

# Miradas interdisciplinarias entre **lengua, lingüística y traducción**

Ioana Cornea  
Noëlle Groult Bois  
V́ctor Mart́nez de Badereau  
coordinadores



Universidad Nacional  
Aut3noma de M3xico

La presente obra está bajo una licencia de:  
<https://creativecommons.org/licenses/by-nc-sa/4.0/deed.es>



## Atribución-NoComercial-CompartirIgual 4.0 Internacional (CC BY-NC-SA 4.0)

This is a human-readable summary of (and not a substitute for) the [license](#). [Advertencia](#).

### Usted es libre de:

**Compartir** — copiar y redistribuir el material en cualquier medio o formato

**Adaptar** — remezclar, transformar y construir a partir del material

La licenciente no puede revocar estas libertades en tanto usted siga los términos de la licencia

### Bajo los siguientes términos:



**Atribución** — Usted debe dar [crédito de manera adecuada](#), brindar un enlace a la licencia, e [indicar si se han realizado cambios](#). Puede hacerlo en cualquier forma razonable, pero no de forma tal que sugiera que usted o su uso tienen el apoyo de la licenciente.



**NoComercial** — Usted no puede hacer uso del material con [propósitos comerciales](#).



**CompartirIgual** — Si remezcla, transforma o crea a partir del material, debe distribuir su contribución bajo la [misma licencia](#) del original.

Esto es un resumen fácilmente legible del:  
texto legal de la licencia completa

En los casos que sea usada la presente obra, deben respetarse los términos especificados en esta licencia.



# Análisis de lecturabilidad en niños de 2° y 3° de primaria

DIANA YAZMÍN DUEÑAS CHÁVEZ

GEMMA BEL-ENGUIX

Instituto de Ingeniería. Grupo de Ingeniería Lingüística. UNAM

ARTURO CUIEL

Universidad Veracruzana

## Resumen

La lectoescritura es una de las habilidades básicas para la educación. Deficiencias en su adquisición pueden afectar el desarrollo educativo de las personas en el largo plazo. Por este motivo, tener medios para evaluar la proficiencia de los niños durante su proceso de adquisición temprana, es crucial para garantizar una educación de calidad. Este trabajo muestra un estudio comparativo entre la complejidad lectora de textos escritos por adultos y por niños mexicanos, como una forma de evaluar la proficiencia de estos últimos durante la adquisición. Para ello, se usaron tres corpus de referencia. Dos de estos, CEELE y el EXCALE, contienen textos creados por niños de 2° y 3° de primaria. El tercero, *Historias Cortas*, es un corpus de control escrito por adultos. Los textos se organizaron siguiendo la escala de perspicuidad propuesta por Szigriszt-Pazos, la cual clasifica un documento según su nivel de dificultad. Dicha escala toma en cuenta factores tales como el número de frases por párrafo, el número de palabras por frase y la estructura silábica de las mismas. La clasificación de Szigriszt-Pazos permitió realizar un estudio pormenorizado de las estructuras silábicas presentes en cada corpus, así como la comparación directa por nivel de dificultad entre los documentos de ambos grupos. Los resultados indican que los elementos cuantitativos comúnmente considerados por las fórmulas de complejidad lectora no son suficientes para determinar la competencia escrita de los niños; problemas como la hiposegmentación, carencia de puntuación y errores ortográficos aumentan artificialmente la dificultad calculada para los textos infantiles. Asimismo, los resultados indican que el estudio de sílabas específicas puede constituir un mejor indicador de proficiencia escrita que los ya mencionados. En el futuro, estos datos pueden contribuir a la elaboración de una escala de capacidad de escritura, allanando el camino para una evaluación sistematizada del proceso de adquisición.

**Palabra clave:** *índice de perspicuidad, adquisición de la escritura, complejidad escrita, escritura infantil*

**Abstract**

Literacy is fundamental for education. Deficiencies in its acquisition may affect the learners' performance in the long run. Thus, having means to evaluate written proficiency as early as possible may be crucial to guarantee a quality education. This chapter presents a comparative study between Spanish-language texts written by Mexican adults and children, using reading complexity as a means to measure proficiency. For the study, three reference corpora were used. Two of them, CEELE and EXCALE, contain texts created by children enrolled in the 2nd and 3rd grades of Mexican elementary school. The third one, Short Stories, is a control corpus produced by adults. The texts were classified by their reading difficulty level, as given by the Szigriszt-Pazos perspicuity formula. To this end, the equation considers features such as the number of phrases per paragraph, the number of words per phrase and, finally, the number of syllables per word. The Szigriszt-Pazos classification allowed for a detailed analysis of the syllabic structures found in each corpus, enabling a direct per-level comparison. The obtained results indicate that the quantitative elements commonly used to measure readability are not enough to determine writing ability in children: problems such as under-segmentation, lack of punctuation and spurious orthography, artificially increase the difficulty scores calculated for the children productions. Moreover, the results suggest that the study of specific syllables is a better indicator of proficiency than the ones already mentioned. In the future, this information may contribute to the creation of a writability formula, paving the way towards novel strategies for the systematic evaluation of the acquisition process.

**Keywords:** perspicuity index, writing acquisition, written complexity, child writing

## Introducción

La escritura y la lectura son habilidades fundamentales que un niño debe aprender en los primeros años de educación. Ambas se consideran la base para todo conocimiento posterior; por tanto, desa-

rollarlas adecuadamente puede repercutir en cómo los niños comprenden, decodifican y establecen vínculos a largo plazo entre su entorno y lo que pasa en el mundo (Hemphill & Tinan, 2008; Defazio, Jones, Tennant & Hook, 2010).

A pesar de estar fuertemente relacionadas, la investigación moderna ha encontrado que existen diferencias fundamentales entre las habilidades lectora y escritora (Allen, Snow, Crossley, Jackson & McNamara, 2014). En general, la habilidad lectora se refiere a la capacidad de un individuo para comprender un texto escrito por terceros. Por otro lado, la habilidad escritora se refiere a la capacidad de un individuo de producir un texto por sí mismo. Ambas actividades conllevan el desarrollo de habilidades cognitivas tanto de bajo nivel (e.g. memoria de trabajo) como de alto nivel (e.g. inferencias contextuales), así como el manejo de conocimiento común entre autores y lectores potenciales (e.g. contexto, vocabulario, etc.).

La diferencia fundamental entre estas habilidades radica en el proceso de *integración y separación* de la información. Así, la comprensión lectora implica la integración paulatina de la información textual en una representación interna, la cual se va construyendo a medida que se desarrolla la lectura (Kintsch, 1998; Mayers & O'Brien, 1998). Para esto, los modelos existentes proponen que los lectores no solo necesitan tener un conocimiento común con el autor, sino que requieren además tener las habilidades cognitivas que les permitan usar dicho conocimiento en aquella representación. Por otro lado, el proceso de escritura sigue el flujo opuesto: consiste en poder segmentar el conocimiento interno del individuo para expresarlo en el texto, de forma que su organización y distribución permita a los lectores reconstruirlo adecuadamente. Notablemente, el modelo de Hayes-Flowers (Hayes & Flowers, 1980; Hayes, 2012) propone que la escritura es un proceso no lineal, en el cual la interacción entre elementos individuales tales como el escritor, el ambiente de escritura y su capacidad de memoria producirán textos con diferente nivel de



calidad. De este modo, aunque ambos procesos utilizan habilidades cognitivas similares, la lectura requerirá adecuarse a la escritura de un tercero para lograr comprensión, mientras que la escritura exigirá comprensión para lograr la planeación del discurso individual.

En este contexto, Ferreiro (1998) plantea que los niños que cursan 2° y 3° de primaria deben tener una noción clara sobre la lengua escrita para lograr un desarrollo óptimo. Lo anterior, como resultado de una educación recibida en grados escolares anteriores en los que han tenido la oportunidad de practicarla. Sin embargo, la autora también señala que a la escritura se le considera como una actividad superficial en la que los alumnos se concentran en colocar la letra correcta en el lugar correspondiente; es decir, aprenden a transcribir el lenguaje oral en lenguaje escrito, sin importar si se comprende el significado de lo que se dice. Por si misma, esta actividad no es suficiente para desarrollar una capacidad de escritura adecuada (Kellogg, 2008).

Si se contara con una enseñanza sólida se esperaría que, después de cierto tiempo, las personas en edad escolar tuvieran un grado de complejidad escrita de acuerdo con la etapa educativa cursada. Zamudio (2008, pp. 10-19) sostiene que los alumnos, al tener que escribir de manera sistemática consonantes y vocales sin haberse adaptado completamente a la relación entre los símbolos y el sonido, intentan imitar una estructura silábica sencilla que se acerque a la relación de las palabras escuchadas con el patrón gráfico CV; por esta razón, las estructuras dentro de sus redacciones suelen formarse por CV-CV-CV. A los niños les resulta difícil romper esta estructura, aunque muchas veces son capaces de detectar la presencia de más consonantes en sílabas de mayor complejidad e identificar cuáles son sus componentes.

Zamudio (2008) y Tapia (2003) mencionan que, de acuerdo con algunos estudios sobre las estructuras silábicas del español, las más sencillas y comunes serán: V, VC, CV, CVC, CCV y CCVC. Estos patrones deberían aparecer con mayor frecuencia dentro de los escritos

de personas con educación primaria, para ir adquiriendo después, dentro de un entorno educativo ideal, construcciones más complejas como VCC, CVCC y CCVCC.

Dentro de los análisis realizados, la complejidad de un patrón silábico se determinó a partir de la combinación entre elementos C y V, la longitud de esos elementos dentro de la sílaba, el número de sílabas que conforman la palabra y la dificultad que esta estructura tendrá para la comprensión de las palabras.

En América Latina, la lectura y escritura son introducidas de manera simultánea. Se parte de la hipótesis de que cada niño adquirió previamente el sistema fonético de su lengua materna y además aprendió a diferenciar entre un dibujo y las letras de un escrito. Por todo esto se plantea que no deberían existir dificultades para discriminar entre dos formas visuales próximas como p-b-d, 1-l-l, 5-s-2, 3-E o v-u. Tampoco debería causar problema la asimilación de la forma de las letras escritas o trazarlas en un papel, pues en este caso el lenguaje está definido como una simple transcripción de lo sonoro a lo visual (Ferreiro, 1998, pp. 14-20).

Este capítulo se enfoca en el análisis de la complejidad escrita y de los patrones silábicos de alumnos de segundo y tercero de primaria. Se quiso determinar si los niños en esta etapa de escolarización cuentan con el nivel de complejidad y estructura escrita requerida para los años de estudio cursados, además de verificar los factores que ayudaron o afectaron a que sus redacciones tengan una buena legibilidad.

El planteamiento del problema se concretó a partir de los siguientes objetivos: a) determinar si los escritos realizados por niños de 2º y 3º de primaria cumplen con el nivel de perspicuidad adecuado para su nivel escolar; b) analizar los factores relacionados con la legibilidad que afectan o propician que las redacciones de niños sean entendibles; c) analizar y comparar la complejidad de la estructura silábica que emplean los niños de 2º y 3º con los parámetros de perspicuidad del español.

El trabajo se basó en tres corpus: CEELE y EXCALE, que sirven para analizar la escritura de los niños, y el corpus *Historias Cortas* (HC), que se elaboró especialmente para llevar a cabo esta investigación y sirvió como corpus de control para comparar las características de los estudiantes con los parámetros generales del español. Paralelamente, para llevar a cabo los experimentos, se usó “Legible”, una herramienta digital específica para la escala de índice de perspicuidad. Este programa se utiliza para medir la lecturabilidad de los textos, comparar diferentes estructuras silábicas y analizar la complejidad de sus patrones.

El resto del capítulo se estructura de la siguiente manera: la sección 2 explica las distintas escalas para medir la complejidad escrita que han sido creadas o adaptadas al español; la sección 3 explica los corpus en que se basa este estudio; la sección 4 explica los resultados de clasificación del corpus según la escala de perspicuidad; el apartado 5 hace un análisis de los patrones vocálicos encontrados en las fuentes investigadas; finalmente, la sección 6 presenta algunas conclusiones y el trabajo futuro.

## **Escalas de medición para la complejidad escrita**

El análisis cuantitativo del contenido textual es un tema que ha generado interés en todas las épocas. Desde principios del siglo XX, diversos autores han buscado maneras de cuantificar la dificultad de lectura. Un ejemplo de esto es el desarrollo de métricas orientadas a caracterizar cuán difícil será para un lector comprender un texto. Desde el punto de vista de la separación lectura-escritura, la comprensión se ha reducido a encontrar fórmulas que predigan cuán complejo será para un lector promedio (según su nivel de estudios) integrar una representación interna de la información textual. Cabe decir que, a conocimiento de los autores, existe poco



trabajo previo dedicado a la creación de métricas de capacidad escritora, por ejemplo, formulas capaces de caracterizar cuantitativamente la capacidad que tiene un escritor para segmentar y organizar información mediante la expresión escrita, con respecto de su nivel de estudios.

Así, con el objetivo de medir la comprensión textual, Thorndike (1921) tomó el vocabulario como el primer componente dentro de una fórmula para calcular la dificultad lectora. Para ello, el autor elaboró una lista con las 10,000 palabras más usuales en la lengua inglesa; a partir de este trabajo, dos años más tarde, se propone la primera fórmula de legibilidad. Finalmente, hacia 1925 se empezaron a desarrollar experimentos usando pasajes de lectura juzgados de diferente nivel, de modo que se pudiesen identificar variables cada vez más representativas de los mismos.

A partir de estos trabajos seminales se comenzaron a crear nuevos métodos orientados a calcular la comprensión de un texto, el conocimiento de las palabras y el razonamiento de su contenido. Un ejemplo notable son las fórmulas de nivel de facilidad lectora de Flesch (1948), creadas para calcular el nivel de comprensión de redacciones en inglés. En general, este método toma en cuenta variables como la longitud de las frases, el número total de palabras, el número total de oraciones y el número total de sílabas.

Actualmente, la complejidad lectora se refiere al nivel de comprensión en cada redacción para determinado nivel educativo o, en otras palabras, la dificultad que encuentra el lector para introducirse en su discurso informativo y el nivel de claridad que para él tiene el mensaje. En esencia, las fórmulas caracterizan la dificultad con respecto a la población: mientras mayor sea el grado de estudios de la persona, mayor, se asume, la capacidad para decodificar escritos con mayor variedad de palabras o diferente exposición de ideas o temas.

Para aplicar este concepto a cualquier tipo de obra o publicación es necesario contar con fórmulas que se adapten a las reglas de un idioma en específico y ofrezcan un análisis rápido y eficaz. A continuación, presentamos algunas fórmulas para el cálculo de la complejidad textual en español.

### *Índice de perspicuidad de Szigriszt-Pazos*

El índice de Perspicuidad de Szigriszt-Pazos (1993) mide la facilidad lectora de un texto. Se realizó como una adaptación de la ecuación de Flesch para el inglés (1948), que tiene en cuenta las aptitudes o defectos de los adultos para comprender diferentes tipos de lectura.

Esta escala se encarga de medir el grado de comprensibilidad y la facilidad de lectura; además demuestra que la velocidad de la lectura y el número de combinaciones de palabras están condicionados por el grado de legibilidad del texto.

El cómputo de sílabas también es sumamente importante. Para medirlas se toman en cuenta grupos de letras agrupadas fonéticamente que ocupan un espacio dentro de la palabra, su largo determina la dificultad de comprensión porque puede resultar en un uso de palabras menos conocidas.

La fórmula de Szigriszt-Pazos para medir la perspicuidad ( $\mathcal{P}$ ) en español es la siguiente:

$$\mathcal{P} = 206.835 - 62.3 \frac{S}{P} - \frac{P}{F}$$

Donde  $P$  corresponde al número total de palabras en el documento,  $S$  denota el número total de sílabas y  $F$  el número total de oraciones.

La Tabla 1 muestra los puntajes de la escala. Cada nivel específica un tipo de publicación relacionado con una edad o grado de

estudios. Esta propuesta indica qué textos serán fáciles de entender para cada persona dependiendo de su nivel educativo.

**Tabla 1.** Niveles en la escala de perspicuidad ( $\mathcal{P}$ ) (Szigriszt-Pazos).

Nivel de perspicuidad ( $\mathcal{P}$ )			
$\mathcal{P}$	Estilo	Tipo de publicación	Estudios
0-15	Muy difícil	Científica, filosófica	Titulados universitarios
16-35	Árido	Pedagógica técnica	Estudios universitarios
36-50	Bastante difícil	Literatura y divulgación	Cursos secundarios
51-65	Normal	Media	Popular
66-75	Bastante fácil	Novela, revista femenina	12 años
76-85	Fácil	Para quioscos	11 años
86-100	Muy fácil	Cómics y viñetas	6 a 10 años

### *Otras escalas para medir la legibilidad textual en español*

Existen otras escalas diseñadas para medir la legibilidad en español. A continuación, se hace una breve presentación de las principales.

**Lectorabilidad** (Fernández Huerta, 1959). Es una adaptación de la fórmula de Flesch, aunque no considera tamaño de letra, uso de mayúsculas y minúsculas, números, abreviaturas o uso de guiones, aspectos tipográficos que influyen mucho en la facilidad lectora. La fórmula es la siguiente:  $L = 206.84 - 0.60P - 1,02F$ , donde  $P$  es el promedio de sílabas por palabra y  $F$  la media de palabras por frase. Esta métrica no sirve para los objetivos de este trabajo debido a que los niveles de clasificación propuestos toman en cuenta a los estudiantes de 4° de primaria como el nivel más fácil.

**Comprensibilidad** (Gutiérrez Polini, 1972). Esta escala fue propuesta en 1972 por Luisa Elena Gutiérrez Polini. Los datos que se toman para el cálculo son el número de palabras, el número de letras y de frases. Se descartó para nuestro trabajo debido a que está

pensada solamente para analizar textos escolares de niños de 6° de primaria.

**Fórmula de comprensibilidad** (Crawford, 1984). Calcula los años de escolaridad necesarios para entender un texto, pero solo funciona con escritos de niños de primaria. Toma en cuenta el número de años de escolarización, el número de oraciones y el número de sílabas por cada cien palabras. Es difícil de aplicar porque no cuenta con una escala.

**Inflesz** (Barrio Cantalejo, 2008). Se trata de una adaptación de la fórmula de perspicuidad de Szigriszt-Pazos que reacomoda los puntajes de los niveles. Se enfoca en publicaciones comerciales, textos utilizados para la educación primaria y secundaria y revistas científicas del ámbito de la salud.

**Legibilidad** ( $\mu$ ). Fue desarrollada por Miguel Muñoz Baquedano en Chile (2008, 2006). La medida considera el número de palabras, la media y la varianza del número de letras en cada una de ellas. Mientras mayor sea la medida de letras por palabra, el texto resulta más difícil.

## Corpus de estudio

Para calcular la perspicuidad en niños de 2° y 3° de primaria se han utilizado dos corpus de referencia en español mexicano, el CEELE y el EXCALE. Además, hemos elaborado un corpus de control denominado *Historias Cortas* (HC). A continuación, se explican los detalles de cada uno de ellos.

### Corpus CEELE

El CEELE (Corpus Electrónico para el Estudio de la Lengua Escrita) fue recolectado por Díaz Argüero *et al.* (2010). Cuenta con una base

de 300 textos escritos por niños de entre 7 y 8 años que estaban iniciando el segundo año de primaria. Para la recolección del corpus se leyó a los niños una historia como base motivacional para que posteriormente ellos pudieran describir su escuela. Con esta instrucción, los escritos quedaron acotados a poner el nombre, edad, escuela, cómo es su trayecto hacia la escuela, cómo es su escuela y qué hacen ahí durante el día. Esta base de datos es electrónica y se respetó al máximo la escritura original de los niños. Asimismo, la página web ofrece un esquema de consulta donde se puede buscar por palabra, sin embargo, dado que los niños tienen diferentes formas para escribir una palabra, el buscador no solo se limita a la variante correcta, sino que muestra todas las coincidencias y variables. Para dar un contexto más amplio se puede elegir el número de ítems para que aparezcan antes y después.

### ***Corpus EXCALE***

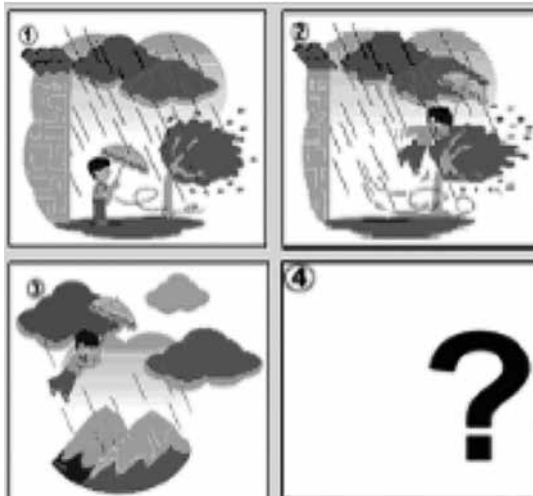
El corpus EXCALE fue elaborado por el Instituto Nacional para la Evaluación de la Educación (INEE). Se trata, en realidad, de un examen sobre los conocimientos de diferentes materias. Las estadísticas, cuestionarios y una parte de la prueba se pueden consultar en línea. Esta base de datos contiene diferentes tipos de escritos realizados por estudiantes de tercero y sexto de primaria, tercero de secundaria y tercero de bachillerato. Con el propósito de dar a conocer lo que los estudiantes mexicanos deben aprender como parte del currículo nacional a lo largo de la educación básica, su finalidad fue determinar el logro educativo a nivel nacional y estatal obtenido al término de un ciclo escolar (Backoff et al., 2010, p. 17).

Los ejercicios no fueron aplicados en todos los colegios, solo participó una muestra representativa de alumnos de las escuelas urbanas públicas, rurales públicas, de educación indígena, telese-

cundaria, secundaria técnica, bachillerato profesional técnico y escuelas privadas más distintivas del país.

En cuanto a los grados de nivel básico, las materias evaluadas fueron: Español, Matemáticas, Formación Cívica y Ética, Ciencias Naturales y Ciencias Sociales. En el módulo de Español, se integraron contenidos sobre el sistema de escritura, forma y propósito de los textos, gramática, retórica, vocabulario, comprensión de textos y producción escrita.

Zamudio *et al.* (2012) plantean que este tipo de exámenes requieren que los textos utilizados para la evaluación sean comparables, por eso es importante saber elegir el tipo y las instrucciones de las consignas de los escritos. El análisis de esta investigación se enfocó únicamente en textos escritos por niños de 3° de primaria. La prueba, realizada al final del ciclo escolar 2006-2007, contenía las siguientes instrucciones: "Observa con atención las imágenes e imagina un final creativo para la historia. Escribe el cuento completo". Las imágenes a las que se refieren las instrucciones se muestran en la figura 1.



**Figura 1.** Imágenes que sirvieron de base para la prueba del EXCALE.



Las pruebas están digitalizadas en formato de imagen, lo que obligó a un procesamiento y transcripción de las mismas. Para decidir si cada texto-imagen se tomaba para su estudio o no, se siguieron los siguientes criterios: a) las palabras debían ser entendibles y coherentes; b) la redacción debía versar sobre el tema de la instrucción; c) los relatos debían tener la estructura narrativa solicitada: inicio, desarrollo y final.

Con los criterios de selección de EXCALE se obtuvo una muestra final de 286 elementos procedentes de todos los estados de la república, excepto Oaxaca, entidad que no tenía registros. La distribución final de los textos analizados en este corpus fue la siguiente: 149 hombres (52%), 137 mujeres (48%), 122 para escuelas urbanas públicas (43%), 48 en rurales públicas (17%), 112 en privadas (39%) y 4 en educación indígena (1%). En la clasificación por edad hubo 1 de 7 años (1%), 9 de 8 años (3%), 182 de 9 años (64 %), 80 de 10 años (28%), 9 de 11 años (3%), 4 de 12 años (1%) y 1 de 13 años (1%).

### **Corpus de *Historias Cortas***

Para tener un análisis más completo de la complejidad en la escritura de niños, se necesitaba un corpus que fuera tomado como base para poder hacer una comparación, es decir, formado por diferentes tipos de texto considerados adecuados para su comprensión dependiendo del nivel escolar o años de estudio que tenían los infantes.

Historias Cortas (HC) está conformado por 70 textos divididos en los siete niveles de dificultad de la escala de perspicuidad de Szigriszt-Pazos (1993), desde el más sencillo hasta el más árido. Para clasificar los escritos se utilizó la herramienta digital llamada "legible" (<https://legible.es/>) donde se pueden realizar análisis de diferentes obras o escritos. "Legible" fue diseñada especialmente para medir la facilidad de comprensión lectora de un texto en el idioma

español por el periodista Francisco Szigriszt-Pazos. Es un analizador automático de legibilidad que cuenta de manera rápida el número de palabras, sílabas, letras, signos de puntuación, palabras no reconocidas, letras del texto ordenadas por frecuencia y tiempo de lectura estimado del texto en minutos.

En la Tabla 1 se puede observar el tipo de publicación típico para cada nivel en la escala, que depende de los años de estudio o escolaridad de las personas, así como el puntaje que corresponde a cada uno de los niveles.

La recolección del corpus se basó en la tabla de perspicuidad. Se obtuvieron 10 fragmentos representativos de diferentes obras dentro de cada nivel que tuvieran un rango de entre 200-250 palabras. Todos los textos fueron recopilados de diversas páginas de internet y tienen formato TXT.

## **Clasificación de los escritos**

El primer análisis que se hizo con los textos de los corpus CEELE y EXCALE fue ver si su lugar en la escala de Szigriszt-Pazos era coherente con la propuesta del autor para niños de 2° y 3° de primaria. Para ello, se normalizaron todos los escritos para que tuvieran un solo párrafo. A continuación, se utilizó la herramienta *online* "Legible" para clasificarlos. Los resultados de este primer experimento se pueden observar en la Tabla 2.

De acuerdo con la escala, se esperaba que la complejidad de producción escrita de los niños de 3° de primaria estuviera situada entre los niveles "Muy fácil" a "Fácil" y la de los alumnos de 2° solo en el primer nivel. Según estas mismas especificaciones, los textos del corpus EXCALE deberían distribuirse hasta "Normal", pues la edad de los alumnos va de los 7 a los 13 años. Con este criterio, 73.77% de los textos, es decir 211, cumplieron con los puntajes re-

queridos en el EXCALE y 143 (49%) con los parámetros del CEELE. En la tabla 3 se aprecia que el puntaje negativo en EXCALE en algunos textos supera al del corpus de HC. Esto quiere decir que algunos textos se consideran mucho más complejos que fragmentos de García Márquez. La longitud de las palabras influyó mucho en la clasificación, pues algunas no tenían la separación correcta. Por ejemplo: "25deoctubredel2005", "segundogrado", "CuantosañosienescomoSellamaendondebibes" o "abesesenmotor".

**Tabla 2.** Clasificación de los escritos de CEELE y EXCALE según la escala de perspicuidad.

Nivel	Número de escritos	
	EXCALE	CEELE
Muy fácil	16	8
Fácil	53	21
Bastante fácil	61	47
Normal	81	67
Bastante difícil	31	35
Árido	25	39
Muy difícil	19	83

**Tabla 3.** Resultados en los niveles extremos de la escala en cada corpus.

CORPUS	Documento	Perspicuidad	No. Palabras	No. Frases	Palabras/ Frase
EXCALE	yuc15	96.51	151	14	11
EXCALE	qr11	-27.09	123	1	123
CEELE	N040403ESM26	99.42	75	17	4.41
CEELE	S041202EscF03	-185.99	243	1	243
CH	Adivinanzas	99.78	236	37	6.38
C H	Otoño_patriarca	-153.19	237	1	237

A simple vista, la mayoría de las redacciones de niños clasificadas en las últimas categorías contienen solo una frase a lo largo del párrafo y el número de palabras en total es mucho mayor en comparación a los niveles más sencillos.

En las columnas de número de frases y número de palabras por frase de la Tabla 3 se tienen desde 14 frases por un párrafo de 151 palabras dentro del primer nivel hasta una frase por párrafo de 243 palabras en el nivel siete. Esto último indica que posiblemente no se utilizaron signos de puntuación o se escribieron palabras menos comunes o más largas debido a que la complejidad en la escala de perspicuidad está determinada por estos factores. A partir de este análisis se puede deducir qué pasa en los demás. Notoriamente, en los corpus de escritos de niños el número de frases va disminuyendo y aumenta el número de palabras por frase. También se pudo observar que el hecho de que grupos de palabras se escriban sin espacios aumenta la dificultad. En el CEELE, la mayoría de los textos quedaron situados en un lugar más bajo del esperado porque los niños casi no utilizaban signos de puntuación para separar sus ideas, agregando aún más complejidad cuando juntan varias palabras y sus descripciones son más largas que el resto.

De todo ello se desprende que, en contra a lo estipulado por Szigriszt, los textos con menor puntaje no reflejan un alto dominio de la lengua ni mucho menos experiencia con la escritura. Al contrario, de esto se pueden observar fallas ortográficas, de cohesión, fluidez y separación de oraciones. Debido a que el índice de perspicuidad se encarga de medir la facilidad de comprensión lectora de los textos, los factores tomados en cuenta para el cálculo se enfocan en las características de los textos realizados para ser comprendidos por determinado nivel educativo y que han sido realizados por escritores con experiencia.

## Análisis de los patrones silábicos

Para tener un análisis más completo acerca de los fenómenos que afectan la complejidad en la escritura de niños, realizamos un análisis de patrones silábicos. La Tabla 4 muestra que los patrones más comunes en los tres corpus que se han estudiado son: CV, CVC, CVV, V, VC.

**Tabla 4.** Distribución porcentual de los patrones silábicos más comunes en los corpus.

	CV	CVC	CCV	CCVC	VV	VC	V	CVVC	CVV
CEELE	49.6	10.6	5.8	2.1	0.6	8.8	10.7	2.5	7.7
EXCALE	48.6	13.5	5	0.7	0.7	8.4	11.2	3.7	7.8
HC	47.7	20.7	5.3	1.5	N.D.	7.5	7	3.2	6.3

Patrones sencillos como CV, CVC y V aparecen de forma regular dentro de cada uno de los niveles de los 3 corpus, aunque con distintas fluctuaciones en cada uno de ellos. La estructura más compleja de las tres, CVC, tiende a aumentar en el corpus HC. El índice de aparición de otros patrones, como VV, VC, CVVC, CCVC es menor a 10%. Estos resultados podrían indicar que tener un nivel de escritura competente no necesita el empleo de tantas combinaciones de patrones silábicos más complejos, sino que recae en el uso de vocabulario específico, pues el hablante es capaz de utilizar más combinaciones de palabras o lenguaje especializado. Algunos patrones encontrados, como cadenas de consonantes sin vocales, no pueden realizarse en español. Se trata de formas como CC, CCC, CCCV. En general, pertenecen a palabras incompletas, raras o mal escritas como: llamn, hulz, Ruth, direcsson, ChangWang. El uso de estas estructuras, inválidas para el español, indica una baja competencia en habilidad escrita. Para el resto de los patrones, se muestra una concurrencia consistente a través de todos los niveles de cada

corpus, lo cual pudiera dar información significativa acerca de la diversidad de vocabulario. Al analizar las combinaciones silábicas más comunes dentro de los patrones CV y CVC, se puede comenzar a medir el nivel de habilidades escritas de los niños.

Por lo que respecta a las realizaciones de los patrones más recurrentes en el corpus, CV y CVC, las figuras 2 y 3 son explicativas de la distribución. Dentro del patrón CV, las sílabas "la", "de", "se", "mi" y "su" son las más empleadas, aunque, según se observa en la Figura 2, su distribución es bastante desigual. Entre las realizaciones que presentan más variabilidad de un corpus a otro se encuentra "mi". Esta sílaba se encuentra representada enfáticamente en el CEELE, mientras que en el EXCALE y en el HC tiene una aparición menor. Un caso contrario es el de "su", con un pico de frecuencia alto en el EXCALE, mientras en los otros corpus mantiene un porcentaje de aparición mucho más modesto.

En general, el valle que se aprecia en la gráfica cuando se representa la ocurrencia de estas sílabas en HC indica que, en la escritura adulta, la realización de los patrones silábicos está mucho más distribuida, existe una variedad mucho mayor, y no existen sílabas con excesos de representación.

El desbalanceo entre CEELE y EXCALE mostrado en el patrón vocálico CV entre "mi" y "su" tiene su paralelo en CVC con "los" (pico en CEELE) y "su" (pico en EXCALE, Figura 3). Es un fenómeno también resaltable que, en este caso, no existe un valle pronunciado en HC, sino que una de las realizaciones silábicas ("del") presenta también un elevamiento en los niveles intermedios de la escala.



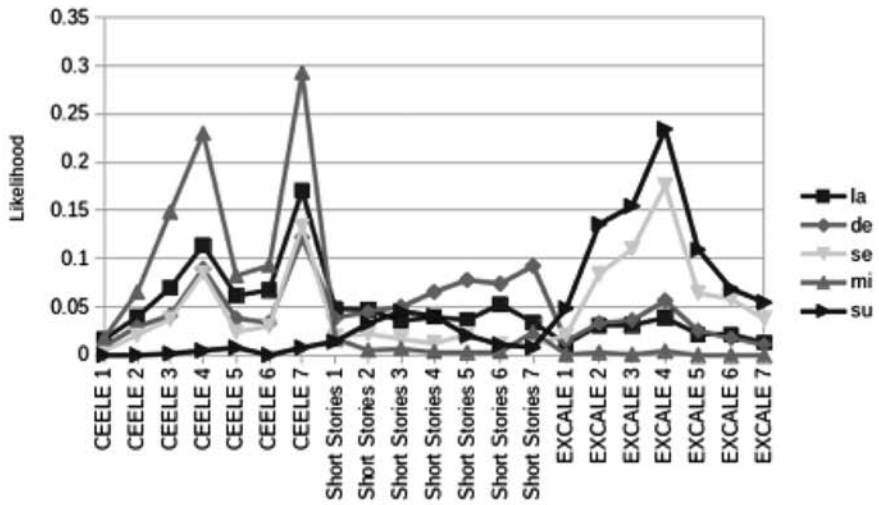


Figura 2. Distribución normalizada a una de las realizaciones más frecuentes del patrón CV en cada uno de los corpus y en cada escala de complejidad.

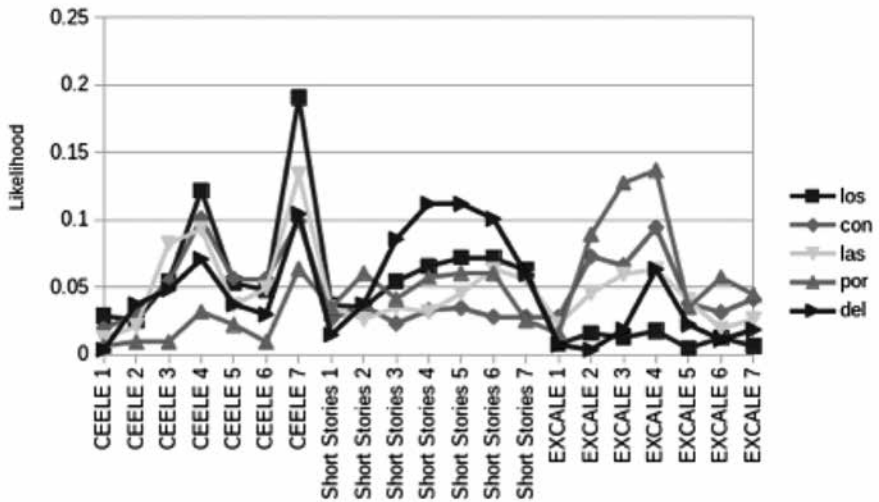


Figura 3. Distribución normalizada a una de las realizaciones más frecuentes del patrón CVC en cada uno de los corpus y en cada escala de complejidad.

## Conclusiones y trabajo futuro

El desarrollo adecuado de las habilidades de lectura y escritura permitirán que los niños de educación primaria adquieran las bases adecuadas para interpretar su propia realidad, organizar y estructurar mejor su pensamiento. Ferreiro (2000) menciona que para 2° y 3° de primaria, los alumnos deberían hacer un uso correcto de aspectos como la relación de los sonidos con su correspondiente grafía, reglas ortográficas, segmentación correcta de palabras, utilización de signos de puntuación y coherencia en los textos. Sin embargo, con los resultados arrojados por el análisis de la fórmula de perspicuidad, esta afirmación no queda del todo confirmada. Al revisar los escritos de CEELE, pertenecientes a niños de 2°, se observaron problemas de hiposegmentación entre varias palabras, omisión de letras y casi nula utilización de signos de puntuación como la coma y el punto, además de falta de coherencia en las ideas de las descripciones; aunque al avanzar de grado, en particular con los escritos de 3° de primaria pertenecientes a EXCALE, estos problemas fueron disminuyendo.

Este trabajo permite comprobar que el nivel de habilidades escritas de los alumnos estudiados cumple con las características esperadas para su nivel educativo. Por ello, la fórmula usada, el índice de perspicuidad de Szigriszt-Pazos, que toma en cuenta aspectos de longitud de oraciones y combinaciones silábicas, no es aplicable a los escritos de escolares de estos grados. Por otro lado, tomando en cuenta el primer objetivo particular de la investigación, se ha comprobado que los textos de los alumnos de 2° y 3° de primaria no alcanzan el nivel adecuado dentro de la escala de perspicuidad debido a que utilizan construcciones no estipuladas dentro de dicho índice.

Los factores que más influyeron en la clasificación fueron la ausencia de signos de puntuación para separar oraciones, el empleo

de gran cantidad de palabras dentro de una oración y la longitud de las palabras. El principal problema parece ser que las escalas para medir la complejidad están basadas en la capacidad de comprensión lectora, y no en la capacidad de competencia escrita que, según se ha demostrado en nuestro análisis, se basa en parámetros muy diferentes.

Queda claro, por otra parte, la importancia de enfocarse en el estudio de los textos de niños para tratar de predecir futuros fracasos en las habilidades escritas y no solo calificar la calidad de la escritura con dictados de palabras o basándose en la ortografía.

En el futuro, se prevé ampliar el estudio para realizar el análisis con escritos en diferentes niveles educativos con el fin de comparar cómo evolucionan las habilidades escritas de los alumnos. También se quiere comparar los escritos de personas que tengan como lengua materna otros idiomas románicos y que estén aprendiendo el español como L2 para comprobar si los problemas encontrados son parecidos. Finalmente, y dadas las dificultades de medición propias de la fórmula de perspicuidad, se propone elaborar una escala para valorar el nivel de producción escrita basada en la competencia de tal habilidad y no en la comprensión.

## Referencias

- Allen, L. K., Snow, E. L., Crossley, S. A., Jackson, G. T., & McNamara, D. S. (2014). Reading comprehension components and their relation to writing. *L'Année Psychologique*, 114 (4): pp. 663–691.
- Backhoff Escudero, E., Sánchez Moguel, E., Peón, A., Zapata, M. & Andrade Muñoz, E. (2010). Comprensión lectora y habilidades matemáticas de estudiantes de educación básica en México: 2000-2005. *Revista electrónica de investigación educativa* 12 (1): pp. 1–15.
- Barrio-Cantalejo, I.M., Simón-Lorda, P., Melguizo, M., Escalona, I., Marijuán, M., & Hernando, P. (2008). Validación de la Escala INFLESZ para evaluar

- la legibilidad de los textos dirigidos a pacientes. En *Anales del Sistema Sanitario de Navarra*, tomo 31: pp. 135–152. SciELO España.
- Crawford, A.N. (1984). A Spanish language Fry-type readability procedure: Elementary level. *Bilingual education paper series. Evaluation, Dissemination and Assessment Center*. Vol. 7(March, 1984): pp. 1- 17.
- Defazio, J., Jones, J., Tennant, F. & Hook, S. A. (2010). Academic Literacy: The Importance and Impact of Writing across the Curriculum--A Case Study. *Journal of the Scholarship of Teaching and Learning*, 10 (2): pp. 34–47.
- Díaz Argüero, C., Zamudio Mesa, C. y Méndez Cruz, C. (2010). *Corpus Electrónico para el Estudio de la Lengua Escrita [CEELE]*. México: Universidad Nacional Autónoma de México.
- Fernández Huerta, J. (1959). Medidas sencillas de lecturabilidad. *Consigna: revista pedagógica de la sección femenina de Falange ET y de las JONS* (214): pp. 29-32.
- Ferreiro, E. (1998). *Alfabetización: teoría y práctica*. México: Siglo XXI.
- Ferreiro, E. (2000). Entre la sílaba oral y la palabra escrita. *Infancia y aprendizaje* 23(89): pp. 25–37
- Flesch, R. (1948). A new readability yardstick. *Journal of applied psychology* 32(3): p. 221.
- Gutiérrez de Polini, L. E. (1972). Investigación sobre lectura en Venezuela. En *Primeras Jornadas de Educación Primaria*. Encuentro llevado a cabo en el Ministerio de Educación, Venezuela.
- Hayes, J. R. & Flower, L. S. (1980). Identifying the organization of writing processes. En Gregg, L. W., Steinberg, E. R. (eds.), *Cognitive processes in writing* (pp.3-30). Hillsdale, NJ: Erlbaum.
- Hayes, J. R. (2012). Modeling and Remodeling Writing. *Written Communication*, 29, pp. 369–388. doi: 10.1177/0741088312451260
- Hemphill, L., & Tivnan, T. (2008). The Importance of Early Vocabulary for Literacy Achievement in High-Poverty Schools. *Journal of Education for Students Placed at Risk (JESPAR)*, 13, pp. 426–451. doi: 10.1080/10824660802427710

- Kellogg, R. T. (2008). Training writing skills: A cognitive developmental perspective. *Journal of Writing Research*, 1, pp. 1–26. <https://doi.org/10.17239/jowr-2008.01.01.1>
- Kintsch, W. (1998). *Comprehension: A paradigm for cognition*. Cambridge: Cambridge University Press.
- Muñoz, M. (2006). Legibilidad y variabilidad de los textos. *Boletín de Investigación Educativa*, 21 (2): pp. 13–26.
- Myers, J. L., & O'Brien, E. J. (1998). Accessing the discourse representation during reading. *Discourse Processes*, 26: pp. 131-157.
- Sánchez Abchi, V., Diuk, B., Borzone, A.M. & Ferroni, M. (2009). El desarrollo de la escritura de palabras en español: Interacción entre el conocimiento fonológico y ortográfico. *Interdisciplinaria* 26 (1): pp. 95–119.
- Szigiszt-Pazos, F. (1993). *Sistemas predictivos de legibilidad del mensaje escrito: fórmula de perspicuidad*. Universidad Complutense de Madrid, Servicio de Reprografía.
- Tapia Poyato, A. M. (2003). Orden de adquisición de segmentos y tipos de sílaba en español. *Cauce* (26): pp. 491–501.
- Thorndike, E. (1921), *The Teacher's Word Book*. Recuperado de <http://www.archive.org>
- Zamudio Mesa, C. (2008). Influencia de la escritura alfabética en la segmentación de sonidos vocálicos y consonánticos. *Lectura y vida: Revista latinoamericana de lectura* 29 (1): p. 10.
- Zamudio Mesa, C., Díaz, C. & Lepe García, E. (2012). *El aprendizaje de los contenidos curriculares de Español. Un análisis de los resultados de los Exámenes de la Calidad y el Logro Educativos (Excale 03, 06 y 09)*. México: INEE.