCOLAR

El TE constituye uno de los géneros más representativos del Discurso Pedagógico (DP), el cual, según Bernstein (1990), reúne en sí la actividad pedagógica en general y define las relaciones entre los participantes sociales de dicho ámbito. Christie (2002), a partir de los planteamientos de Bernstein (1990), sostiene que el DP está compuesto por géneros que se realizan en un registro regulativo –instrucciones que guían los procesos de enseñanza-aprendizaje– e instruccional –conocimientos y habilidades que los estudiantes deben adquirir–. Estos géneros pueden asociarse y redundar en macrogéneros, los cuales, en términos de Rose (2014), se configuran por Géneros de Conocimiento (GCO) y por Géneros Curriculares (GCU). Los GCO divulgan el saber formulado en el marco de una cierta disciplina, en tanto los GCU se vinculan con la dirección de la actividad pedagógica, es decir, el establecimiento de objetivos, la formulación de instrucciones, entre otros. Así, el TE se puede conceptualizar como un macrogénero curricular, cuyo macropropósito es instruir acerca de los conocimientos declarativos y procedurales propios de una disciplina (Ibáñez et al., 2017).

Ibáñez et al. (2017), en su estudio acerca de los textos escolares chilenos de Segundo Ciclo de Enseñanza Básica, identificaron 15 CGO en un total de 11.139 instancias analizadas. Entre los CGO identificados se hallan: Nota (NOT), Espacio Reflexivo (ER), Exposición de Contenidos (EC), Definición (DEF), Noticia de Divulgación Científica (NDC), Síntesis Conceptual (SC), Recuento Histórico (RH), Biografía (BI), Fuente Histórica (FH), Mapa (MAP), Guía Procedimental (GP), Comprobación Empírica (CE), Expresiones Frecuentes (EF), Espacio Metalingüístico (EM) y, por último, Dato Estratégico (DE).

De estos géneros, EC y NOT son dos de los más frecuentes en la asignatura de Lenguaje y Comunicación. De hecho, entre ambos CGO se concentra el 50% de los 1.629 textos identificados en dicha asignatura. En cuanto a sus características, la EC es un “género pedagógico cuyo propósito comunicativo es presentar temáticas propias de una asignatura. Su organización discursiva es expositiva o narrativa y se configura semióticamente a partir del modo verbal y gráfico” (Ibáñez et al., 2017, p. 468). Por su parte, la NOT es un “género pedagógico cuyo propósito comunicativo es presentar información específica acerca de una temática. Su organización discursiva es expositiva. Se configura semióticamente a partir del modo

verbal y gráfico” (Ibáñez et al., 2017, p. 468). En la Figura 1 se puede apreciar un ejemplo del género EC; en la Figura 2, por su parte, se presenta un ejemplo correspondiente al género NOT –consignado en un rectángulo–.

Entrevistas para reflexionar

¿Qué tenemos en común con los monos?



La entrevista de opinión

La entrevista de opinión es un tipo de conversación semiestructurada que busca obtener la opinión o punto de vista del entrevistado sobre un tema de interés común. A diferencia de otros tipos de entrevista, en esta los entrevistados son generalmente personajes reconocidos en algún área del conocimiento. Así, pueden entregar una opinión autorizada sobre el tema tratado. Las entrevistas de opinión se realizan en distintos medios de comunicación masiva, tales como televisión, radio, diarios y revistas.

Figura 1. Género Exposición de Contenido (EC).

Aplico mi aprendizaje

A continuación, en un grupo de cuatro compañeros o compañeras desarrollarán un radioteatro. Para ello:

Tipos de lenguaje no verbal
Dependiendo del medio de comunicación, existen distintos tipos de lenguaje no verbal. Cuando dos o más personas conversan, este lenguaje se compone de vestimenta, distancia entre los hablantes, gestos y expresiones.
En el caso específico del radioteatro, entre los elementos no verbales aparecen los ruidos y sonidos de efectos especiales (ruido de pasos, por ejemplo).

Paso 1 Escuchen el siguiente radioteatro para tenerlo como modelo. Para ello, ingresen al sitio web <http://codigos.auladigital.cl> e introduzcan el código TL8P244. Mientras lo escuchan, fíjense en el lenguaje paraverbal y el lenguaje no verbal (efectos especiales).

Paso 2 Lean el guion que aparece en la página siguiente y subrayen en qué momentos deben incluir elementos no verbales e intencionar elementos paraverbales. Para los primeros, anoten qué efectos especiales deben hacer (pasos, puertas al abrirse, etc.). Para los segundos, cómo deben expresar cada frase (qué entonación, qué tipo de voz, etc.).

Paso 3 Asignen los roles para desarrollar el radioteatro. Es decir, quiénes serán el narrador, el hombre y los efectos especiales.

Paso 4 Ensayen y graben varias veces el radioteatro hasta que queden conformes con un resultado. Presten mucha atención al lenguaje no verbal y al paraverbal, ya que con ellos podrán generar una atmósfera determinada.

Figura 2. Género Nota (NOT).

Como se expuso anteriormente, estos géneros aglutinan prácticamente la mitad de las instancias de GCO identificadas en la asignatura de Lenguaje y Comunicación. En consecuencia, son los géneros que los estudiantes deben enfrentar con mayor frecuencia para aprender los contenidos de la asignatura, lo que releva la importancia de determinar su grado de lecturabilidad. A continuación, se describen las dimensiones de análisis y respectivos indicadores utilizados en el presente estudio para dicho propósito, además de las herramientas empleadas.

3. INDICADORES DE LECTURABILIDAD Y HERRAMIENTAS INFORMÁTICAS DE ANÁLISIS

Enseguida, se revisan los indicadores extensión de la palabra, nominalización (dimensión de la palabra); extensión de la oración, subordinación (dimensión oracional); y relaciones de coherencia y marcadores del discurso (dimensión discursiva). Los indicadores extensión de la palabra y de la oración son los que poseen un tratamiento más extenso en español (Alliende, 2002); además, aún son empleados en los análisis en otras lenguas, a pesar de la incorporación de nuevos indicadores (Dufty et al., 2006; Crossley et al., 2014; Pander Maat et al., 2014; McNamara et al., 2014). Por su parte, nominalización, subordinación y marcadores asociados a relaciones de coherencia permiten conseguir una visión más detallada de la lecturabilidad, ya que enriquecen la dimensión de la palabra y la oración (Dufty et al., 2006; Crossley et al., 2014; Pander Maat et al., 2014; McNamara et al., 2014) e integran una dimensión –discursiva– que permite vislumbrar aspectos asociados a la comprensión profunda de los textos (Kleijn, 2018).

3.1. Dimensión de la palabra

3.1.1. Extensión de la palabra

En el caso del español, las palabras compuestas por dos sílabas o menos son de fácil decodificación comparadas con aquellas que poseen tres sílabas o más (Alliende, 2002). En este sentido, *terror* –te-rror: 2 sílabas– es más sencilla que una palabra como *tremebundo* –tre-me-bun-do: 4 sílabas–. Este fenómeno se atribuye a que las palabras más breves son utilizadas frecuentemente, al contrario que las palabras más extensas (Kleijn, 2018). Por tanto, los textos compuestos por palabras de dos sílabas o menos son, en principio, más familiares y, por lo tanto, más sencillos de procesar y comprender (McNamara et al., 2014; Crossley et al., 2014; Guamán, 2014).

3.1.2. Nominalización

La nominalización es un procedimiento léxico-gramatical usual en el discurso académico y especializado, que consiste en transformar verbos o adjetivos en sustantivos (Halliday & Matthiessen, 2014; Cinto, 2009). Por ejemplo, el verbo *comprender*, por medio de sufijación, se transforma en el sustantivo *comprensión*. Este procedimiento plantea dificultades al procesamiento y la comprensión del discurso, sobre todo en el caso de la nominalización de verbos –deverbal–, ya que invisibiliza a los participantes del proceso, con lo cual se incrementa la abstracción del discurso (García et al., 2005), como se observa en los siguientes ejemplos:

- (1) Durante la Segunda Guerra Mundial, *Fritz Haber inventó* el salitre sintético.
- (2) Durante la Segunda Guerra Mundial, ocurrió la *invención* del salitre sintético.

En (1) se explicita quién inventó el salitre sintético –Fritz Haber–; por el contrario, en (2) el sujeto desaparece debido a la nominalización del verbo *inventar*. Además, las nominalizaciones deverbales introducen ambigüedad en el texto al entablar relaciones anafóricas, las que, en ciertas circunstancias, son difíciles de determinar (García et al., 2005; Pander Maat et al., 2014), como se aprecia en el siguiente ejemplo:

- (3) En el reino vegetal, la producción de semillas y frutos implica la transferencia del polen desde los estambres hasta el estigma de las flores, donde se fecundan sus óvulos. Agentes tanto bióticos como abióticos participan de la *polinización*.

Para comprender (3) es necesario que el lector establezca una relación anafórica entre el término *polinización* y la información planteada en la primera oración. Sin embargo, establecer exitosamente este tipo de relaciones resulta complejo, sobre todo para estudiantes de educación primaria y secundaria, ya que la plena adquisición de la nominalización se desarrolla en los últimos años de la adolescencia o en los primeros de la edad adulta (Cinto, 2009).

3.2. Dimensión de la Oración

3.2.1. Extensión de la oración

El éxito de los lectores en el procesamiento del nivel sintáctico requiere la integración de los diversos elementos que componen la oración. Esta integración resulta ardua en el caso de las estructuras que poseen nueve o más ítems –palabras– (San-

tos, 2015). Por ello, una construcción como: “El fundamento último de la realidad constituye un enigma para al Hombre” –n.º de palabras: 12– impone cierto nivel de dificultad para algunos lectores, incrementando su carga cognitiva. Esto se debe a que los lectores deben utilizar una mayor cantidad de recursos cognitivos adicionales para la búsqueda, recuperación e integración de los elementos en la memoria operativa en relación con el procesamiento de estructuras oracionales más breves (Kleijn, 2018).

3.2.2. Subordinación

La subordinación o hipotaxis consiste en la relación entre un elemento dependiente y su dominante (Halliday & Matthiessen, 2014), como se observa en (4):

(4) El mueble *que sirve de librero* cayó con el temblor.

La oración *sirve de librero* –introducida por el pronombre relativo *que*– es dependiente de la oración *El mueble cayó con el temblor* en la cual está incrustada. Estudios experimentales han constatado que este fenómeno lingüístico implica un incremento en la dificultad del procesamiento y de la comprensión del texto (Pander Maat et al., 2014). Además, se sostiene que las estructuras oracionales subordinadas relentizan la generación de inferencias (Holguín & Rodríguez, 2017). Para que un lector procese exitosamente una estructura subordinada, debe ser hábil, capaz de precisar la función y relación de las palabras en una oración (Campbell, 2018).

3.3. Dimensión Discursiva

3.3.1. Relaciones de coherencia

Las relaciones de coherencia son nexos de significado entre unidades cognitivas denominadas eventos (E). Cada E puede estar compuesto por dos unidades: uno o más conceptos (C) y una acción (A) –proceso o estado–. Cada E se construye a partir del contenido expresado en un segmento discursivo –p. ej. oración– (Ibáñez, et al., 2015). Por ejemplo, a partir del segmento discursivo *Juan corrió*, los lectores debiesen construir un E compuesto por un concepto (Juan) y una acción (corrió). Las relaciones de coherencia pueden clasificarse en tres categorías principales: aditivas, temporales y causales (Kleijn, 2018). Enseguida, se ejemplifica cada una de estas relaciones:

(5) Aditiva: $\underbrace{\text{Pedro (C) canta (A)}}_{E1}$ y $\underbrace{\text{María (C) baila (A)}}_{E2}$.

(6) Temporal: $\underbrace{\text{Pedro (C) tomó (A) su bolso (C)}}_{E1}$, luego $\underbrace{\text{él (C) salió (A)}}_{E2}$.

(7) Causal: $\underbrace{\text{María (C) comió (A)}}_{E1}$ porque ella $\underbrace{\text{(C) tenía (A) hambre (C)}}_{E2}$.

Las relaciones de coherencia imponen diferentes exigencias cognitivas. En principio, las relaciones aditivas son más fáciles de procesar y comprender que las relaciones temporales. Estas, a su vez, son más sencillas respecto a las relaciones causales (Kleijn, 2018). Sin embargo, la investigación llevada a cabo en las últimas décadas ha constatado que las relaciones causales suelen procesarse más rápidamente, resultan en una representación mental de mejor calidad y se recuerdan mejor que otros tipos de relaciones (Trabasso & Sperry, 1985; Van den Broek, 1990a; Goldman & Varnhagen, 1986; Graesser & Clark, 1985; Sanders & Noordman, 2000).

3.3.2. Marcadores del discurso

Cuando las relaciones de coherencia están explicitadas en la superficie textual por un marcador discursivo, como un conector –porque, pues– o frase guía –por lo anterior, por tal motivo–, tanto el procesamiento como la comprensión del texto se ven generalmente beneficiadas. La evidencia empírica ha constatado que los marcadores, en tanto instructores de procesamiento (Noordman & Vonk, 1997; Cozijn et al., 2011) contribuyen a facilitar el procesamiento, disminuyendo los tiempos de lectura y las regresiones; asimismo, suelen aumentar la calidad de la representación mental, haciéndola más coherente y permanente en la memoria (Sanders & Noordman, 2000; Degand & Sanders, 2002; Van Silfhout et al., 2014). Los marcadores también operan como un *proxy*, es decir, un aspecto superficial, observable y cuantificable del texto, que tentativamente permite establecer el tipo de relación de coherencia predominante en un texto (McNamara et al., 2014). Reconociendo que en la literatura no existe consenso absoluto sobre cómo definir y clasificar los marcadores del discurso (Taboada, 2006), en el presente estudio se asume que son una clase funcional de expresiones léxicas extraídas desde diferentes clases morfológicas y sintácticas –p. ej. conjunciones– (Das, 2014). De este modo, en los ejemplos (5), (6) y (7), *y*, *luego* y *porque* corresponden a marcadores discursivos.

3.4. Herramientas informáticas de análisis de lecturabilidad

Durante la primera década del nuevo milenio, en los Estados Unidos (McNamara et al., 2014), Europa (Pander Maat et al., 2014) y Asia (Tanaka-Ishii et al., 2010), se comprobó que una forma eficiente de abordar la lecturabilidad era por medio de herramientas automáticas, tales como Coh-Metrix (Graesser, et al., 2004; McNamara et al., 2014) y T-Scan (Pander Maat et al., 2014). Estos instrumentos fueron elaborados con base en investigaciones que complementaron los estudios psicolingüísticos de corte experimental con datos obtenidos a partir de estudios de corpus (Graesser et al., 2004).

En esta investigación en particular, se optó por las herramientas Inflesz (Barrio-Cantalejo, 2007; Barrio-Cantalejo et al., 2008) y NooJ (Silberztein, 2016). Inflesz es un programa diseñado específicamente para determinar la lecturabilidad de los textos escritos en español. Opera de acuerdo con nueve parámetros para evaluar la lecturabilidad. Entre ellos, se destaca: n.º de sílabas, n.º de palabras, n.º de oraciones, promedio sílabas / palabras y promedio palabras / oración. Con base en estos datos, entrega un grado de lecturabilidad de acuerdo con su propia escala –Escala Inflesz–; adicionalmente, considera los índices Flesch-Szigriszt, Flesch-Fernández Huerta y Correlación Word (Barrio-Cantalejo, 2007; Barrio-Cantalejo et al., 2008). Entre los méritos de esta herramienta –que justifican su elección– se considera el hecho de que fue diseñada específicamente para el español, a partir de un corpus de textos construido de manera aleatoria y representativa; asimismo, se trata de un *software* de libre disponibilidad, de rápida y fácil aplicación (Barrio-Cantalejo, 2007; Barrio-Cantalejo et al., 2008).

Por su parte, NooJ es un *software* que procesa automáticamente textos escritos en lenguaje natural, de modo que facilita la descripción exhaustiva de diversas lenguas. También, NooJ permite integrar diferentes niveles de descripción –léxico, sintáctico, discursivo–, lo que facilita abordar distintos fenómenos lingüísticos (Silberztein, 2016). Considerando la versatilidad de la herramienta, que permite su uso en español sin ningún inconveniente; la facilidad de su uso, tras un entrenamiento acotado; y la posibilidad de estudiar fenómenos de interés en distintos niveles –tal como se precisó– (Silberztein, 2016) se optó por emplear este programa en particular en el contexto de esta investigación. En el siguiente apartado se ahonda en el funcionamiento de ambos programas.

4. MÉTODO

4.1. Corpus

El corpus estuvo constituido por los textos pertenecientes a los géneros EC y NOT presentes en el TE de Lenguaje y Comunicación de Octavo Año Básico en Chile, los que corresponden a un total de 60 textos: 34 corresponden a EC (5.093 palabras) y 26, a NOT (2.709 palabras). Estos textos fueron extraídos del corpus conformado por Ibáñez et al. (2017), el cual recopiló 100 TE correspondientes a diversas asignaturas del Segundo Ciclo de Educación Básica distribuidos por el Ministerio de Educación de Chile entre los años 2012 y 2016. Los textos escolares analizados en el presente estudio fueron publicados por las editoriales Piedra de Sol, Santillana, Ediciones SM Chile, Cal y Canto y Zig-Zag.

4.2. Procedimiento de análisis

A continuación se detalla el procedimiento de análisis realizado con las herramientas seleccionadas y para cada una de las medidas consideradas para el cálculo de la lecturabilidad de los GCO seleccionados.

4.2.1. Análisis mediante la herramienta Inflesz

En primer lugar, se empleó el programa Inflesz (Barrio-Cantalejo, 2007; Barrio-Cantalejo et al., 2008) para determinar la extensión de las palabras y de las oraciones. De acuerdo con esos datos, el programa asigna un puntaje y una etiqueta a los textos a analizar según la siguiente escala:

Tabla I. Escala Inflesz.

Puntos	Grado
Menos de 40	Muy difícil
40-55	Algo difícil
55-65	Normal
65-80	Bastante fácil
Más de 80	Muy fácil

Más adelante, en la Figura 3, se aprecia la visualización de un texto correspondiente a EC en el programa Inflesz, que será sometido al cálculo de lecturabilidad de acuerdo con los parámetros descritos en la Tabla I.

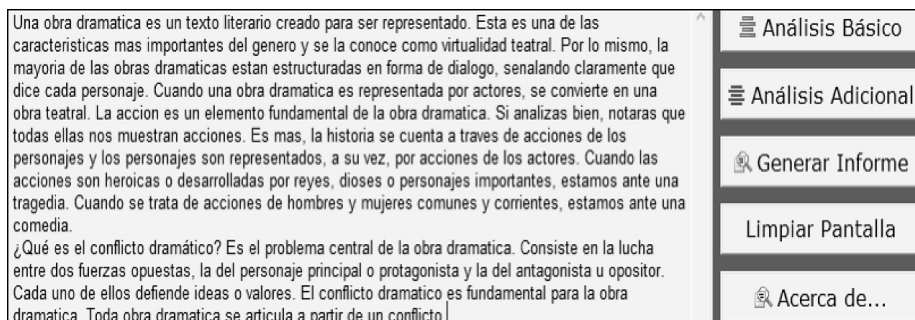


Figura 3. Visualización de un texto EC para ser analizado en Inflesz.

Como se observa en la Figura 3, a la derecha del texto, el programa otorga una serie de opciones. La opción *análisis básico* es la que brinda el cálculo de lecturabilidad de acuerdo con la Escala Inflesz –Tabla I– usada en este estudio. Junto con otorgar el cálculo de lecturabilidad, en la Figura 4 se distinguen otros aspectos que provee Inflesz: cantidad de sílabas, palabras y oraciones. A partir de estos datos, el *software* estima el promedio de sílabas por palabra y el promedio de palabras por oración. A su vez, con base en esta última información, Inflesz calcula el grado de lecturabilidad del texto de acuerdo con su escala. La información respecto del número de sílabas por palabras, palabras por oración y grado en la Escala Inflesz fue registrada en una planilla Excel.

Sílabas:	431
Palabras:	202
Frases:	15
Promedio sílabas / palabra:	2,13
Promedio palabras / frase:	13,47
Índice Flesch-Szigiszt:	60,44
Grado en la Escala Inflesz:	Normal

Figura 4. Información Inflesz.

4.2.2. Análisis mediante la herramienta NooJ

Tras realizar el análisis con Inflesz, se procedió con la búsqueda de nominalizaciones; oraciones subordinadas sustantivas y adjetivas; marcadores discursivos y relaciones de coherencia. Este análisis se realizó con el programa NooJ (Silberstein, 2016), el cual funciona según las pautas que le brindan diccionarios y gramáticas

elaborados previamente por el investigador. Los diccionarios son listas de unidades lingüísticas a las cuales se les asigna una etiqueta que corresponde a una categoría, por ejemplo, morfológica (p. ej. sustantivo), sintáctica (p. ej. sujeto) o discursiva (p. ej. marcador). Luego, el programa identifica dichas unidades en un texto y les asigna la etiqueta correspondiente. Por su parte, las gramáticas son modelos elaborados a partir de un patrón morfológico regular que se asocia con distintas palabras. Por ejemplo, se podría elaborar una gramática que busque palabras terminadas en *-mente*, para identificar los adverbios que provienen de adjetivos transformados por sufijación. En los siguientes apartados se presenta el detalle de los indicadores de lecturabilidad buscados con NooJ.

- Identificación de subordinadas

Se construyó un diccionario que incluyó los pronombres relativos. De ellos, *que* tuvo una doble etiqueta: pronombre y, además, conjunción. De este modo, fue factible identificar oraciones subordinadas sustantivas y adjetivas. El ejemplo (8) corresponde a una estructura subordinada sustantiva y (9), a una adjetiva:

- (8) El presidente dijo *que impulsará una reducción de impuestos*.
- (9) El presidente, *que fue recientemente elegido*, impulsará una reducción de impuestos.

En (8), *que* opera como un conector que introduce la oración subordinada sustantiva (destacada en cursiva). Por su parte, en (9), *que* opera como un pronombre que remite a *presidente*, al mismo tiempo que funciona como sujeto de la subordinada adjetiva (destacada en cursiva).

- Identificación de nominalización

Se confeccionó una gramática para identificar todas las palabras terminadas en *-sión*, *-ción*, *-miento* y *-dor*, las que corresponden a las terminaciones de las nominalizaciones más recurrentes en el discurso especializado (Cademártori et al., 2006).

- Identificación de marcadores y relaciones de coherencia

Se elaboró un diccionario de marcadores del discurso con más de 220 entradas a partir de las taxonomías propuestas por Calsamiglia y Tusón (1999), Zorraquino y Portolés (1999) y Cuenca (2010). Con él, se determinó la presencia (P) o no presencia (NP) de marcadores discursivos en el corpus. Luego, se estableció si el tipo de relación de coherencia predominante era aditiva (A), temporal (T) o causal (C) en cada texto de acuerdo con el tipo de marcador más frecuente.

4.2.3. Aplicación de diccionarios y gramática

A continuación se describe el procedimiento de aplicación de diccionarios y gramática que permitieron identificar oraciones subordinadas, nominalizaciones, marcadores del discurso y relaciones de coherencia. La Figura 5 muestra cómo se visualiza el análisis en NooJ:

<p>Una obra dramática es un texto literario creado para ser representado. Esta es una de las características más importantes del género y se la conoce como <u>virtualidad teatral</u>. Por lo mismo, la mayoría de las obras dramáticas están estructuradas en forma de diálogo, señalando claramente qué dice cada personaje. Cuando una obra dramática es representada por actores, se convierte en una obra teatral. La acción es un elemento fundamental de la obra dramática. Si analizas bien, notarás que todas ellas nos muestran acciones. Es más, la historia se cuenta a través de acciones de los personajes y los personajes son representados, a su vez, por acciones de los actores. Cuando las acciones son heroicas o desarrolladas por reyes, dioses o personajes importantes, estamos ante una tragedia. Cuando se trata de acciones de hombres y mujeres comunes y corrientes, estamos ante una comedia. ¿Qué es el conflicto dramático? Es el problema central de la obra dramática. Consiste en la lucha entre dos fuerzas opuestas, la del personaje principal o protagonista y la del antagonista u opositor. Cada uno de ellos defiende ideas o valores. El conflicto dramático es fundamental para la obra dramática. Toda obra dramática se articula a partir de un conflicto.</p>					
0	311	403	459	471	485
y,MARC+aditivo	cuando,MARC+temporal+causal	acción,NOM	si,MARC+causal	bien,MARC+aditivo	que,PRON+relativo CONJ

Figura 5. Análisis NooJ.

Como se aprecia en la parte inferior de la Figura 5, con NooJ se identificó las nominalizaciones requeridas, los pronombres u otros términos que constituyen indicadores de una oración subordinada y también, los marcadores discursivos. A diferencia del análisis realizado con Inflesz, la búsqueda de unidades lingüísticas ejecutada con NooJ requirió ratificación humana con la finalidad de otorgar mayor validez al análisis. Por ejemplo, en el caso de la Figura 5, el programa identifica el marcador *y*, el cual tiene aparejada la etiqueta MARC (marcador) + aditivo (tipo) (área inferior izquierda de la Figura 5). Sin embargo, es necesario determinar si efectivamente se trata de un marcador asociado a una relación de coherencia. Esto se corrobora dirigiéndose a los segmentos ligados por el marcador *y*, que, para mayor claridad, se subrayaron (área central superior de la Figura 5).

El procedimiento de ratificación se aplicó a cada marcador identificado; de igual modo, cada oración subordinada y nominalización deverbal se examinó individualmente. Estos datos se registraron en la misma planilla Excel en que se consignó la información otorgada por Inflesz. Así, en este documento se tomó nota,

por cada texto del corpus, de la presencia (P) o no presencia (NP) de marcadores, del tipo de relación de coherencia predominante –aditiva (A), temporal (T) o causal (C)–, el número de nominalizaciones y la cantidad de oraciones subordinadas, además del grado Inflesz, número de sílabas por palabra y la cantidad de palabras por oración.

5. RESULTADOS Y DISCUSIÓN

A continuación se presentan, en primer lugar, los resultados correspondientes al grado de lecturabilidad del corpus analizado según la Escala Inflesz, señalando el promedio de sílabas por palabra y de palabras por oración. Luego, se presentan los resultados correspondientes a nominalizaciones, subordinaciones, marcadores y relaciones de coherencia obtenidos con NooJ. La información relativa a subordinaciones y nominalizaciones se presenta según las categorías de la Escala Inflesz, ya que se enmarca dentro de las mismas dimensiones que contempla esa escala –de la palabra y la oración–. Por el contrario, los resultados correspondientes a marcadores y relaciones de coherencia se exhiben al margen de las categorías Inflesz pues pertenecen a una dimensión que excede el análisis de dicho programa.

5.1. Grado de lecturabilidad Inflesz

El Gráfico 1 presenta la distribución de los textos presentes en el corpus de acuerdo con los grados establecidos en la Escala Inflesz:

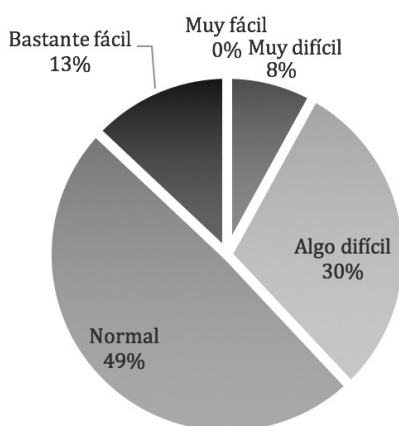


Gráfico 1. Distribución de los textos del corpus según su grado de lecturabilidad Inflesz.

De acuerdo con el Gráfico 1, el grado prevalente (49%) corresponde a «normal». Según la Escala Inflesz, este tipo de texto está destinado a sujetos entre 12 y 16 años. Considerando que el TE de Lenguaje y Comunicación de Octavo Año Básico corresponde a jóvenes entre 13 y 14 años, esta porción del corpus sí se ajusta a sus destinatarios en términos de sílabas por palabra y palabras por oración. Sin embargo, la mayor parte de los textos (51%) se sitúa sobre o debajo el nivel de los estudiantes. En concreto, 30% de los textos se ubica en el grado «algo difícil», que corresponde a las edades entre 15 y 16 años; 13%, en el grado «bastante fácil», que coincide con las edades entre 9 y 12 años; 8%, en el grado «muy difícil», que va desde los 17 años en adelante. Por tanto, una parte importante del TE de Lenguaje y Comunicación no se adecua a los estudiantes en cuanto a su grado de lecturabilidad Inflesz.

Con base en lo anterior, cabe afirmar que una parte importante del TE de Lenguaje y Comunicación no se adecua a los estudiantes. Estos resultados coinciden con los obtenidos en investigaciones llevadas a cabo en otros países con Inflesz o herramientas afines (Kasule, 2011; Guamán, 2014), donde también se estableció que los textos escolares no se ajustan al nivel de comprensión que en principio poseen los estudiantes. Las implicancias son serias, en tanto se ha probado que existe una relación directa entre la lecturabilidad de los TE y la motivación de los estudiantes para el aprendizaje (Guamán, 2014; McNamara et al., 2014), de modo que se entorpecen los procesos de enseñanza-aprendizaje cuando no tiene lugar un *match* adecuado entre textos y alumnos. Esta situación resulta particularmente delicada cuando se trata del desarrollo de la comprensión de textos. En este caso, emparejar lectores con TE comprensibles para ellos constituye uno de los presupuestos de la adquisición y desarrollo de habilidades de lectura (Crossley et al., 2019).

5.2. Promedios de sílabas por palabras y palabras por oración

El Gráfico 2 exhibe los promedios de sílabas por palabra y palabras por oración:

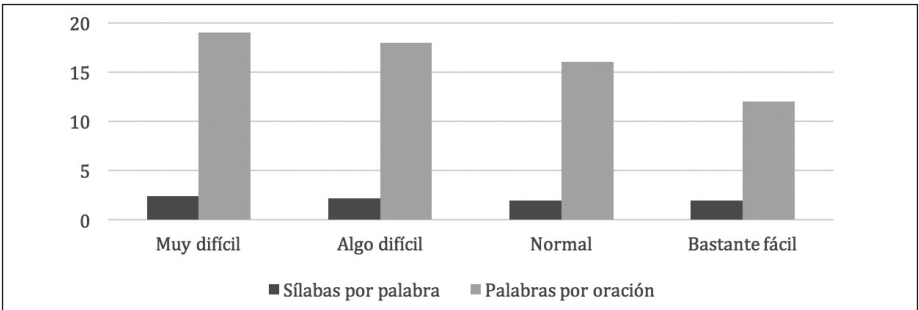


Gráfico 2. Promedios de sílabas por palabra y palabras por oración Inflesz.

Tal como se observa en el Gráfico 2, el promedio de sílabas por palabra es relativamente bajo y constante a lo largo de las cuatro categorías. En la categoría «muy difícil» el promedio fue 2,4; en la categoría «algo difícil», 2,2; en las categorías «normal» y «bastante fácil», 2. Estos valores se acercan o incluso coinciden con el número de sílabas promedio que poseen las palabras empleadas de manera cotidiana en español. Debido a esto, se podría asumir que las palabras presentes en el corpus son fáciles de procesar y comprender (Alliende, 2002). La familiaridad de las palabras constituye un factor fundamental en la comprensión, en tanto *input* inicial con base en el que se elaboran unidades de significado que, a su vez, corresponden al presupuesto necesario de construcción del Modelo de Situación (Perfetti & Stafura, 2014).

Por el contrario, en la extensión de las oraciones se observa un elevado número de palabras a través de las cuatro categorías. En la categoría «muy difícil» el promedio fue 19; en la categoría «algo difícil», 18; en la categoría «normal», 16; finalmente, en la categoría «bastante fácil», 12. Estos valores implican una dificultad para la capacidad limitada de la memoria de trabajo (McNamara & Magliano, 2009), que, de hecho, suele oscilar entre los cinco y los nueve ítems –palabras– de información (Santos, 2015). La memoria de trabajo desempeña un rol central en el procesamiento y comprensión de textos, en tanto permite conservar disponible la información importante a lo largo del proceso de lectura para su integración (Bartek et al., 2011; Gibson, 1998; Kleijn, 2018). Si las exigencias durante el proceso de comprensión aumentan –como en este caso–, los recursos disponibles para conservar activos los ítems en la memoria operativa decrecen, lo que se traduce en un perjuicio del análisis sintáctico, la integración semántica o ambos (Guzmán et al., 2016).

5.3. Dimensión de la palabra: nominalización

En la Tabla II se consigna la cantidad de nominalizaciones en relación con el total de palabras de cada uno de los conjuntos de textos agrupados según su grado de lecturabilidad Inflesz:

Tabla II. Presencia de nominalización en relación con el total de palabras.

Categorías Inflesz	Nominalización
Muy difícil	5%
Algo difícil	2%
Normal	2%
Bastante fácil	2%

La Tabla II exhibe que la presencia de nominalizaciones es escasa a través de las cuatro categorías. En los textos agrupados en el grado «muy difícil», las nominalizaciones corresponden al 5% de un total de 385 palabras; por su parte, en los textos situados en los grados «algo difícil» (1.891 palabras), «normal» (4.550 palabras) y «bastante fácil» (976 palabras) la cantidad de nominalizaciones no superan el 2% del total.

Así, los resultados obtenidos con el uso de NooJ son coherentes con Inflesz. Con esta herramienta, se estableció que el promedio de sílabas por palabra fue de 2.1, cifra que constituye un indicador de que el vocabulario del TE de Lenguaje y Comunicación es sencillo (Alliende, 2002). Por tanto, es consistente que las nominalizaciones, rasgo distintivo de textos que poseen vocabularios más complejos (García et al., 2005; Cademártori et al., 2006; Cinto, 2009; Pander Maat et al., 2014), tuviesen escasas ocurrencias en el corpus analizado.

Como se consignó en el apartado teórico, las nominalizaciones suponen un incremento de la ambigüedad, la abstracción y el establecimiento de relaciones anafóricas, muchas veces, difíciles de determinar por parte de lectores inexpertos, categoría en la que precisamente se ubican –en principio– los estudiantes de Segundo Ciclo de Enseñanza Básica (Cinto, 2009; García et al., 2005). A ello se debe sumar que la nominalización permite condensar grandes cantidades de información, cuyo análisis es complejo para los lectores no experimentados (Cinto, 2009). En este contexto, la escasa presencia de nominalizaciones en el TE de Lenguaje y Comunicación favorece tanto el procesamiento como la comprensión de los estudiantes.

5.4. Dimensión oracional: subordinación

En la Tabla III se presenta el porcentaje de subordinación en relación con el total de oraciones de cada uno de los conjuntos de textos agrupados según su grado de lecturabilidad Inflesz:

Tabla III. Presencia de subordinación en relación con el total de oraciones.

Categorías Inflesz	Subordinación
Muy difícil	50%
Algo difícil	46%
Normal	45%
Bastante fácil	38%

Los resultados revelan una presencia considerable de subordinación en relación con el total de oraciones en cada categoría Inflesz. En la categoría «muy difícil», la subordinación se presenta, en promedio, en el 50% de un total de 22 oraciones; en la categoría «algo difícil», en el 46% de 113 oraciones; en la categoría «normal», en el 45% de 291 oraciones; por último, en la categoría «bastante fácil» corresponde al 38% de 92 oraciones.

Los resultados conseguidos con NooJ son consistentes con la información otorgada por Inflesz. Con NooJ se logró establecer que la presencia de estructuras oracionales subordinadas corresponde a un porcentaje elevado en relación con el promedio de oraciones por texto. A su vez, con Inflesz se identificó un promedio general de 16 palabras por oración, un número elevado (Santos, 2015), que permite asumir, precisamente, que un texto posee estructuras sintácticas complejas (Alliende, 2002; Guamán, 2014).

Las estructuras sintácticas que poseen oraciones subordinadas implican –en principio– un reto superior para los lectores. Estas estructuras poseen una mayor longitud de dependencia sintáctica entre algunos de sus elementos constituyentes –por ejemplo, el núcleo del sintagma nominal con función de sujeto y el verbo principal– respecto a una oración simple, debido a la incorporación de una categoría incrustada o dependiente, como se precisó en el apartado teórico (Bartek et al., 2011; Gibson, 1998; Kleijn, 2018). El incremento de la extensión de la longitud de dependencia, con ocasión de la integración de dichas estructuras dependientes, exige a los lectores más recursos cognitivos de procesamiento, recuperación e integración de elementos respecto a estructuras más simples (Kleijn, 2018). Ante este panorama, su elevada presencia en el TE de Lenguaje y Comunicación constituye un desafío para los estudiantes.

5.5. Dimensión discursiva: marcadores del discurso

De acuerdo con el análisis realizado, el 87% del corpus (52 textos) sí presenta marcadores del discurso. Por el contrario, el 13% del corpus (8 textos) no los presenta. Estos resultados implican que los textos analizados favorecen, en mayor medida, la identificación, procesamiento y comprensión de las relaciones de coherencia, ya que las explicitan en la superficie textual (Sanders & Noordman, 2000; Degand & Sanders, 2002; Van Silfhout et al., 2014; Moncada, 2018; Ibáñez, et al., en prensa).

Dado el rol fundamental que desempeñan las relaciones de coherencia en la construcción de una representación mental coherente del texto, su explicitación por medio de marcadores discursivos debiese poseer un efecto favorable, a su vez, sobre el procesamiento y la comprensión (Das, 2014). En este sentido, existen

investigaciones que han demostrado que las relaciones de coherencia se procesan más rápidamente al estar señalizadas en la superficie textual; asimismo, otras investigaciones han dejado en evidencia que las relaciones explícitamente marcadas conducen a un mejor desempeño en evaluaciones de comprensión (Sanders & Noordman, 2000).

5.6. Dimensión discursiva: relaciones de coherencia

Como se manifestó en apartados precedentes, se utilizaron los marcadores discursivos como un *proxy* para identificar las relaciones de coherencia en el corpus. Por tanto, la información que se presenta a continuación no considera los textos que carecían de marcadores (8 textos). En los textos en que predominó más de un tipo de marcador, se empleó una etiqueta compuesta: Aditiva-Temporal, Aditiva-Causal, Temporal-Causal o Temporal-Causal-Aditiva.

De acuerdo con el análisis realizado, en la mayor parte del corpus predominan las relaciones de coherencia aditiva con un 48% (25 textos). En la tercera mayoría con un 15% (8 textos), también están presentes las relaciones aditivas, aunque, en este caso, comparten predominancia con relaciones causales. A partir de estos resultados se puede inferir que la mayor parte de los textos analizados implica, para los lectores, un procesamiento fácil, pero, al mismo tiempo, un recuerdo más débil del contenido, en tanto las relaciones de coherencia aditiva se traducen en una representación del discurso menos organizada y, asimismo, dificultosa de recuperar desde la memoria (Sanders & Noordman, 2000). La dificultad de recuperación implica, no solo un enlentecimiento del procesamiento y una menor calidad de la representación del significado del texto, sino que, principalmente, un impedimento para aprender del texto que se lee.

6. CONCLUSIONES

El objetivo que orientó esta investigación fue caracterizar la lecturabilidad de los textos pertenecientes a los géneros Exposición de Contenidos y Nota presentes en el TE de Lenguaje y Comunicación de Octavo Año de Enseñanza Básica en Chile. En este sentido, se sometió cada texto a un doble análisis con los programas Inflexz (Barrio-Cantalejo, 2007; Barrio-Cantalejo et al., 2008) y NooJ (Silberztein, 2016). El primero permitió identificar el grado de lecturabilidad de cada texto, la extensión de palabras y la extensión de oraciones. El segundo, establecer la presencia de nominalizaciones, estructuras subordinadas y, sobre todo, integrar dos indicadores normalmente excluidos de los estudios de lecturabilidad en español: los marcadores discursivos y las relaciones de coherencia.

Se reveló que la mayor parte del corpus, de acuerdo con la Escala Inflesz, no se adecua a sus destinatarios. Se determinó, asimismo, que una de las dificultades de los estudiantes al enfrentar el TE de Lenguaje y Comunicación de Octavo Año de Enseñanza Básica reside en la dimensión de la oración, ya que la extensión de las oraciones presentó números elevados en relación con la capacidad limitada de la memoria de trabajo (McNamara & Magliano, 2009; Santos, 2015). Por su parte, la extensión de las palabras se mantuvo relativamente constante y baja a través de cada una de las categorías Inflesz. Esto último implica que el vocabulario de los textos analizados es, en principio, fácil de procesar y comprender (Alliende, 2002).

El análisis con la herramienta NooJ corroboró que el fundamento de la dificultad de los textos se hallaba, principalmente, en la dimensión oracional. Los textos analizados exhiben una presencia elevada de estructuras sintácticas subordinadas y, por el contrario, escasas nominalizaciones. Respecto a los marcadores discursivos, se determinó que la mayor parte del corpus sí los presentaba; esto facilita la identificación, procesamiento y comprensión de las relaciones de coherencia que vertebran el discurso (Sanders & Noordman, 2000; Degand & Sanders, 2002; Van Silfhout et al., 2014). En cuanto al tipo de relación de coherencia predominante, se observó que, en la mayor parte del corpus analizado, las relaciones aditivas son las más frecuentes. Esto supone un recuerdo más débil del contenido del texto y, por tanto, un aprendizaje menos significativo (Van den Broek, 1990b; Sanders & Noordman, 2000).

De acuerdo con los resultados de esta investigación, se puede afirmar que una porción relevante del TE de Lenguaje y Comunicación no cumpliría plenamente con los lineamientos declarados por el propio Ministerio de Educación chileno en cuanto a que el TE representa un instrumento de equidad y enriquecimiento para el estudiante y las familias (MINEDUC, 2012). Difícilmente el TE podría permitir un acceso equitativo a los diversos conocimientos y habilidades si no presenta los rasgos lingüísticos y discursivos adecuados en función de su propósito y sus destinatarios –los estudiantes–. Resultaría pertinente que el Ministerio de Educación evalúe los TE en función de rasgos de lecturabilidad e introduzca las mejoras correspondientes. Actualmente, MINEDUC dispone de un departamento de evaluación orto-tipográfica, evaluación en el marco de la cual solo se cuantifican los errores tipográficos y ortográficos, inconsistencias en los índices y errores de correspondencia (MINEDUC, 2012). Además, como se planteó en la introducción, el principal propósito de la asignatura de Lenguaje y Comunicación es optimizar la comprensión de los estudiantes (MINEDUC, 2012). Al respecto, se concluye que el TE no estaría contribuyendo cabalmente a su misión.

Junto con evidenciar las características de lecturabilidad de dos de los CGO más frecuentes en el TE de Lenguaje y Comunicación, esta investigación contribuye a los estudios de lecturabilidad en español incorporando los marcadores discursivos y las relaciones de coherencia, indicadores no considerados en estu-

dios anteriores en esta lengua. De forma típica, los análisis de lecturabilidad en español se han restringido a la dimensión de la palabra y la oración, y, en ambos casos, muchas veces no se consideran otros indicadores fuera de la extensión de palabras y oraciones. Sin embargo, como se planteó en los primeros apartados de este artículo, dada la naturaleza de la comprensión lectora, necesariamente la lecturabilidad de los textos debe juzgarse de acuerdo con un mayor número de dimensiones e indicadores. Al incorporar la dimensión discursiva en este estudio, se ha logrado dicho propósito y se ha brindado una visión integral de la complejidad que revisten los CGO más frecuentes en el TE de Lenguaje y Comunicación.

Desde un punto de vista metodológico, destacamos el uso combinado de herramientas computacionales para el análisis de la lecturabilidad en español. Inflex y NooJ facilitaron el análisis de cada uno de los indicadores contemplados en esta investigación; al mismo tiempo, introdujeron un mayor grado de precisión en la identificación de las unidades lingüísticas y discursivas de interés para la caracterización de la lecturabilidad del corpus de estudio. Sin perjuicio de lo anterior, constituiría un aporte contundente para los análisis de lecturabilidad en español avanzar en la elaboración de un programa de libre disposición que integre las funcionalidades de ambas herramientas e identifique con un grado aún mayor de precisión las unidades de interés.

Adicionalmente, se contempla como proyección llevar a cabo investigación psicolingüística que permita establecer empíricamente el impacto de las nominalizaciones, subordinaciones, marcadores discursivos, relaciones de coherencia y otros aspectos lingüísticos y discursivos en relación con la comprensión lectora en español. Así, se podría obtener una visión más certera de la lecturabilidad a través de la confección de índices que se articulen con los que ya existen para el español –extensión de palabras y oraciones–. Este índice debiese integrarse en el mismo programa señalado en el párrafo precedente, para facilitar el análisis de lecturabilidad, tanto para investigadores como para quienes son responsables de mediar la lectura en distintos contextos –p. ej. los profesores– (McNamara et al., 2014).

REFERENCIAS

- Alliende, Felipe. 2002. *La legibilidad de los textos: Manual para la evaluación, selección y elaboración de textos*. Santiago, Chile: Editorial Andrés Bello.
- Barrio-Cantalejo, Inés. 2007. Legibilidad y salud. Los métodos de medición de la legibilidad y su aplicación al diseño de folletos educativos sobre salud. Tesis de Doctorado. Madrid, España: Universidad Autónoma de Madrid.
- Barrio-Cantalejo, Inés, Simón-Lorda, Pablo, Melguizo, Miguel, Escalona, Isabel, Marijuán, María y Hernando, Pablo. 2008. Validación de la escala IN-

- FLESZ para evaluar la legibilidad de los textos dirigidos a pacientes. *Anales del Sistema Sanitario de Navarra*, 31(2), 135-152. http://scielo.isciii.es/scielo.php?script=sci_arttext&pid=S1137-66272008000300004
- Bartek, Brian, Lewis, Richard, Vasishth, Shravan y Smith, Mason. 2011. In search of on-line locality effects in sentence comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 37(5), 1178-1198. <https://doi.org/10.1037/a0024194>
- Bernstein, Basil. 1990. *The structuring of pedagogic discourse: Class, codes y control*, Volume IV. London, England: Routledge.
- Cademártori, Yanina, Parodi, Giovanni y Venegas, René. 2006. El discurso escrito y especializado: Caracterización y funciones de las nominalizaciones en los manuales técnicos. *Literatura y Lingüística*, 17, 243-265. <http://dx.doi.org/10.4067/S0716-58112006000100015>
- Calsamiglia, Helena y Tusón, Amparo. 1999. *Las cosas del decir. Manual de análisis del discurso* (2da Edición.) Barcelona, España: Editorial Ariel.
- Campbell, Yilmar. 2018. Sintaxis y comprensión de textos: interpretación de oraciones subordinadas. Recuperado de servicio.bc.uc.edu.ve/educacion/revista/a1n18/1-18-7.pdf
- Christie, Frances. 2002. *Classroom discourse analysis: A functional perspective*. London, England: Continuum.
- Cinto, María. 2009. La nominalización. Obstáculo para la comprensión lectora. *Educación, Lenguaje y Sociedad*, 6, 171-186. <http://www.biblioteca.unlpam.edu.ar/pubpdf/ieles/n06a09cinto.pdf>
- Cozijn, Reinier, Noordman, Leo y Vonk, Wietske. 2011. Propositional integration and world-knowledge inference: Processes in understanding because sentences. *Discourse Processes*, 48(7), 475-500. <https://doi.org/10.1080/0163853X.2011.594421>
- Crossley, Scott, Allen, Laura, Kyle, Kristopher y McNamara, Danielle. 2014. Analyzing discourse processing using a simple natural language processing tool. *Discourse Processes*, 51(5-6), 511-534. <https://doi.org/10.1080/0163853X.2014.910723>
- Crossley, Scott, Kyle, Kristopher y McNamara, Danielle. 2016. The tool for the automatic analysis of text cohesion (TAACO): Automatic assessment of local, global, and text cohesion. *Behavior Research Methods*, 48, 1227-1237. <https://link.springer.com/article/10.3758/s13428-015-0651-7>
- Crossley, Scott, Skalicky, Stephen, Dascalu, Mihai, McNamara, Danielle y Kyle, Kristopher. 2017. Predicting text comprehension, processing, and familiarity in adult readers: New approaches to readability formulas. *Discourse Processes*, 54(5-6), 340-359. <https://doi.org/10.1080/0163853X.2017.1296264>
- Crossley, Scott, Skalicky, Stephen, Dascalu, Mihai. 2019. Moving beyond classic readability formulas: new methods and new models. *Journal of Research in*

- Reading*, 42(3-4), 541-561. <https://doi.org/10.1111/1467-9817.12283>
- Cuenca, María. 2010. *Gramática del texto*. Madrid, España: Arco/Libros, S.L.
- Das, Debopam. 2014. Signalling of Coherence Relations in Discourse. Tesis de Doctorado. Columbia Británica, Canada: Simon Fraser University.
- Degand, Liesbeth y Sanders, Ted. 2002. The impact of relational markers on expository text comprehension in L1 and L2. *Reading and Writing: An Interdisciplinary Journal*, 15, 739-757. <https://doi.org/10.1023/A:1020932715838>
- DuBay, William. 2007. *Smart language: Readers, readability, and the grading of text*. California, USA: Impact Information.
- Dufty, David, Graesser, Arthur, Louwerse, Max y McNamara, Danielle. 2006. Assigning grade levels to textbooks: Is it just readability? 28th Annual Conference of the Cognitive Science Society, Vancouver, Canadá.
- Dyson, Dana y Schellenberg, Kathryn. 2017. Access to justice: The readability of legal services corporation legal aid internet services. *Journal of Poverty*, 21(2), 142-165. <https://doi.org/10.1080/10875549.2016.1186773>
- García, Marta, Hall, Beatriz y Marín, Marta. 2005. Ambigüedad, abstracción y polifonía del discurso académico: Interpretación de las nominalizaciones. *Revista Signos*, 38(57), 49-60. <http://dx.doi.org/10.4067/S0718-09342005000100004>
- Gibson, Edward. 1998. Linguistic complexity: locality syntactic dependencies. *Cognition*, 98, 1-76.
- Goldman, Susan y Varnhagen, Connie. 1986. Memory for embedded and sequential story structures. *Journal of Memory and Language*, 25(4), 401-418. [https://doi.org/10.1016/0749-596X\(86\)90034-3](https://doi.org/10.1016/0749-596X(86)90034-3)
- Graesser, Arthur y Clark, Leslie. 1985. *Structures and Procedures of Implicit Knowledge*. Norwood, NJ: Ablex Publishing Corporation.
- Graesser, Arthur, McNamara, Danielle, Louwerse, Max y Cai, Zhiqiang. 2004. Coh-Metrix: Analysis of text on cohesion and language. *Behavior Research Methods, Instruments, y Computers*, 36, 193-202. <https://doi.org/10.3758/BF03195564>
- Guamán, Alonso. 2014. La legibilidad de los textos escolares como elemento motivador del aprendizaje de los alumnos de Educación General Básica Superior. Tesis de Doctorado. Piura, Perú: Universidad Nacional de Piura.
- Guzmán, Belén, Véliz, Mónica y Reyes, Fernando. 2016. Memoria operativa, comprensión lectora y rendimiento escolar. *Literatura y Lingüística*, 35, 379-404. <https://www.redalyc.org/articulo.oa?id=35252277019>
- Hall, Sophie, Maltby, John, Filik, Ruth y Paterson, Kevin. 2014. Key skills for science learning: the importance of text cohesion and reading ability. *Educational Psychology*, 36(2), 191-215. <https://doi.org/10.1080/01443410.2014.926313>
- Halliday, Michael y Matthiessen, Christian. 2014. *Halliday's introduction to Functional Grammar* (4th Edition). New York, USA: Routledge.
- Holguín, John y Rodríguez, Marcos. 2017. Inferencias causales de profesores que

- realizan la lectura de textos con implicancia causal y subordinación oracional. *Fides et Ratio*, 14(14), 75-98. http://www.scielo.org.bo/pdf/rfer/v14n14/v14n14_a06.pdf
- Ibáñez, Romualdo, Moncada, Fernando y Santana, Andrea. 2015. Variación disciplinar en el discurso académico de Biología y del Derecho: Un estudio a partir de las relaciones de coherencia. *Onomázein*, 32, 101-131. <https://www.redalyc.org/pdf/1345/134544960010.pdf>
- Ibáñez, Romualdo, Moncada, Fernando, Cornejo, Felipe y Arriaza, Valeria. 2017. Los géneros del conocimiento en textos escolares de educación primaria. *Caleidoscopio*, 15(3), 462-476. <https://doi.org/10.4013/cld.2017.153.06>
- Ibáñez, Romualdo, Moncada, Fernando y Arriaza, Valeria. 2018. Recontextualización del conocimiento en textos escolares chilenos. *Revista Signos*, 51(98), 430-456. <https://scielo.conicyt.cl/pdf/signos/v51n98/0718-0934-signos-51-98-00430.pdf>
- Ibáñez, Romualdo, Moncada, Fernando y Cárcamo, Benjamín. En prensa. Texto escolar y comprensión. Efecto de las características de las relaciones de coherencia en estudiantes con diferente nivel socioeconómico. *Revista Española de Lingüística Aplicada*.
- Kasule, Daniel. 2011. Textbook readability and ESL learners. *Reading & Writing, Journal of the Reading Association of South Africa*, 2(1), 63-76. <https://hdl.handle.net/10520/EJC131508>
- Kleijn, Suzanne. 2018. *Clozing in on readability: How linguistic features affect and predict text comprehension and on-line processing*. Tesis de Doctorado. Utrecht, the Netherlands: Utrecht University.
- Long, Lance y Christensen, William. 2011. Does the Readability of Your Brief Affect Your Chance of Winning an Appeal? *The Journal of Appellate Practice and Process*, 12(1), 144-162.
- McNamara, Danielle, Graesser, Arthur, McCarthy, Philip y Cai, Zhiqiang. 2014. *Automated evaluation of text and discourse with Coh-Metrix*. New York, USA: Cambridge University Press.
- McNamara, Danielle y Magliano, Joe. 2009. Toward a comprehensive model of comprehension. *Psychology of Learning and Motivation - Advances in Research and Theory*, 51, 297-384. [https://doi.org/10.1016/S0079-7421\(09\)51009-2](https://doi.org/10.1016/S0079-7421(09)51009-2)
- Meneses, Alejandra, Montenegro, Maximiliano, Ruiz, Marcela. 2014. Textos escolares para aprender ciencias: Habilidades, contenidos y lenguaje académico. En Centro de Estudios MINEDUC, *Evidencias para políticas públicas en educación: Selección de investigaciones Sexto Concurso FONIDE* (pp. 233-277). Santiago: Gráfika.
- MINEDUC. 2012. Bases curriculares para la enseñanza básica. Disponible en: http://www.curriculumenlineamineduc.cl/605/articles-22394_programa.pdf. Acceso en: 26/11/2017.

- MINEDUC, 2018. Textos escolares oficiales. Disponible en: <https://www.curriculumnacional.cl/portal/Tipo/Recursos-Digitales/Textos-escolares-oficiales/>
- Moncada, Fernando. 2018. Interacción entre conectores y conocimiento previo en el procesamiento de la coherencia causal. *Círculo de Lingüística Aplicada a la Comunicación*, 76, 179-196. <https://doi.org/10.5209/CLAC.62504>
- Noordman, Leo y Vonk, Wietske. 1997. The different functions of a conjunction in constructing a representation of the discourse. En J. Costermans y M. Fayol (Eds.), *Processing interclausal relationships. Studies in the production and comprehension of text* (pp. 75-93). Mahwah: Lawrence Erlbaum Associates, Inc.
- Pander Maat, Henk, Kraf, Rogier, Van den Bosch, Antal, Dekker, Nick, Van Gompel, Maarten, Kleijn, Suzanne y Van der Sloot, Ko. 2014. T-Scan: A new tool for analyzing Dutch text. *Computational Linguistics in the Netherlands Journal*, 4, 53-74. <http://dSPACE.library.uu.nl/handle/1874/322949>
- Perfetti, Charles y Stafura, Joseph. 2014. Word knowledge in a theory of reading comprehension. *Scientific Studies of Reading*, 18(1), 22-37. <https://doi.org/10.1080/10888438.2013.827687>
- Rose, David. 2014. Analyzing pedagogic discourse: An approach from genre and register. *Functional Linguistics*, 1(11), 1-32. <https://functionallinguistics.springeropen.com/articles/10.1186/s40554-014-0011-4>
- Sanders, Ted y Noordman, Leo. 2000. The role of coherence relations and their linguistic markers in text processing. *Discourse Processes*, 29(1), 37-60. https://doi.org/10.1207/S15326950dp2901_3
- Santos, Sara. 2015. Memoria operativa y comprensión lectora. Tesis de Grado. Valladolid, España: Universidad de Valladolid.
- Silberstein, Max. 2016. *Formalizing natural languages: The NooJ approach*. London, Great Britain: ISTE Ltd.
- Skierkowski, Dorothy, Florin, Paul, Harlow, Lisa, Machan, Jason y Ye, Yinjiao. 2019. A readability analysis of online mental health resources. *American Psychologist*, 74(4), 474-483. <https://doi.org/10.1037/amp0000324>
- Soaje de Elias, Raquel. 2018. Textos escolares: consideraciones didácticas. *Educación y Educadores*, 21(1), 73-92. <https://www.redalyc.org/jatsRepo/834/83455923004/html/index.html>
- Taboada, Maite. 2006. Discourse markers as signals (or not) of rhetorical relations. *Journal of Pragmatics*, 38(4), 567-592. <https://doi.org/10.1016/j.pragma.2005.09.010>
- Tanaka-Ishii, Kumiko, Tezuka, Satoshi y Terada, Hiroshi. 2010. Sorting texts by readability. *Computational Linguistics*, 36(2), 203-227. <https://doi.org/10.1162/coli.09-036-R2-08-050>
- Trabasso, Tom y Sperry, Linda. 1985. Causal relatedness and importance of story events. *Journal of Memory and Language*, 24(5), 595-611. [https://doi.org/10.1016/0749-596X\(85\)90048-8](https://doi.org/10.1016/0749-596X(85)90048-8)

- Van den Broek, Paul. 1990a. The causal inference maker: Towards a process model of inference generation in text comprehension. En D. Balota, G. Flores y K. Rayner (Eds.), *Comprehension processes in reading* (pp. 423-446). Hillsdale, New York: Lawrence Erlbaum Associates.
- Van den Broek, Paul. 1990b. Causal inferences in the comprehension of narrative texts. En: A. C. Graesser y G.H. Bower (Eds.), *The psychology of learning and motivation: Inferences and text comprehension* (pp.175-194). San Diego, USA: Academic Press.
- Van Silfhout, Gerdineke, Evers-Vermeul, Jacqueline y Sanders, Ted. 2014. Connectives as processing signals: How students benefit in processing narrative and expository texts. *Discourse Processes*, 52(1), 47-76. <https://doi.org/10.1080/0163853X.2014.905237>
- Zorraquino, María y Portolés, José. 1999. Los marcadores del discurso. En I. Bosque y V. Demonte (Eds.) *Gramática descriptiva de la lengua española* (vol. II) (pp. 4051-4214). Madrid, España: Espasa.