

**EVALUACIÓN AUTOMATIZADA Y JUICIO EXPERTO EN LA ESTIMACIÓN
DE LA COMPLEJIDAD TEXTUAL: UN ANÁLISIS APLICADO A TEXTOS
ESCOLARES DE CIENCIAS**

AVALIAÇÃO AUTOMATIZADA E JULGAMENTO ESPECIALIZADO NA ESTIMATIVA
DA COMPLEXIDADE TEXTUAL: UMA ANÁLISE APLICADA A TEXTOS ESCOLARES
DE CIÊNCIAS

AUTOMATED ASSESSMENT AND EXPERT JUDGMENT IN ESTIMATING TEXTUAL
COMPLEXITY: AN ANALYSIS APPLIED TO SCIENCE SCHOOL TEXTS

Nathaly Rayen Reyes Peralta
Universidad de Concepción, Chile
nathalyreyes@udec.cl
0009-0002-2269-6822

Pamela Francesca Ramírez-Peña
Universidad de Concepción, Chile
pamerami@udec.cl
0000-0001-6295-5820

Christian Soto
Universidad de Concepción, Chile
christiansoto@udec.cl
0000-0003-2118-792X

Diego Palma
Universidad de Concepción, Chile
dpalma@udec.cl
0000-0003-1540-7164

Resumen

La comprensión lectora en contextos escolares se ve afectada no solo por las características del lector, sino también por los rasgos lingüísticos de los textos. En este marco, la complejidad textual se ha convertido en un foco de estudio relevante, especialmente para mejorar los criterios de selección de materiales escolares. En el presente estudio se analiza la relación entre los juicios docentes sobre la complejidad de textos científicos escolares y los índices lingüísticos automatizados generados por TRUNAJOD, una herramienta computacional desarrollada para evaluar la lecturabilidad de textos en español.

Se empleó un diseño cuantitativo, no experimental y correlacional. Participaron docentes especialistas en lectura y escritura, quienes evaluaron un corpus de 80 textos del área de ciencias, extraídos de textos escolares de enseñanza media. Los textos fueron clasificados por nivel educativo (1º–2º y 3º–4º medio), y los datos fueron analizados mediante regresiones jerárquicas, pruebas t de Student y coeficiente de concordancia de Kendall. Los resultados indican que los índices con mayor poder predictivo del juicio de complejidad son de naturaleza léxica (familiaridad, frecuencia, diversidad, imaginabilidad) y cohesiva (solapamiento de verbos, marcadores de causa). Asimismo, se evidenció una variación en los criterios de evaluación según el nivel escolar, observándose un foco más amplio en niveles iniciales y uno más restringido al léxico en niveles superiores. Se concluye que los juicios docentes reflejan sensibilidad a las propiedades lingüísticas del texto, aunque mediados por su formación disciplinar. El estudio aporta evidencia para el diseño de herramientas híbridas de evaluación textual y plantea futuras investigaciones con corpus ampliados y evaluadores de distintas áreas disciplinares.

Palabras clave: Complejidad textual; juicio de expertos; lecturabilidad; Textos escolares; lingüística computacional.

Resumo

A compreensão leitora em contextos escolares é influenciada não apenas pelas características do leitor, mas também pelos traços linguísticos dos textos. Nesse contexto, a complexidade textual tornou-se um foco relevante de estudo, especialmente para aprimorar os critérios de seleção de materiais didáticos. Este estudo analisa a relação entre os julgamentos de professores sobre a complexidade de textos escolares de ciências e os índices linguísticos automatizados gerados pelo TRUNAJOD, uma ferramenta computacional desenvolvida para avaliar a legibilidade de textos em espanhol.

Foi utilizado um delineamento quantitativo, não experimental e correlacional. Participaram professores especialistas em leitura e escrita, que avaliaram um corpus de 80 textos da área de ciências, extraídos de livros didáticos do ensino médio. Os textos foram classificados por nível educacional (1º–2º e 3º–4º anos do ensino médio), e os dados foram analisados por meio de regressões hierárquicas, testes t de Student e coeficiente de concordância de Kendall. Os resultados indicam que os índices com maior poder predictivo do julgamento de complexidade são de natureza lexical (familiaridade, frequência, diversidade, imaginabilidade) e coesiva (sobreposição de verbos, marcadores de causa). Além disso, observou-se uma variação nos critérios de avaliação conforme o nível escolar, com um foco mais amplo nos níveis iniciais e um foco mais restrito no léxico nos níveis superiores. Conclui-se que os julgamentos dos professores refletem sensibilidade às propriedades linguísticas dos textos, embora mediados por sua formação disciplinar. O estudo fornece evidências para o desenvolvimento de ferramentas híbridas de avaliação textual e propõe futuras investigações com corpora ampliados e avaliadores de diferentes áreas disciplinares.

Palavras-chave: Complexidade textual; julgamento de especialistas; legibilidade; textos escolares; linguística computacional.

Abstract

Reading comprehension in school contexts is influenced not only by the reader's characteristics but also by the linguistic features of the texts. Within this framework, textual complexity has become a relevant research focus, particularly for improving the criteria used to select educational materials. This study analyzes the relationship between teachers' judgments regarding the complexity of school science texts and the automated linguistic indices generated by TRUNAJOD, a computational tool designed to assess readability in Spanish.

A quantitative, non-experimental, correlational design was used. Reading and writing specialists participated by evaluating a corpus of 80 science texts extracted from secondary school textbooks. The texts were grouped by educational level, and the data were analyzed using hierarchical regressions, Student's t-tests, and Kendall's coefficient of concordance. Results indicate that the most predictive indices of perceived textual complexity are lexical (familiarity, frequency, diversity, imageability) and cohesive (verb overlap, causal markers). Furthermore, variation in evaluation criteria was observed by educational level, with broader linguistic attention in lower grades and a narrower focus on lexical elements in higher grades. It is concluded that teachers' judgments reflect sensitivity to the linguistic properties of texts, although mediated by their disciplinary background. This study provides evidence for the development of hybrid tools for textual evaluation and encourages further research using expanded corpora and evaluators from diverse subject areas.

Keywords: Textual complexity; expert judgment; readability; school texts; computational linguistics.

Recibido 11/6/25

Aceptado 15/7/25

1. Introducción

La complejidad textual es un factor relevante para la implementación de actividades didácticas orientadas al desarrollo de las habilidades de comunicación escrita, especialmente en niveles de enseñanza media donde se espera una evolución progresiva de habilidades de lectura crítica y escritura académica (Campos *et al.*, 2014). Estudios recientes han evidenciado que muchos de los textos incluidos en libros escolares no se ajustan al nivel lector de los estudiantes, lo que genera barreras de acceso a los géneros del conocimiento y afecta la equidad educativa (Rojas *et al.*, 2020).

En este contexto, comprender y caracterizar la complejidad textual desde una perspectiva lingüística resulta crucial para contribuir al desarrollo de la comprensión lectora en todas las áreas del currículum a través del uso de textos adecuados a cada nivel educativo.

Estudios previos indican que los textos escolares no solo deben evaluarse por su contenido, sino también por la carga lingüística que implican. Rasgos como la diversidad clausal, el desarrollo sintáctico, la concreción léxica y la cohesión referencial han sido reconocidos como dimensiones relevantes para estimar la dificultad de los textos desde una perspectiva psicolingüística (Aravena y Quiroga, 2023).

Desde la psicolingüística, la comprensión lectora se entiende como una tarea cognitiva compleja que implica la construcción e integración de representaciones mentales del texto (Kintsch, 1998(a); McNamara, 2004; Fuentes *et al.*, 2021). Si bien los lectores juegan un rol activo en la interpretación, el texto proporciona restricciones y guías para dicha comprensión. Pese a su relevancia, las características textuales han sido menos exploradas sistemáticamente en el campo educativo, lo que ha derivado en la ausencia de criterios objetivos para seleccionar textos según nivel lector (Ortúzar, 2011). En respuesta a este desafío, el procesamiento automático del lenguaje y la lingüística computacional han permitido la creación de herramientas capaces de cuantificar rasgos del texto asociados a su complejidad. Entre ellas, destaca TRUNAJOD, software desarrollado en Chile, que permite evaluar textos escolares en español mediante 46 índices lingüísticos, distribuidos en dimensiones léxicas, sintácticas, semánticas y discursivas (Palma *et al.*, 2021). Esta herramienta permite capturar aspectos como densidad léxica, diversidad, cohesión referencial, narratividad y frecuencia léxica, aportando una medición más robusta que las fórmulas tradicionales.

En el contexto educativo, los juicios docentes constituyen una fuente válida para estimar la dificultad de un texto, ya que integran experiencia profesional y conocimiento disciplinar. No obstante, su carácter subjetivo y variable exige ser complementado con mediciones estandarizadas. Por ello, este estudio se propuso identificar qué índices lingüísticos, generados automáticamente por TRUNAJOD, predicen la percepción de complejidad textual por parte de docentes especialistas en lectura y escritura, a partir de un corpus de textos escolares explicativos. Con ello, se pretende aportar evidencia empírica sobre la relación entre características lingüísticas objetivas de los textos y el juicio experto de docentes especialistas en lectura y escritura.

2. Lecturabilidad y complejidad textual

La lecturabilidad se refiere al grado de comprensibilidad de un texto determinado por sus características lingüísticas (Klare, 1963), y ha sido conceptualizada como la facilidad o dificultad que un texto presenta para ser comprendido por un lector promedio (Crossley *et al.*, 2011). Este concepto ha cobrado especial relevancia en investigaciones que vinculan el análisis textual con el proceso de comprensión lectora, particularmente en el ámbito educativo (Parodi, 1986; Dubay, 2004).

Tradicionalmente, la lecturabilidad se ha estimado mediante fórmulas matemáticas basadas en variables superficiales como la longitud de palabra y oración, como es el caso del índice de Flesch (1948).

No obstante, estas métricas han sido criticadas por su escaso fundamento en teorías cognitivas de la comprensión lectora (McNamara y Graesser, 2011). En respuesta, herramientas actuales basadas en procesamiento del lenguaje natural (PLN) como *Coh-Metrix*, *TAACO* y *TERA* han ampliado el análisis hacia niveles más profundos del texto, considerando dimensiones como cohesión, narratividad, y relaciones semánticas (Jackson *et al.*, 2016; Crossley *et al.*, 2015).

En el contexto hispanohablante, una contribución destacada al análisis de la complejidad textual es el desarrollo de TRUNAJOD, un software diseñado para evaluar textos escolares en español mediante 46 índices distribuidos en cinco dimensiones: léxico, sintaxis, cohesión referencial, cohesión profunda y género discursivo (Palma *et al.*, 2021). Esta herramienta permite estimar el nivel escolar más adecuado para un texto, integrando tanto métricas tradicionales como indicadores alineados con los procesos cognitivos implicados en la lectura. En el plano léxico, se consideran variables como la diversidad, la densidad y la frecuencia léxica, fundamentales para caracterizar la carga informativa y la accesibilidad del vocabulario (Castillo Fadic y Sologuren Insúa, 2020). En el plano sintáctico, se incluyen medidas como la longitud oracional y el número de cláusulas subordinadas, aspectos clave para estimar la complejidad estructural del texto y su impacto en la memoria de trabajo del lector (Véliz, 1999; Zhang *et al.*, 2013). En conjunto, estos índices permiten observar cómo la configuración lingüística del texto incide en su comprensibilidad, proporcionando una base empírica sólida para su evaluación objetiva. Como advierten Perfetti *et al.* (2005), tanto la carga informativa como la complejidad léxica influyen directamente en la dificultad que perciben los lectores al enfrentar un texto.

A partir de lo anterior, se establece la necesidad de analizar cómo los rasgos lingüísticos de los textos inciden en la percepción de su complejidad por parte de docentes especializados en lectura y escritura. En este sentido, el presente estudio tiene un carácter exploratorio y busca identificar qué características textuales predicen los juicios de complejidad emitidos por profesores de enseñanza media, considerando además posibles diferencias según el nivel escolar. Esta investigación se inscribe, por tanto, en una línea de trabajo que articula análisis lingüístico automatizado con la interpretación experta de profesionales del ámbito educativo, con el fin de avanzar hacia una comprensión más precisa y fundamentada de la complejidad textual en contextos escolares.

3. Juicio de expertos en la validación de la complejidad textual

En el ámbito educativo y científico, el juicio de expertos constituye una metodología ampliamente reconocida para evaluar materiales, instrumentos o dimensiones complejas que requieren interpretación cualitativa fundamentada (Cabrero y Llorente, 2013; Pérez-Iribar *et al.*, 2017). Esta técnica se basa en la selección de personas con un conocimiento profundo y especializado sobre el objeto de estudio, y ofrece un marco metodológico riguroso para validar

contenidos y procesos, siendo especialmente útil en estudios exploratorios que requieren captar percepciones informadas sobre fenómenos complejos (Marín-González *et al.*, 2021).

En el caso de la evaluación de la complejidad textual, los docentes especialistas en lectura y escritura representan una fuente experta altamente calificada, ya que su experiencia pedagógica les permite integrar elementos disciplinares y didácticos en la valoración de la complejidad textual. No obstante, dada la naturaleza subjetiva de sus juicios, se requiere complementarlos con herramientas objetivas y sistemáticas que aporten mediciones confiables sobre los rasgos lingüísticos de los textos, tal como se plantea en el presente estudio.

4. Metodología

4.1. Diseño

Se desarrolló un estudio de alcance exploratorio, cuyo propósito fue identificar los rasgos lingüísticos que inciden en la percepción de complejidad textual. El diseño metodológico adoptado fue cuantitativo, no experimental, de tipo correlacional (Hernández Sampieri *et al.*, 2014).

Se analizó la relación entre índices lingüísticos de lecturabilidad, obtenidos mediante una herramienta de procesamiento automático, y los juicios de complejidad textual emitidos por docentes del área de Lenguaje y Comunicación.

4.2. Corpus

Se utilizó un corpus compuesto por 80 textos del área de ciencias, extraídos de libros escolares vigentes en el sistema educativo chileno para los niveles de 1º a 4º medio, grupo en el cual se encuentran estudiantes que tienen entre 14 a 18 años aproximadamente. Los textos se organizaron en dos grupos de acuerdo con el nivel educativo: un conjunto de 40 textos correspondientes a 1º y 2º medio, y otro de 40 textos correspondientes a 3º y 4º medio.

Los textos seleccionados tienen una extensión aproximada de entre 250 y 350 palabras y abordan temáticas propias del currículo nacional, como la célula, los nutrientes o los beneficios de las proteínas, entre otros.

Los textos científicos son altamente pertinentes para el currículo escolar, ya que en la enseñanza media los estudiantes se enfrentan regularmente a materiales de esta índole, caracterizados por una alta densidad informativa y un vocabulario técnico especializado (Ochoa Sierra, 2021). Comprender estos textos exige habilidades lectoras específicas, como la activación de conocimientos previos, la precisión interpretativa y el manejo de una baja ambigüedad semántica (Muñoz Calvo *et al.*, 2013).

Por estas razones, este tipo de material resulta especialmente adecuado para estudios que buscan explorar la relación entre rasgos lingüísticos medidos por índices de lecturabilidad y el juicio experto de docentes especialistas en lectura.

4.3. Participantes

Participaron en el estudio 20 docentes de enseñanza media en ejercicio, con especialidad en el área de Lenguaje y Comunicación. Los participantes fueron asignados a uno de los grupos de textos (10 docentes por grupo), según su experiencia en los niveles educativos.

El rol que asumió cada docente fue de experto en lectura y escritura, lo que permitió que la percepción de complejidad textual se fundamentara a partir de rasgos lingüístico-discursivos.

4.4. Instrumentos

TRUNAJOD

Corresponde a una herramienta de análisis automático de textos para el sistema escolar chileno. Este software permite analizar la complejidad textual, a través de la identificación de rasgos lingüísticos y discursivos que inciden en el acceso al significado. Los resultados se agrupan en cinco índices de lecturabilidad: cohesión referencial, cohesión profunda, narratividad, sintaxis y léxico (Palma *et al.*, 2021). A continuación, se presenta una breve descripción de estos índices:

- A) Cohesión referencial: da cuenta de la repetición o solapamiento de las ideas en el texto, favoreciendo la conexión entre oraciones. Este tipo de cohesión se logra principalmente a través de mecanismos léxicos y se evidencia a través de los subíndices de solapamiento léxico, reiteración de sustantivos y uso de verbos sinónimos.
- B) Cohesión profunda: da cuenta de la relación entre las ideas dentro del texto a través del uso de conectores que establecen distintas categorías de conexión (conexiones temporales, causales, aditivas, lógicas y adversativas). Este tipo de cohesión considera expresiones que indican causa, tiempo, adición, contraste o lógica, permitiendo que la información presentada se entrelace de manera estructurada. Un texto con buena cohesión profunda guía al lector a través de una progresión coherente de eventos, argumentos o ideas.
- C) Narratividad: evalúa la presencia de elementos propios de textos narrativos, como la organización en torno a personajes, acciones, emociones y relaciones causales. Este índice considera que los textos con estructura narrativa son más fáciles de procesar, comprender y recordar, a diferencia de los textos informativos. Se analizan aspectos como el uso de verbos que indican acciones o pensamientos, así como la densidad y longitud de las cláusulas que configuran la historia. TRUNAJOD proporciona los subíndices de densidad de cláusula, densidad de promedio, longitud de cláusula, densidad de pronombres, entre otros.
- D) Sintaxis: da cuenta de la complejidad estructural de las oraciones, considerando elementos como la longitud de cláusulas y oraciones, y la presencia de subordinación. Una sintaxis más elaborada impone mayores

demandas a la memoria de trabajo del lector, lo que puede dificultar la comprensión. TRUNAJOD proporciona como subíndices promedio longitud de palabras, promedio longitud oracional, promedio longitud de cláusula, densidad de cláusula, palabras antes de la raíz, densidad de verbos.

- E) Léxico: analiza la diversidad, densidad y complejidad del vocabulario empleado en el texto. Evalúa, por ejemplo, la proporción de palabras abstractas frente a concretas, así como la variedad y frecuencia léxica. Un léxico más variado y menos frecuente suele indicar mayor complejidad, ya que exige al lector un mayor esfuerzo para procesar y comprender el significado de las palabras. TRUNAJOD proporciona los subíndices de densidad léxica, diversidad léxica, frecuencia de la palabra y contenido de la palabra.

Escala de percepción de complejidad textual

Para evaluar la percepción de los participantes sobre la complejidad textual, se empleó un instrumento de respuesta cerrada con cinco niveles ordinales de dificultad percibida: muy fácil, fácil, medio, difícil, muy difícil. Además, el instrumento recogió información cualitativa sobre los rasgos del texto que fundamentan la clasificación.

4.5. Procedimiento

Los textos se seleccionaron y organizaron en dos grupos en función de los niveles educativos. A continuación, se realizó un análisis automático de lecturabilidad de cada texto, empleando TRUNAJOD. La aplicación de la escala de percepción de complejidad textual se realizó de forma presencial, en una sesión de trabajo por docente.

4.6. Análisis de datos

Análisis preliminar

Dado que TRUNAJOD genera una gran cantidad de índices lingüísticos, se realizó un análisis exploratorio previo para reducir la multicolinealidad. Se examinaron las correlaciones bivariadas entre predictores, y se eliminaron los índices con alta redundancia, priorizando aquellos que representaban constructos distintos (criterio teórico) y relevantes para los objetivos del estudio (criterio metodológico). Posteriormente, se verificó el cumplimiento del supuesto de normalidad univariada mediante el análisis de los coeficientes de asimetría y curtosis, siguiendo las recomendaciones de Tabachnick y Fidell (2013). Todas las variables consideradas se encontraban dentro de los rangos aceptables, lo que permitió continuar con pruebas estadísticas paramétricas.

Para la detección de valores atípicos en los juicios de complejidad textual emitidos por los docentes, se utilizó el gráfico de caja y bigotes. No se identificaron observaciones extremas en las clasificaciones de complejidad realizadas por los jueces, por lo que se conservaron las 80 observaciones. Finalmente, se corroboró que no existieran valores perdidos en el conjunto de datos y se normalizaron los datos correspondientes a los índices lingüísticos.

Análisis principal

Se calculó el coeficiente de concordancia de Kendall para evaluar el grado de acuerdo entre los docentes en los juicios de complejidad textual. Posteriormente, se realizó un análisis de regresión jerárquica para identificar los índices lingüísticos que predicen los juicios de complejidad textual. Finalmente, para comparar los índices lingüísticos entre los dos grupos de textos, se empleó la prueba t de Student.

5. Resultados

El coeficiente de concordancia de Kendall fue $W = 0.52$, lo que indica que existe un nivel de concordancia moderada entre los docentes al clasificar la complejidad de los textos. Se realizó un análisis de regresión jerárquica para identificar qué índices predicen los juicios docentes sobre complejidad textual. El modelo fue estadísticamente significativo, $F(13, 66) = 10.69$, $p < .001$, y explicó un 61.5 % de la varianza en la percepción de complejidad textual (R^2 ajustado = .615). Los predictores significativos incluyeron índices léxicos como la familiaridad ($\beta = 0.186$, $p = .003$), la frecuencia ($\beta = 0.152$, $p = .008$), la imaginabilidad ($\beta = -0.347$, $p < .001$) y la diversidad de lemas ($\beta = 0.174$, $p = .003$). Asimismo, se observaron efectos significativos de índices asociados a la cohesión, como el solapamiento de verbos ($\beta = 0.176$, $p = .010$) y los conectores causales ($\beta = -0.158$, $p = .023$), junto con la proporción de palabras funcionales ($\beta = -0.128$, $p = .038$) y la diversidad de pronombres ($\beta = -0.203$, $p = .001$). Estos resultados sugieren que los docentes fundamentan sus juicios de complejidad principalmente en propiedades léxicas y cohesivas del texto (Tabla 1).

Índice de lecturabilidad	M	DE	β	Error estándar	t	p
Tipo de lema	0.533	0.050	0.174	0.056	3.138	0.003**
Familiaridad léxica	5.575	0.211	0.186	0.061	3.080	0.003**
Frecuencia léxica	0.068	0.314	0.152	0.055	2.740	0.008**
Solapamiento de verbos	0.060	0.090	0.176	0.066	2.661	0.010*
Densidad léxica	0.543	0.022	0.143	0.062	2.327	0.023*
Densidad de adjetivos	0.102	0.022	0.114	0.074	1.537	0.129
Diversidad gramatical en la oración	0.424	0.058	-0.066	0.066	-1.003	0.319
Palabras antes de la raíz	6.515	2.012	-0.069	0.059	-1.172	0.245
Diversidad de adverbios	0.839	0.161	-0.104	0.055	-1.894	0.063
Palabras funcionales	2.729	0.617	-0.128	0.061	-2.115	0.038*
Conectores causales	0.496	0.243	-0.158	0.068	-2.327	0.023*
Diversidad de pronombres	0.399	0.124	-0.203	0.059	-3.429	0.001**
Imaginabilidad	4.089	0.302	-0.347	0.077	-4.505	0.000**

Tabla 1: Resultados del modelo de regresión entre variables de lecturabilidad y la percepción de complejidad textual

Nota. $N = 80$. β = coeficiente estandarizado. DE = desviación estándar.
 $p < .05^*$, $p < .01^{**}$

Al analizar el conjunto de textos correspondiente a los niveles de 1º y 2º medio, el análisis de regresión evidenció un modelo estadísticamente significativo, compuesto por cinco índices ($F(5, 34) = 6.72, p < 0.0001$), el cual explica el 42 % de la varianza en los juicios de complejidad textual (R^2 ajustado = 0.423).

Estos resultados indican que, en conjunto, los predictores incluidos en el modelo —familiaridad léxica, densidad léxica, superposición de sinónimos verbales, diversidad gramatical en la oración y densidad de pronombres— contribuyen significativamente a la evaluación que los docentes realizan sobre la complejidad de los textos escolares (véase Tabla 2).

Índice de lecturabilidad	M	DE	β	Error estándar	t	p
Familiaridad léxica	5.569	0.240	-0.323	0.073	-4.41	< 0.001**
Densidad léxica	0.543	0.022	0.073	0.076	0.95	0.347
Superposición de sinónimos verbales	0.544	0.024	0.212	0.086	2.46	0.019*
Diversidad gramatical en la oración	0.422	0.057	0.157	0.083	1.88	0.068
Densidad de pronombres	0.016	0.009	0.105	0.079	1.32	0.196

Tabla 2: *Textos 1º–2º medio: modelo de regresión entre variables de lecturabilidad y percepción de complejidad textual.*

Nota. $N = 40$. β = coeficiente estandarizado de regresión. DE = desviación estándar.
 $p < .05^*$, $p < .001^{**}$

En el conjunto de textos correspondiente a 3º y 4º medio, el análisis de regresión evidenció un modelo estadísticamente significativo, $F(3, 36) = 14.03, p < .001$, y explicó el 50.1% de la varianza en los juicios (R^2 ajustado = .501). Los tres índices lingüísticos con mayor incidencia en la percepción de dificultad fueron: diversidad léxica lematizada, familiaridad de las palabras y frecuencia léxica, todos con un nivel de significancia estadística similar ($p < 0.01$) (Tabla 3).

Índices de lecturabilidad	M	DE	β	Error estándar	t	p
Diversidad de lemas	0.521	0.049	-0.333	0.090	-3.691	0.001*
Familiaridad léxica	5.582	0.181	-0.297	0.103	-2.871	0.007*
Frecuencia léxica	0.094	0.291	-0.287	0.095	-3.019	0.005*

Tabla 3: *Textos 3º–4º medio: modelo de regresión entre variables de lecturabilidad y percepción de complejidad textual.*

Nota. $N = 40$. β = coeficiente estandarizado de regresión. DE = desviación estándar.
 $p < .05^*$

Los resultados obtenidos a partir de los modelos de regresión demuestran la relevancia de los rasgos léxicos en la percepción de complejidad textual. El análisis comparativo de los índices léxicos mediante la prueba t de Student reveló diferencias estadísticamente significativas entre los textos destinados a 1º–2º y 3º–4º medio en cuatro dimensiones (Tabla 4): diversidad de lemas

($t(38) = 2.15, p = .034$), diversidad de adjetivos ($t(38) = 2.30, p = .024$), diversidad de palabras funcionales ($t(38) = 1.99, p = .006$) y concreción léxica ($t(38) = 3.59, p < .001$). Estos resultados sugieren que los textos comparados difieren en aspectos léxicos y gramaticales, los cuales podrían incidir en su complejidad percibida.

Índice de lecturabilidad	$t(38)$	p (bilateral)
Diversidad de lemas	2.148	0.034
Diversidad de adjetivos	2.297	0.024
Palabras funcionales	1.990	0.006
Concreción léxica	3.591	< .001

Tabla 4: *Análisis comparativo de índices léxicos entre grupos escolares*
Nota. Prueba t bilateral para muestras independientes. $gl = 38$.

6. Discusión

Los resultados de este estudio muestran que ciertos rasgos lingüísticos medidos por herramientas automatizadas de análisis textual, como TRUNAJOD, inciden significativamente en los juicios de complejidad textual emitidos por docentes especialistas en lectura y escritura. El modelo general de regresión evidenció que los índices más relevantes para predecir el juicio de complejidad fueron léxicos (familiaridad, frecuencia, imaginabilidad, diversidad lematizada), lo que concuerda con estudios previos que destacan la centralidad del léxico en los procesos de reconocimiento y comprensión (Graesser *et al.*, 2004; Castillo Fadic y Sologuren Insúa, 2020; Crossley *et al.*, 2015).

A partir de la centralidad asignada al acceso al léxico y la carga informativa en la construcción de la representación mental del texto (Kintsch, 1998(b)), el hallazgo de que la familiaridad léxica y la diversidad de lemas funcionen como predictores en ambos grupos de textos escolares, podría considerarse una evidencia de que los docentes especialistas en lectura y escritura reconocen la complejidad semántica de los textos a partir de índices léxicos.

Asimismo, los resultados evidenciaron diferencias en los patrones de evaluación de los grupos de textos. En el grupo de 1º–2º medio, además del léxico, se detectaron indicadores de cohesión referencial (solapamiento de verbos) y estructuras sintácticas (disimilaridad de categorías gramaticales), lo cual sugiere una atención más amplia a distintos niveles del texto. Esta tendencia podría explicarse por el hecho de que los docentes perciben que los estudiantes en etapas iniciales requieren apoyos de cohesión más evidentes y estructuras sintácticas menos exigentes (McNamara y Kintsch, 1996). Por el contrario, en el grupo de 3º–4º medio, los juicios se concentraron exclusivamente en aspectos léxicos, posiblemente por la densidad conceptual de los textos, a partir de lo cual es posible inferir que la percepción de los docentes de la complejidad textual se centra en la accesibilidad terminológica de los textos.

Cabe destacar que el coeficiente de concordancia de Kendall ($W = 0.52$) sugiere un acuerdo moderado entre evaluadores, lo que resulta aceptable para

estudios exploratorios que implican evaluaciones subjetivas (Marín-González *et al.*, 2021; Pérez-Iribar *et al.*, 2017). No obstante, se evidencia la necesidad de complementar el juicio de expertos como estrategia metodológica con mediciones objetivas y sistemáticas que permitan, en este caso, objetivar los rasgos lingüísticos evaluados.

En conjunto, los resultados aportan evidencia empírica sobre la relación entre características lingüísticas de los textos y la percepción docente de complejidad textual, lo cual puede ser útil tanto para la selección informada de materiales escolares como para el diseño de instrumentos de evaluación y adaptación de textos.

Además, el análisis comparativo entre niveles escolares muestra que la complejidad léxica no solo varía entre textos, sino también en cómo es interpretada por los evaluadores en función del contexto curricular.

7. Conclusión

La investigación reportada en este artículo permitió determinar que los índices de lecturabilidad, obtenidos mediante análisis automatizado con TRUNAJOD, inciden significativamente en los juicios que docentes especialistas en lectura y escritura realizan sobre la complejidad textual de textos del área de ciencias, provenientes de textos escolares. En particular, se constató que las dimensiones léxicas y de cohesión son las que ejercen mayor influencia, lo que coincide con hallazgos previos en el ámbito de la psicolingüística y la lingüística computacional.

Asimismo, se evidenció que la percepción de complejidad textual varía según el nivel educativo al que está dirigido el texto. En los niveles de 1º a 2º medio, los docentes consideraron una gama más amplia de dimensiones lingüísticas, mientras que en 3º a 4º medio su evaluación se centró casi exclusivamente en el léxico. Esta diferencia podría estar mediada por el grado de conocimiento disciplinar de los evaluadores, lo que refuerza la importancia de considerar el perfil de los jueces en este tipo de estudios.

La concordancia moderada obtenida entre los evaluadores ($W = 0.52$) valida el uso del juicio experto como herramienta complementaria en investigaciones que buscan comprender cómo se percibe y evalúa la dificultad de los textos desde una perspectiva docente. Los hallazgos de esta investigación no solo aportan evidencia empírica relevante para la validación de herramientas de análisis textual, sino que también pueden guiar el diseño de criterios pedagógicos para la selección de materiales escolares pertinentes para cada nivel educativo.

Entre las principales limitaciones del estudio se encuentra el tamaño del corpus textual y el hecho de que los docentes participantes no eran especialistas en el área de ciencias, lo que podría haber condicionado su foco evaluativo en aspectos léxicos por sobre otros niveles discursivos. Además, el uso de juicios individuales, aunque valioso en la fase exploratoria, podría beneficiarse de futuros diseños que incorporen técnicas de consenso. Futuros estudios podrían ampliar el corpus, incorporar otros géneros discursivos o comparar los juicios docentes con datos empíricos de comprensión lectora en estudiantes, con el fin

de seguir profundizando en las relaciones entre características textuales objetivas, percepción experta y desempeño lector real.

Referencias bibliográficas

- Aravena, María y Riva Quiroga. 2023. ¿Cómo crece la sintaxis en los años escolares? Un estudio de la frase nominal en textos explicativos escritos por estudiantes secundarios. *Logos: Revista de Lingüística, Filosofía y Literatura*, 33, 1, 20-39 [en línea]. Disponible en: <https://doi.org/10.15443/RL3302>
- Cabrero, Julio y María del Carmen Llorente. 2013. La aplicación del juicio de experto como técnica de evaluación de las tecnologías de la información (TIC), *Revista de Tecnología de Información y Comunicación en Educación*, 7, 2: 11-22 [en línea]. Disponible en: <https://revistaeduweb.org/index.php/eduweb/article/view/206>
- Campos, Daniela, Paula Contreras, Bernardo Riff, Mónica Véliz y Alejandro Reyes. 2014. Complejidad textual, lecturabilidad y rendimiento lector en una prueba de comprensión en escolares adolescentes. *Universitas Psychologica*, 13, 3 [en línea]. Disponible en: <https://revistas.javeriana.edu.co/index.php/revPsycho/article/view/4803>
- Castillo Fadić, María Natalia y Enrique Sologuren Insúa. 2020. Léxico frecuente, riqueza léxica y estereotipos sobre la lectura de profesores en formación. *Logos: Revista De Lingüística, Filosofía y Literatura*, 30, 1, 69-85 [en línea]. Disponible en: <https://doi.org/10.15443/RL3006>
- Crossley, Scott A., David B. Allen y Danielle S. McNamara. 2011. Text readability and intuitive simplification: A comparison of readability formulas. *Reading in a Foreign Language*, 23, 1, 84-102.
- Crossley, Scott A., Kristopher Kyle y Danielle S. McNamara. 2015. The tool for the automatic analysis of text cohesion (TAACO): Automatic assessment of local, global, and text cohesion. *Behavior Research Methods*, 48, 1227-1237 [en línea]. Disponible en: <https://link.springer.com/journal/13428>
- Dubay, William H. 2004. *The Principles of Readability*. CA: Plain-Language Services.
- Flesch, Rudolph. 1948. A new readability yardstick. *Journal of Applied Psychology*, 32, 221-233.
- Fuentes Román, Ivonne, Elizabeth Donoso Osorio y Carolina Iturra Herrera. 2021. Actividades e interacciones en clases para promover la comprensión lectora. *Lingüística*, 37, 1, 9-27 [en línea]. Disponible en: <https://doi.org/10.5935/2079-312x.20210002>
- Graesser, Arthur, Danielle McNamara, Max Louwerse y Zhiqiang Cai. 2004. Coh-Metrix: Analysis of test on cohesion and language. *Behavior Research Methods*, 36, 193-202 [en línea]. Disponible en: <https://link.springer.com/article/10.3758/BF03195564>
- Hernández Sampieri, Roberto, Carlos Fernández Collado y Pilar Baptista Lucio. 2014. *Metodología de la Investigación* (6 ed., 88-101). México: McGraw-Hill.

- Jackson, G. Tanner, Laura K. Allen y Danielle S. McNamara. 2016. Common Core TERA: Text Ease and Readability Assessor, en Scott A. Crossley y Danielle S. McNamara (eds.), *Adaptive Educational Technologies for Literacy Instruction*. Taylor & Francis, Routledge: NY.
- Kintsch, Walter. 1988(a). The role of knowledge in discourse comprehension: A construction-integration model. *Psychological Review*, 95, 2, 163-182.
- Kintsch, Walter. 1998(b). *Comprehension: A paradigm for cognition*. Cambridge University Press.
- Klare, George R. 1963. The measurement of readability. Ames, Iowa: Iowa State University Press.
- Marín-González, Freddy, Judith Pérez-González, Alexa Senior-Naveda y Jesús García-Guliany. 2021. Validación del diseño de una red de cooperación científico-tecnológica utilizando el coeficiente K para la selección de expertos. *Información Tecnológica*, 32, 2, 79-88 [en línea]. Disponible en: <https://doi.org/10.4067/S0718-07642021000200079>
- Muñoz Calvo, Elena, Lilia Muñoz Muñoz, Mercedes García González y Luis Granado Labrada. 2013. La comprensión lectorade textos científicos en el proceso de enseñanza-aprendizaje. *Humanidades Médicas*, 13, 3, 772-804.
- McNamara, Danielle S. y Walter Kintsch. 1996. Learning from texts: Effects of prior knowledge and text coherence. *Discourse Processes*, 22, 247-288.
- McNamara, Danielle S. 2004. Aprender del texto: Efectos de la estructura textual y las estrategias del lector. *Revista Signos*, 37, 55, 19-30 [en línea]. Disponible en: <https://dx.doi.org/10.4067/S0718-09342004005500002>
- McNamara, Danielle S. y Arthur Graesser. 2011. Coh-Metrix: An automated tool for theoretical and applied natural language processing. *Applied natural language processing and content analysis: Identification, investigation, and resolution* [en línea]. Disponible en: <http://dx.doi.org/10.4018/978-1-60960-741-8.ch011>
- Ochoa Sierra, Ligia. 2021. Complejidad estructural en textos escolares descriptivoexpresivos: estudio de corpus. *Folios*, 54 [en línea]. Disponible en: <https://doi.org/10.17227/folios.54-11977>
- Ortúzar, Pablo. 2011. *Informe textos escolares: Calidad, formato y mercado de los textos escolares en Chile (IES)*. Santiago: IES. [en línea]. Disponible en: <https://es.scribd.com/document/354406925/Informe-Textos-escolares-pdf>
- Palma, Diego, Christian Soto, Monica Veliz, Bruno Karelovic y Bernardo Riffo. 2021. TRUNAJOD: A text complexity library to enhance natural language processing. *Journal of Open Source Software*, 6, 60, 3153 [en línea]. Disponible en: <https://joss.theoj.org/papers/10.21105/joss.03153>
- Parodi, Giovanni. 1986. Validación y desarrollo del gráfico de Fry para el castellano hasta el nivel universitario. *Signos*, 19, 24, 125-132.
- Pérez-Iribar, Giorver, Maylin Beleño-Fuentes, Carlos Nuñez-Peña, Marcos Orquera-Cadena. 2017. Revisión. Valoración del Resultado Científico de la Investigación. Una Experiencia desde la Aplicación del Criterio de Experto, *OLIMPIA, Revista de la Facultad de Cultura Física de la Universidad de Granma*, 14, 46, 154-168.

- Perfetti, Charles, Nicole Landi y Jane Oakhill. 2005. The Acquisition of Reading Comprehension Skill, en M. Snowling & Ch. Hulme (eds.), *The Science of Reading*, Oxford: Blackwell, 227-247.
- Rojas Villarroel, Daniel, Romualdo Ibáñez Orellana, Fernando Moncada Nahuelquín y Andrea Santana Covarrubias. 2020. Los géneros del conocimiento en el texto escolar de Lenguaje y Comunicación: Un análisis semiautomático de su lecturabilidad. *RLA. Revista de Lingüística Teórica y Aplicada*, 58, 2, 41-67 [en línea]. Disponible en: <https://doi.org/10.29393/RLA58-14GCDR40014>
- Tabachnick, Barbara y Linda S. Fidell. 2013. *Using Multivariate Statistics* (6th ed.). Boston, MA Pearson.
- Véliz, Mónica. 1999. Complejidad sintáctica y modo del discurso. *Estudios Filológicos*, 34, 181-192.
- Zhang, Lixiao, Zaiying Liu y Jun Ni. 2013. Feature-Based Assessment of Text Readability. *2013 Seventh International Conference on Internet Computing for Engineering and Science* [en línea]. Disponible en: <https://dl.acm.org/doi/10.1109/ICICSE.2013.18>

Nota

La revisión teórica, descripción metodológica, discusión y revisión general del artículo ha estado a cargo de Mg. Nathaly Reyes Peralta y Dra. Pamela Ramírez Peña.

El análisis de datos y resultados a cargo de Mg. Diego Palma Sánchez. La supervisión del estudio y revisión del artículo ha estado a cargo del Dr. Christian Soto Fajardo.

Agradecimientos:

La autora, Mg. Nathaly Reyes Peralta, agradece a ANID por el financiamiento recibido a través de la Beca de Doctorado Nacional nº 21241404.

La autora, Dra. Pamela Ramírez-Peña, agradece a ANID Proyecto FONDECYT Postdoctorado nº 3230193 por el apoyo a la investigación.

Disponibilidad de los datos

Los datos de esta investigación se encuentran disponibles, previa solicitud a los autores dado que el estudio fue realizado en el marco del proyecto Fondef IT17I0051.

Nota de aceptación

Este texto ha sido aceptado para publicación por el único Director-Editor de la revista, Adolfo Elizaincín, quien ha actuado de acuerdo a lo establecido en la "Declaración de comportamiento ético" de la revista *Lingüística* https://www.mundoalfal.org/sites/default/files/revista/Declaracion_comp_etico.pdf, primer párrafo del capítulo "Obligaciones del Director-Editor". A esta declaración deben adherir, explícitamente, el Director-Editor, los árbitros y los autores.