

Complejidad estructural en textos escolares descriptivo-expositivos: estudio de corpus

Structural Complexity
in Descriptive-Expository
School Texts: Corpus
Study

Complexidade estrutural
em textos escolares
descriptivo-expositivos:
estudo de corpus

Ligia Ochoa Sierra* <https://orcid.org/0000-0002-8761-4819>



Para citar este artículo

Ochoa Sierra, L. (2021). Complejidad estructural en textos escolares descriptivo-expositivos: estudio de corpus. *Folios*, (54). <https://doi.org/10.17227/folios.54-11977>

Artículo recibido
22 • 06 • 2020

Artículo aprobado
28 • 01 • 2021

* Doctora en lingüística de la Universidad Autónoma de Madrid, España y Doctora en Educación por la UNED, Madrid, España. Profesora del Departamento de Lingüística de la Universidad Nacional de Colombia.
Correo: lochoas@unal.edu.co

Resumen

Este artículo de investigación tiene como objeto de estudio la complejidad lingüística estructural, específicamente la relacionada con los textos escolares descriptivo-expositivos. De esta forma, en el texto se indagó acerca de los factores (dimensiones) lingüísticos que permiten medir la complejidad y clasificar los textos de una muestra. Se analizaron 80 textos escolares a partir de una serie de rasgos (agrupados en dos niveles: morfosintáctico y semántico) y a través de un análisis factorial. Los resultados muestran que los factores *cantidad*, *prototipicidad* e *informatividad*, resultaron significativos y el factor *variedad* parcialmente productivo. Frente a los rasgos, se valida la hipótesis de que es necesario incluir en el análisis no solo rasgos sintácticos sino también semánticos e incluso pragmáticos; no obstante, estos resultados son provisionales pues se requiere repensar los criterios y las hipótesis explicativas dadas.

Palabras clave

lingüística descriptiva; sintaxis; semántica; pragmática; complejidad textual.

Abstract

This research article deals with linguistic complexity, specifically regarding descriptive-expository school texts. The article focuses on the linguistic factors (dimensions) that allow to measure complexity and classify sample texts, where 80 school texts were analyzed by factor analysis according to a set of features, grouped at the morpho-syntactic and semantic levels. Results show that factors such as *quantity*, *prototypicality*, and *informativity* were significant, and the factor *variety* was partially productive. As to the features, it was confirmed by the hypothesis that the analysis should include not only syntactic, but also semantic and even pragmatic features. However, these are interim results as the criteria and the explanatory hypotheses require further analysis.

Keywords

descriptive linguistics; syntax; semantics; pragmatics; textual complexity

Resumo

O objetivo desta e artigo de pesquisa é estudar a questão da complexidade lingüística, especificamente aquela relacionada a textos escolares descritivos-expositivos. Investigamos fatores lingüísticos (dimensões) que nos permitem medir a complexidade e classificar os textos de uma amostra. Oitenta textos escolares foram analisados a partir de uma série de características (agrupados em dois níveis: morfossintático e semântico) e através de uma análise fatorial. Os resultados mostram que a quantidade de fatores, a prototipagem e a informatividade foram significativas e a variedade de fatores foi parcialmente produtiva. Contra os traços, a hipótese de que é necessário incluir na análise não apenas traços sintáticos, mas também semânticos e até pragmáticos é validada. Entretanto, estes resultados são provisórios, uma vez que os critérios e as hipóteses explicativas dadas precisam ser repensadas.

Palavras chave

lingüística descritiva; sintaxe; semântica; pragmática; complexidade textual

Introducción

Se puede decir que los textos descriptivos-expositivos aparecen a lo largo de la escolaridad en todos los grados y son asignados para su lectura por los profesores sin tener en cuenta, por lo general, las características sintácticas y semánticas de dichos textos. En ocasiones, como lo señalan Berendes et ál. (2018), su complejidad no aumenta gradualmente en forma sistemática a lo largo de la escolaridad.

Este artículo presenta los resultados finales de una investigación cuyo objetivo fue indagar acerca de la complejidad estructural de los textos escolares descriptivos-expositivos, con el fin de reconocer los criterios que la determinan, qué factores están asociados a ella, qué textos de este mismo tipo son más complejos y cómo se presenta la complejidad por grados y por áreas escolares. Para efectos de esta investigación, se asumió el concepto de *complejidad estructural* dado por Palloti (2015), entendida como la propiedad formal de textos que tiene que ver con el número de elementos lingüísticos y sus patrones relacionales.

Es de entender que la complejidad de un texto puede depender de múltiples factores, entre ellos su estructura y estrategias metacognitivas que posee el lector (McNamara, 2004), tipo de texto y asignatura (Graesser y McNamara, 2011), las habilidades del lector y la naturaleza de las tareas asignadas (Campos et al., 2014), el nivel educativo (Lu y Ai, 2015), tipo de registro –oral o escrito– (Nippold et ál., 2017), el tema (Yoon, 2017) y el conocimiento previo (Zunino, 2017). Por tanto, un texto que es complejo para un lector puede resultar simple para otro; y análisis estadísticos, frente al comportamiento en una prueba, pueden dar indicios acerca de cómo el lector percibió el texto.

Diversas investigaciones educativas estudian la relación entre complejidad lingüística estructural y comprensión lectora; para ello usan o bien sistemas informáticos que tienen ya sus criterios para medir la complejidad (*KoRpus*, *Coh-Matrix*, *lexile*) o establecen sus propios criterios, y con base

en estos, identifican el nivel de complejidad de un texto y cómo esta variable influye en los procesos de comprensión lectora. Sin embargo, cuando se valora un texto a través de programas informáticos o índices de complejidad estandarizados (Szigriszt-Pazos, Inflesz, legibilidad μ) o propios, sucede que el mismo texto sale ubicado en diferentes escalas: para un medidor puede ser un texto normal o incluso fácil y para otro, muy difícil. Uno de los factores que determinan esta discrepancia tiene que ver con el hecho de que falta investigación lingüística que respalde los sistemas de medición y que permita tener criterios fiables.

De otra parte, existe disenso entre los investigadores acerca de algunos criterios lingüísticos como indicadores de complejidad, es decir, algunos rasgos que se consideran indicadores de complejidad han sido cuestionados. Así, por ejemplo, aunque se ha asociado la longitud de la oración con la complejidad, se ha demostrado que oraciones incluso cortas sin un indicador claro de la relación entre dos oraciones suelen ser más complejas que una oración larga con indicadores claros (Frantz et ál., 2015; Ochoa, 2016). Algo similar sucede con la presencia de los marcadores o conectores discursivos; para unos investigadores su presencia complejiza el texto y para otros no. Por ejemplo, Vyatkina (2013) concluyó que la coordinación es un factor de complejidad sintáctica importante; no obstante, investigadores como Van Silfhout, Evers-Vermeul y Sanders (2015) consideran que “los conectores conducen a un procesamiento más rápido de la información posterior, así como tiempos de lectura más cortos de información del texto anterior” (p. 47). En una posición intermedia, Hoek et ál. (2017) consideran que la coherencia se puede marcar de manera explícita (mediante conectivos) o implícita, pero que algunas relaciones implícitas aumentan la complejidad del texto cuando estas no son esperadas por el lector. Igualmente, la presencia del subjuntivo como marca de complejidad ha sido cuestionada por investigaciones (Bulté y Housen, 2014) que han demostrado que el uso del subjuntivo decrece

en jóvenes y adultos, y en contextos académicos en los que se privilegian las nominalizaciones o los términos deverbales (Norris y Ortega, 2009).

Finalmente, un vacío señalado en varias investigaciones es la falta de investigación de la complejidad en diversas tipologías textuales; se considera que tipos y géneros textuales distintos requieren medidas distintas (Frantz et ál., 2015; Ochoa, 2016).

Por tanto, cuanto más investigación lingüística se tenga de la complejidad estructural o absoluta, más depurados serán los criterios y, por ende, mayor confiabilidad habrá en los resultados de pruebas de comprensión lectora, donde se indaga acerca de la relación entre complejidad y lectura. Esto permitirá a su vez mayor unificación en las mediciones de los textos. Igualmente, los resultados pueden aportar datos para futuras investigaciones lingüísticas o psicolingüísticas y, por supuesto, para aplicaciones en otros campos.

Así mismo, tener este insumo lingüístico permite a diversos actores educativos (docentes, coordinadores de área, coordinadores académicos y editores) contar con unos criterios para seleccionar los textos que usan para desarrollar y evaluar los procesos de comprensión lectora, escalonarlos adecuadamente y evitar que los alumnos estén expuestos a textos muy sencillos o demasiado complejos para su edad, perfil y nivel escolar. En relación con los textos escolares se espera que su complejidad aumente de forma progresiva de grado a grado y de ciclo a ciclo.

En la revisión bibliográfica que se hizo para establecer el estado de la cuestión de esta investigación, se encontraron alrededor de 150 textos académicos en los que aparecían los lexemas *complejidad textual*, *lingüística*, *semántica* o *sintáctica*. En estos textos se buscaron los criterios usados para medir la complejidad lingüística y los criterios para identificar qué hace que un texto sea más complejo que otro.

A continuación se reseñan brevemente algunos de estos trabajos.

Criterios para medir la complejidad lingüística

La mayoría de investigaciones, tanto teóricas como aplicadas, usan criterios sintácticos y léxicos para medir la complejidad (Bulté y Housen, 2014; Kurdi, 2017; Pallotti, 2014). Los primeros a partir de medidas como: longitud de palabras, frases y oraciones; cantidad de palabras, frases y oraciones; tipos de oración (simples, compuestas), y complejidad de la oración (Lahuerta, 2017; Mancilla, Polat y Akcay, 2017); los segundos a partir de consideraciones como: diversidad léxica, frecuencia léxica (asiduidad de una palabra en la lengua) y la relación *tipo/token* –variedad de palabras únicas (tipo) que aparecen en un texto en relación con el número total de palabras (*tokens*)–. Pallotti (2014) incorpora, además, la complejidad morfológica, referida a la cantidad de formas que toman las palabras para marcar categorías gramaticales y funciones.

Pocos autores tienen en cuenta aspectos semánticos; sin embargo, la cohesión ha sido tenida en cuenta por algunos investigadores (p. ej., Plakans y Bilki, 2016; Solnyshkina, Zamaletdinov, Gorodetskaya y Gabitov, 2017) y también se han considerado aspectos pragmáticos (De Marneffe, Manning y Potts, 2012; Juola, 2008; Andrews y Lamb, 2017)

A propósito, el nivel semántico y el pragmático han sido tenidos en cuenta en programas informáticos que miden la complejidad textual, con miras a analizar la dificultad de los textos en los procesos de comprensión y producción. La mayoría de estos programas analizan preferentemente textos en inglés (cf. *Lexile[®] Coh-Metrix[®]*, *Reader-Bench[®]*, *KoRpus: An R Package for Text Analysis[®]*, s. f.).

De otro lado, investigaciones aplicadas en relación con la enseñanza de lenguas extranjeras también presentan criterios para medir la complejidad lingüística, como la longitud media de la oración y la cláusula (Lu, 2010; Cunningham et ál., 2012; Youn, 2014), subordinación (Youn, 2014; Mancilla et ál., 2017) y complejidad del SN (Lu, 2010; Youn, 2014; Schilk y Schaub, 2016); complejidad a nivel de la palabra, frecuencia y familiaridad (Cunningham et ál., 2012)

Criterios para determinar qué hace que un texto sea más complejo que otro

Un trabajo pionero en este ámbito es el presentado en el *Plan curricular del Instituto Cervantes. Niveles de referencia para el español* (Instituto Cervantes, 2006), en el que se señalan algunas características de los textos según los niveles A, B y C propuestos por el MCER (Marco Común Europeo de Referencia para las lenguas: aprendizaje, enseñanza, evaluación). De acuerdo con este Marco, en el nivel A, los textos tienen una estructura clara y sencilla; la información es concisa y explícita; formato accesible, preferiblemente están impresos. En el nivel B, tienen una estructura clara tanto en forma como contenido; la información es explícita, aunque puede aparecer implícita si resulta fácil desentrañar el sentido; además, están presentados en formato accesible con títulos y subtítulos. En el nivel C, los textos son largos, están dirigidos a un receptor con estudios especializados; tratan sobre temas abstractos; pueden ser literarios o no literarios; la información que contienen puede estar enunciada en clave de ironía y de humor, de manera que, para descifrar el contenido, puede necesitarse recurrir a referencias y conocimientos culturales compartidos entre el emisor y el receptor y puede ser necesario un esfuerzo para reconstruir su estructura; pueden abordar temas técnicos o especializados y emplear registros específicos fuera de lo habitual; los textos orales pueden contener coloquialismos, frases hechas o expresiones idiomáticas.

En lo referente a qué hace un texto más complejo que otro, Ochoa (2016) analiza textos usados en pruebas nacionales colombianas a partir de una serie de rasgos sintácticos, semánticos y pragmáticos, como longitud del texto, modos verbales, polisemia, vocabulario concreto vs. abstracto, marcas relacionadas con el emisor, receptor, propósito comunicativo. El análisis evidencia que factores como cantidad, variedad, explicitud son decisivos para determinar la complejidad.

Por su parte, Sheehan, Flor, Napolitano y Ramineni (2015) investigaron si los libros de texto pensados para niños de primer grado estadouni-

denses y producidos por una editorial específica durante los años 1962 y 2013 se han hecho más complejos o no. La medición se hizo con el *software TextEvaluator*. Los resultados muestran que la complejidad ha aumentado o se mantiene constante, debido a que han incluido una proporción creciente de pasajes informativos, cambio acompañado por las siguientes modificaciones específicas: a) aumento en la proporción de palabras que tienden a aparecer con menos frecuencia en el texto impreso; b) aumento en la proporción de palabras que son más características del texto académico como opuesto a la ficción o la conversación; c) niveles más bajos de cohesión referencial; d) niveles más bajos de narratividad, y e) menos instancias de un estilo interactivo/conversacional (p. 1).

En programas informáticos también se consideran criterios importantes si las palabras y estructuras son menos familiares o poco usuales (Coh-Metrix, s. f.), si se usan palabras abstractas (Coh-Metrix, s. f.) y si los textos son menos cohesivos o con pocas marcas de relación semántica (Coh-Metrix, s. f.; Ochoa, 2016).

Así mismo, se encuentra en la revisión bibliográfica que existen unos géneros discursivos más complejos que otros. El discurso argumentativo (Yang et ál., 2015; Meneses y Ow, 2016; Polio y Yoon, 2018) y el discurso académico aparecen como los más complejos, en tanto que el discurso cotidiano está en el lado opuesto (Nippold et ál., 2017); y el discurso narrativo, en medio (*cf.* Nini, 2015). Para Meneses, Ow y Benítez (2012), “los textos narrativos orales son más largos y tienen una mayor diversidad léxica que los textos explicativos escritos” (p. 65). Pero a su vez los textos expositivos poseen mayores relaciones intraclausales “como una estrategia para condensar y empaquetar información dentro de la cláusula”, además de que, en estos textos, predomina la hipotaxis.

Por su parte, Aravena y Hugo (2016) encuentran que el modo discursivo es un factor determinante en la complejidad, “dado que los textos explicativos reúnen la mayor proporción de relaciones interclausulares complejas, específicamente hipotácticas” (p. 9).

En cuanto a los criterios que determinan la complejidad de un tipo de texto, cabe señalar que Frantz, Starr y Bailey (2015, p. 389) parten de la hipótesis de que la variable *asignatura* podría determinar la complejidad. Ponen como evidencia estudios previos realizados por ellos mismos y por otros investigadores. Por ejemplo, en uno de 2007, usando

el verbo como unidad de análisis, encontraron que hay una tendencia a la complejidad en los textos de ciencias.

En resumen, en la tabla 1 se muestran los criterios empleados por distintos investigadores, programas informáticos y entidades gubernamentales para determinar la complejidad estructural.

Tabla 1. Criterios usados para medir la complejidad

Nivel	Criterios
Sintáctico	Formato del texto, longitud de la palabra y la oración, clases de oraciones (simples, compuestas por coordinación, compuestas por subordinación), complementos nominales (adjetivos, sintagmas preposicionales), orden sintáctico.
Semántico	Cantidad y variedad de información, diversidad léxica, palabras nocionales versus palabras gramaticales, léxico especializado versus léxico frecuente.
Pragmático	Cohesión, presuposición.

Fuente: elaboración propia.

Haciendo un balance de la literatura revisada se puede concluir que son preponderantes los estudios de corte sintáctico para medir la complejidad lingüística, aunque se han ido incorporando criterios semánticos y pragmáticos. Los programas informáticos resultan interesantes y completos, pues tienen en cuenta varios niveles de lengua, pero, desafortunadamente, la mayoría de ellos están creados para el análisis de textos en lengua inglesa, por tanto, la transposición al español resulta problemática.

Es necesario, entonces, incorporar criterios semánticos y pragmáticos, estudiar la complejidad en tipologías textuales específicas, validar los criterios y factores lingüísticos que permiten medir la complejidad, tareas que se pueden lograr a la luz de un modelo factorial estadístico que permita una mejor comprensión del problema.

Metodología

La investigación es un estudio de alcance descriptivo, con un diseño de investigación cuantitativo no experimental.

El corpus

Esta investigación se realizó sobre un corpus de estudio conformado por textos pensados para niños, jóvenes y adultos. En total se analizaron 80 textos, discriminados en 4 asignaturas: 20 de Lengua Materna, 20 de Ciencias Sociales, 20 de Ciencias Naturales y 20 de Matemáticas, y 5 grados escolares: 3.º, 5.º, 7.º, 9.º y 11.º. Por cada grado y área se analizaron 4 textos. Para la búsqueda de los datos se acudió inicialmente al Icfes, pero esta entidad no permitió el acceso a la base de datos de los textos empleados en las pruebas. Por consiguiente, se decidió buscar los textos en editoriales reconocidas, cuya producción es amplia y referenciada para uso en Colombia, a saber: Santillana, Norma, Libros y Libres, Ediciones SM y Educar.

Aunque algunos autores diferencian entre los textos descriptivos y expositivos (*cfr.* Werlich, 1976; Adam, 1992; Berasain, s. f.) se decidió agrupar los textos descriptivos y expositivos, ya que en la selección y el análisis del corpus de esta investigación hubo muy pocos textos descriptivos *puros*; por lo

general, estaban formados no solo por secuencias descriptivas, sino también por secuencias expositivas, incluso en algunos textos no era posible diferenciar la descripción de la exposición. De otra parte, en los grados superiores, no se encontraron textos descriptivos. Este hecho coincide con lo señalado por Disselhoff (2007) quien señala que las secuencias descriptivas no aparecen solas y que “si la[s] hay, entonces solo marginalmente pertenece[n] a los tipos de texto escolares de la escuela primaria, ya que no hay ningún tipo de texto de este nombre fuera de la escuela en cultura pura” (p. 167). De otra parte, en términos comunicativos ambos tienen la misma función: para Feilke (2017), describir es una forma de explicar-informar, y el mismo propósito comunicativo tienen los textos expositivos.

Criterios de análisis

El corpus se analizó inicialmente de manera intuitiva y después con base en una serie de rasgos tomados de este primer análisis y de la revisión bibliográfica. Estos rasgos eran: número de: palabras, renglones, párrafos, oraciones, sustantivos solos, sustantivos con adjetivo(s), sustantivos con oración relativa, sustantivos con sintagma preposicional, sustantivos con apóstrofos, verbos utilizados, modos verbales, tiempos verbales, enunciadores, personas gramaticales, oraciones coordinadas, oraciones subordinadas, complementos circunstanciales, modo de la oración (afirmativa/negativa, etc.), voz (activa o pasiva); vocabulario concreto vs. abstracto, uso de las figuras literarias, polisemia, significado denotativo vs. connotativo, sustantivos nuevos o no repetidos, adjetivos nuevos, adverbios nuevos, verbos nuevos, deixis (personal, espacial, temporal), marcas del emisor (explícito o no), el receptor, el propósito comunicativo, la estructura textual, diversidad léxica (variabilidad y frecuencia de palabras en textos de la misma longitud), marcas de modalización y de cohesión. Asimismo, mediante el programa informático *KoRpus*⁸, se midió la longitud promedio de la oración (número de palabras por oración), la longitud promedio de la palabra (número de caracteres por palabra), el número promedio de sílabas por palabra y la relación tipo-*token* (TypeToken

Ratio (TTR): total de formas (*token* frente a total de formas distintas, *tipo*).

Como se ve, se tomaron criterios de los niveles morfosintáctico, semántico y pragmático, a partir de los cuales se construyó una matriz para el análisis, la cual se aplicó inicialmente a cuarenta textos.

La implementación de la matriz mostró que varios de los criterios no eran productivos en la medida en que no se encontraron en el corpus: la mayoría de textos estaban construidos en tiempo presente, modo indicativo, voz activa, modo afirmativo, significados connotativos, un solo enunciador, tercera persona; no había (o era muy escasa la presencia) sustantivos con apóstrofos, figuras literarias, polisemia. Eran textos sin explicitud (sin marcas) del emisor, receptor, propósito comunicativo, estructura textual y no había o era muy escasa la presencia de marcas de modalización. Este es un primer hallazgo que resulta significativo con miras a caracterizar el corpus analizado: textos descriptivos- expositivos escolares.

Con base en el resultado anterior, se usaron 25 criterios (rasgos):

- Nivel morfosintáctico. Oraciones totales, oraciones simples, oraciones coordinadas, sustantivos solos, sustantivos con adjetivos, sustantivos con oraciones relativas, sustantivo con sintagma preposicional, cantidad verbos (est), cantidad complementos verbales (est), longitud promedio de la oración, longitud promedio de la palabra, número promedio de sílabas.
- Nivel semántico. Cantidad de palabras, cantidad de renglones, cantidad de párrafos, variedad de elementos verbales (elementos), variedad de complementos verbales utilizados (elementos), sustantivos abstractos, sustantivos concretos, marcas de cohesión.
- Pragmático. Información nueva versus vieja: porcentaje de sustantivos nuevos, porcentaje de verbos nuevos, porcentajes de adjetivos nuevos, TTR, diversidad léxica.

Análisis cuantitativo

Antes de exponer la metodología estadística, es preciso señalar que el propósito principal de este artículo era identificar factores (dimensiones) no observables pero subyacentes a los datos, si estos existen, y clasificar los textos de la muestra según su nivel de complejidad lingüística. En esta vía, se consideró procedente implementar un *análisis factorial exploratorio* y, a partir de los resultados, construir un índice para asignar, a cada texto de la muestra, un puntaje que permitiera clasificarlos de acuerdo con su complejidad lingüística.

Análisis factorial exploratorio

En estadística multivariada (Peña, 2002), el análisis factorial exploratorio (AFE) es un método estadístico introducido para descubrir la estructura fundamental de un conjunto de variables relativamente grande; vale decir, busca identificar variables latentes o dimensiones no observables a partir de un conjunto de variables medidas por el investigador. A continuación, se exponen brevemente los aspectos más importantes en la implementación de un AFE.

Conveniencia

Este método parte del análisis de la matriz de correlación de las variables; para que sea apropiado, se espera, en general, un número importante de correlaciones significativamente grandes, mayores a 0,30. La revisión de las correlaciones cuando el número de variables es considerablemente grande, además de ser tedioso, puede llegar a ser complejo; si se cuenta con n variables, es necesario examinar $n(n-1)/2$ coeficientes. Así, por ejemplo, con 10 variables es necesario revisar 45 coeficientes, mientras que si el número de variables asciende a 25 los coeficientes a examinar son 300. Por consiguiente, resulta más conveniente analizar la matriz de correlación como un todo. Dos mecanismos estadísticos fueron usados para este efecto:

- *El contraste de esfericidad de Bartlett*. Esta es una prueba estadística que facilita probar la hipótesis de que no hay factores comunes extraíbles de los datos; en otros términos,

equivale a probar que la matriz de correlación es la identidad, vale decir, que las variables en estudio son incorreladas.

- *La medida de suficiencia muestral de Kaiser (MSA)*. Esta estadística permite cuantificar el grado de intercorrelación entre las variables; más precisamente, comparar los coeficientes de correlación observados con los coeficientes de correlación parcial, presentando un rango de variación entre 0 y 1. Una clasificación comúnmente aceptada para la evaluación de la adecuación del modelo factorial y su interpretación es (Kaiser 1974):

1,00 \geq MSA $>$ 0,90 excelente

0,90 \geq MSA $>$ 0,80 bueno

0,80 \geq MSA $>$ 0,70 aceptable

0,70 \geq MSA $>$ 0,60 mediocre o regular

0,60 \geq MSA $>$ 0,50 malo

MSA \leq 0,50 Inaceptable o muy malo

Este indicador también se calcula de forma individual para cada variable, siguiendo los mismos parámetros de la medida global para su interpretación.

Estimación de un AFE

Dado el objetivo de la investigación, la estimación de factores pasa por el hecho de reemplazar los elementos de la diagonal de la matriz de correlación, por el coeficiente de determinación que resulta de regresar cada variable de análisis con respecto al resto de variables; por definición, el coeficiente de determinación mide la varianza que es compartida por las variables, base de un modelo de análisis factorial común. Los factores que se extraen de este análisis se basan únicamente en la varianza común.

Criterios para la determinación del número de factores a extraer. No existe en la actualidad una base cuantitativa precisa para establecer el número de factores a extraer. No obstante, se han desarro-

llado criterios aproximados para tal fin, los cuales se relacionan a continuación.

- *Criterio de autovalores.* Consiste en extraer tantos factores como autovalores mayores que 1.
- *Criterio de proporción de varianza explicada.* Radica en exigir que el número de factores a extraer acumulen una determinada proporción del total de varianza explicada; en general, se considera que un nivel satisfactorio puede ser del 75 % o el 80 %.
- *El contraste de caída libre.* El criterio para determinar cuántos factores extraer reside en identificar el cardinal del autovalor para el cual la pendiente de la gráfica de “trazado de sedimentación” se torna cercana a 0.
- *Determinación a priori.* En este criterio se asume que el investigador sabe antes de iniciar el análisis el número de factores a extraer.

Interpretación de los factores

La interpretación de los factores extraídos y la adopción de una solución factorial definitiva requieren, en general, implementar las siguientes etapas.

1. En un primer paso, se estima la matriz (inicial) de factores no rotados, procedimiento que además de indicar un número preliminar de factores a ser extraídos, ofrece una primera visión acerca de la estructura subyacente a los datos; la interpretación de dicha estructura (solución) se basa en las correlaciones (cargas factoriales) entre los factores y cada una de las variables originales. Para que la solución se considere adecuada, la matriz de cargas factoriales debe satisfacer las siguientes condiciones:
 - Cada factor debe tener unos pocos pesos altos y los otros próximos a cero.
 - Cada variable debe presentar una carga alta sobre un único factor.

- Los factores deben presentar distribuciones diferentes de cargas factoriales altas y bajas.

Así, puesto que hay más variables que factores comunes, cada factor tendrá una correlación alta con un grupo de variables y baja con las restantes. Por tanto, examinando las características de un grupo de variables asociado con un factor se pueden hallar rasgos comunes que permitan identificar el factor asignándole un nombre.

2. Si, como sucede por lo general en la práctica, la solución inicial no proporciona un patrón adecuado de las cargas factoriales, se hace necesario el uso de algún método de rotación para obtener una solución factorial más simple y teóricamente más significativa. Una rotación consiste simplemente en girar, en el origen, los ejes de referencia de los factores con el fin de obtener un determinado ángulo. Cuando los ejes se giran permaneciendo ortogonales (90 grados) se habla de una rotación ortogonal; cuando el ángulo de rotación es diferente la rotación se denomina oblicua.
3. Una tercera etapa se presenta cuando se plantea la necesidad de especificar nuevamente el modelo factorial debido a:
 - Eliminación de variables.
 - Aplicación de un método de rotación diferente.
 - Extracción de un número diferente de factores.
 - Uso de un método de extracción diferente.

Índice de complejidad lingüística

Con el propósito de clasificar los textos de acuerdo con las características observadas, se construyó un índice de complejidad lingüística, a partir de los resultados del AFE. Este proceso requirió los siguientes pasos:

- a. *Identificación de variables relacionadas con cada factor.* A partir de la matriz de factores,

seguramente rotada, se toman las variables originales cuyas cargas factoriales indiquen una asociación estadísticamente significativa con cada factor.

- b. *Construcción de un índice parcial para cada factor.* Cada índice se estima como la suma de los valores de las variables identificadas como relacionadas con cada factor, dividida por el máximo de estas sumas. Esquemáticamente, si se extraen m factores, sean IPF_i el índice parcial para el factor i ($i=1, 2, \dots, m$) y VF_{1i}, \dots, VF_{ni} (n diferente para cada factor), las variables originales relacionadas al factor i , entonces:

$$IPF_i = \text{Suma}(VF_{1i}, \dots, VF_{ni}) / \text{Máximo}(VF_{1i}, \dots, VF_{ni}) \quad (1)$$

Por construcción, estos son índices estandarizados, ya que presentan un rango de variación entre 0 y 1.

- c. *Estimación de un índice general.* Este índice corresponde a una suma ponderada de los índices parciales, con pesos dados por la proporción de varianza explicada por cada factor con respecto a la varianza total explicada por los m factores extraídos; así, notando los pesos por p_1, \dots, p_m , se tiene:

$$IG = p_1 \cdot IPF_1 + \dots + p_m \cdot IPF_m \quad (2)$$

Donde IG es un índice general variando entre 0 y 1, con 0 indicando una complejidad más baja (mínima) y 1 una complejidad más alta (máxima), bajo el supuesto de que a mayores valores de las variables mayor complejidad lingüística del texto.

- d. *Estratificación en niveles de complejidad.* Con el fin de establecer niveles de complejidad lingüística es necesario construir estratos o grupos de textos usando el índice general y

la técnica estadística de análisis de clúster. Esta técnica emplea algoritmos para construir estratos (grupos) cuya característica es que sus elementos son homogéneos (muy parecidos entre sí) pero difieren en gran medida de los elementos de los demás estratos; esto es lo que en la literatura se encuentra como “grupos internamente homogéneos y externamente heterogéneos” (Hair, Anderson, Tatham y Black, 1999; Díaz y Morales, 2016).

Resultados

En este acápite se presentan los resultados de la aplicación de la metodología antes expuesta; en esa vía, es necesario, en primera instancia, describir de manera breve la información empleada en el estudio. Los datos empleados corresponden a una muestra de 80 textos, distribuidos en las áreas de ciencias, lenguaje, matemáticas y sociales y pertenecientes a los grados 3.o, 5.o, 7.o, 9.o y 11.o de educación básica primaria y secundaria; para cada una de las unidades de observación, los textos, se realizaron mediciones sobre 25 características (ver tabla 2).

Adecuación de los datos

El primer punto a tener en cuenta en el análisis de resultados es la verificación de la idoneidad del conjunto de datos disponible para la implementación de un AFE. De la tabla 3, el valor general de la medida de Kaiser, 0, 83, indica que los datos disponibles sustentan de forma apropiada la aplicación de un análisis factorial; esta conclusión es apoyada por la significancia arrojada por la estadística de Bartlett ($<0,0001$), a partir de la cual se puede concluir que, para cualquiera de los niveles usuales de significancia (0,01; 0,05 y 0,10) se rechaza la hipótesis de diagonalidad de la matriz de correlación, lo que implica que si existen relaciones significativas entre las variables y más de una dimensión subyacente en los datos.

Tabla 2. Medida de Kaiser de suficiencia muestral (MSA) y prueba de Esfericidad de Barlett

Medida de Kaiser de suficiencia muestral por variable	MSA	0,830
Prueba de esfericidad de Bartlett	Chi-Cuadrado	3044,3247
	GL	300
	Significancia	<.0001

Fuente: elaboración propia.

Observando la medida de suficiencia muestral (MSA) para cada una de las variables (tabla 3), se encuentra que esta no es significativa únicamente para la longitud promedio de la oración (0,45), por lo que esta variable es eliminada del análisis. Asimismo, se observa una comunalidad – varianza común o compartida con las demás variables– ini-

cial estimada muy baja para el número promedio de sílabas (0,37), razón por la cual esta variable también es retirada del análisis. La supresión de estas variables, en modo alguno afecta la adecuación de los datos; su idoneidad es corroborada al recalcular la medida de suficiencia muestral y la prueba de esfericidad de Bartlett.

Tabla 3. Medida de Kaiser de suficiencia muestral por variable (MSA) y comunalidades iniciales

Variable	Descripción	Msa	Comunalidad Inicial
Pbras	Cantidad de palabras	0,89853	0,96640
Rglns	Cantidad de renglones	0,94138	0,88822
Orac_tot	No. total de oraciones	0,85725	0,97713
Orac_sim	No. de oraciones simples	0,87192	0,94937
Orac_coor	No. de oraciones coordinadas	0,63052	0,63817
Parraf	No. de párrafos	0,79164	0,79978
Sust_solos	No. de sustantivos solos	0,80643	0,94928
Sust_adj	No. de sustantivos con adjetivos	0,90083	0,92852
Sust_orac_rel	No. de sustantivos con oración relativa	0,75496	0,82005
Sust_sintag_prep	No. de sustantivos con sintagma preposicional	0,89730	0,87597
Verbos_ele	No. de verbos utilizados (Elementos)	0,82904	0,99995
Compl_ele	No. de complementos (Elementos)	0,88868	0,99638
Long_prom_orac	Longitud promedio de la oración	0,45049	0,60369
Long_prom_pbra	Longitud promedio de la palabra	0,66433	0,69733
Num_prom_sil	Número promedio de sílabas	0,69544	0,36956
Sust_abst	No. de sustantivos abstractos	0,75189	0,96190
Sust_conc	No. de sustantivos concretos	0,60095	0,89660
Porc_sust_nuev	Proporción de sustantivos nuevos	0,76890	0,73086

Porc_adjet_nuev	Proporción de adjetivos nuevos	0,78528	0,57965
Porc_verb_nuev	Proporción de verbos nuevos	0,79013	0,58224
Marc_coh	No. de marcas de cohesión	0,84051	0,74238
Verbos_est	No. de verbos utilizados (Estructura)	0,82550	0,99995
Compl_est	No. de complementos (Estructura)	0,88664	0,99638
TTR	Relación tipo-token	0,84140	0,84617
Div_lex	Diversidad léxica	0,81005	0,63449

Fuente: elaboración propia.

Número de factores a extraer

El uso del criterio de los autovalores indica que tres factores deben ser extraídos (ver tabla 4); estos explican un 85,4 % del total de la varianza, un porcentaje significativo, por lo que también se cumple a cabalidad el criterio de proporción de varianza. De acuerdo con el gráfico de sedimentación (figura 1)

se podría extraer un factor más que correspondería al cuarto autovalor, 0,71, un valor no muy cercano a 1, y que presenta, además, un aporte a la varianza no significativo (0,037). Por consiguiente, con el propósito de satisfacer el principio de parsimonia y facilitar la interpretación de los resultados se retienen tres factores.

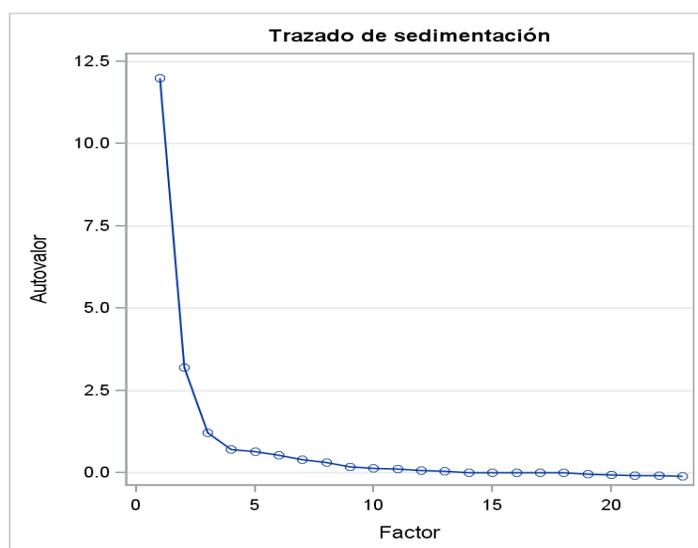


Figura 1. Gráfica para aplicar el criterio de autovalores

Fuente: elaboración propia.

Tabla 4. Autovalores de la matriz de correlación

Autovalores preliminares: Total = 19.2212475 Promedio = 0.83570641				
	Autovalor	Diferencia	Proporción	Acumulada
1	12,0017766	8,8067294	0,6244	0,6244
2	3,1950472	1,9830858	0,1662	0,7906
3	1,2119614	0,5041155	0,0631	0,8537
4	0,7078459	0,0635059	0,0368	0,8905
5	0,6443400	0,1009656	0,0335	0,9240
6	0,5433744	0,1345899	0,0283	0,9523
7	0,4087845	0,1034838	0,0213	0,9736
8	0,3053007	0,1347150	0,0159	0,9894
9	0,1705857	0,0351121	0,0089	0,9983
10	0,1354736	0,0126231	0,0070	1,0054
11	0,1228505	0,0420518	0,0064	1,0118
12	0,0807987	0,0311455	0,0042	1,0160
13	0,0496532	0,0396314	0,0026	1,0185
14	0,0100218	0,0067271	0,0005	1,0191
15	0,0032946	0,0033605	0,0002	1,0192
16	-0,0000659	0,0011787	0,0000	1,0192
17	-0,0012446	0,0031182	-0,0001	1,0192
18	-0,0043628	0,0294160	-0,0002	1,0189
19	-0,0337788	0,0196785	-0,0018	1,0172
20	-0,0534572	0,0250488	-0,0028	1,0144
21	-0,0785061	0,0110545	-0,0041	1,0103
22	-0,0895606	0,0193249	-0,0047	1,0057
23	-0,1088855		-0,0057	1,0000

Fuente: elaboración propia.

Modelo factorial de rotación

Como es común en este tipo de análisis, de los resultados obtenidos para el modelo factorial sin rotación no se observa una estructura simple. En primer lugar, ninguna variable está correlacionada significativamente –carga factorial mayor a 0,5– con el factor 3; y, adicionalmente, la variable relación tipo *token* (TTR) tiene cargas estadísticamente significativas y de signos contrarios sobre los factores 1 y 2, -0,66257 y 0,57648, respectivamente. Esta falta de interpretabilidad llevó a la estimación de un modelo factorial con rotación¹ (tabla 5), cuyos resultados

satisfacen los requerimientos de una estructura factorial simple, es decir, cada variable cargando significativamente a un solo factor y las cargas en los otros factores no significativas mostrando, en general, valores pequeños; adicionalmente, los factores presentan distribuciones diferentes para las cargas altas y bajas. Puesto que las variables *número de sustantivos con relación relativa* y *número de oraciones coordinadas* no cargan significativamente a ningún factor no se reportan en la matriz de factores rotados.

¹ Se usó una rotación oblicua, lo que implica que los factores no tienen por qué ser ortogonales entre sí. Se aplicó el método Promax que consiste en modificar los resultados de una rotación ortogonal hasta crear una solución con cargas factoriales lo más cercanas a la estructura ideal.

Tabla 5. Modelo factorial de rotación

		Factor1	Factor2	Factor3
Parraf	No. de párrafos	0,93338		
Orac_sim	No. de oraciones simples	0,79109		
Pbras	Cantidad de palabras	0,75454		
Sust_solos	No. de sustantivos solos	0,72418		
Rglhs	Cantidad de renglones	0,68991		
Sust_sintag_prep	No. de sustantivos con sintagma preposicional	0,67967		
Marc_coh	No. de marcas de cohesión	0,63776		
Sust_adj	No. de sustantivos con adjetivos	0,59579		
Sust_abst	No. de sustantivos abstractos	0,57942		
Orac_tot	No. total de oraciones	0,53638		
TTR	Relación tipo-token	-0,72542		
Verbos_ele	No. de verbos utilizados (Elementos)		0,71850	
Verbos_est	No. de verbos utilizados (Estructura)		0,71801	
Sust_conc	No. de sustantivos concretos		0,63302	
Div_lex	Diversidad léxica		0,63036	
Compl_est	No. de complementos (Estructura)		0,62689	
Compl_ele	No. de complementos (Elementos)		0,61875	
Porc_verb_nuev	Proporción de verbos nuevos			0,78946
Porc_sust_nuev	Proporción de sustantivos nuevos			0,75647
Porc_adjet_nuev	Proporción de adjetivos nuevos			0,67155
Long_prom_pbra	Longitud promedio de la palabra			0,54167

Fuente: elaboración propia.

Por último, siguiendo la metodología propuesta, es necesario anotar que se realizaron varios ejercicios en los cuales se aumentó o disminuyó el número de factores, se emplearon diferentes métodos de rotación (ortogonales y oblicuos) y se eliminaron variables sin obtenerse estructuras de factores más simples ni con interpretaciones más plausibles.

El factor 1 puede interpretarse como un factor de tamaño; donde, salvo relación tipo token (TTR), las variables muestran una relación directa (positiva) con el factor. Extraña, sin embargo, que la cantidad de verbos (número de verbos utilizados-estructura) y de complementos verbales (est) aparezcan en un segundo factor, los cuales se esperarían en el primer factor si este se asocia con cantidad. La mayoría de rasgos del segundo factor tiene que ver con el verbo y complementos verbales oracionales. Es interesante

este hallazgo en tanto es sabido la importancia del verbo como núcleo generador no solo de estructuras sintácticas (valencias) sino de roles semánticos. La diversidad léxica y el rasgo de sustantivos concretos llaman la atención en este grupo. No obstante, el factor es unitario desde el principio de la prototipicidad: los textos descriptivos-explicativos escolares típicamente presenta un número importante de verbos y complementos verbales, variedad de ellos, sustantivos concretos (predominio en el corpus analizado) y diversidad léxica. El factor 3 resulta más unitario en tanto agrupa tres categorías y su relación con información nueva. En este grupo el único rasgo no esperable es la longitud promedio de la palabra, que podría asociarse más con la prototipicidad, como rasgo discriminatorio del tipo de texto. Sin embargo, la longitud promedio de la palabra de la

muestra está por encima de 0,5, lo que da cuenta de palabras léxicas y no gramaticales (artículo, palabras de relación), es decir, palabras nocionales.

Un índice de complejidad lingüística

En la tabla 6 se presentan las estadísticas básicas para el índice estimado, de acuerdo con lo expresado

en la ecuación (2), según nivel de complejidad. El valor promedio del índice para el conjunto de textos es 0,344 con una desviación estándar de 0,166. La distribución por nivel es asimétrica a derecha con 38 textos clasificados como de nivel muy bajo, 28 en nivel bajo, 10 en nivel medio y apenas 4 en nivel alto.

Nivel de complejidad	Índice				
	No. Textos	Mínimo	Máximo	Media	Desviación Estándar
Total	80	0,140	1,000	0,344	0,166
Muy bajo	38	0,140	0,283	0,222	0,043
Bajo	28	0,287	0,469	0,371	0,053
Medio	10	0,481	0,618	0,528	0,048
Alto	4	0,779	1,000	0,860	0,097

Tabla 6. Índice de complejidad lingüística

Fuente: elaboración propia.

De otra parte, cuando se analiza la distribución por áreas y nivel de complejidad (tabla 7), se encuentra que es el área de matemáticas la que presenta una mayor frecuencia de textos con un nivel muy bajo de complejidad, un total de 13, correspondientes al 32, 2 % de los textos en este nivel; ciencias con 9 textos,

32, 1 %, es el área con más textos de complejidad baja, mientras que sociales con el 40 % de los textos, es el área con más textos de complejidad media. En el nivel de complejidad alto solo clasifican textos del área de lenguaje.

Nivel de complejidad	Total textos	%	Área											
			Ciencias			Lenguaje			Matemáticas			Sociales		
			Total	% HOR.	% VER.	Total	% HOR.	% VER.	Total	% HOR.	% VER.	Total	% HOR.	% VER.
Total	80	100,0	20	25,0	100,0	20	25,0	100,0	20	25,0	100,0	20	25,0	100,0
Muy bajo	38	47,5	8	21,1	40,0	8	21,1	40,0	13	34,2	65,0	9	23,7	45,0
Bajo	28	35,0	9	32,1	45,0	6	21,4	30,0	6	21,4	30,0	7	25,0	35,0
Medio	10	12,5	3	30,0	15,0	2	20,0	10,0	1	10,0	5,0	4	40,0	20,0
Alto	4	5,0				4	100,0	20,0						

Nota. % HOR indica participaciones por fila (nivel de complejidad), mientras que %VER hace referencia a participaciones por columna (área)

Tabla 7. Clasificación de textos por área según nivel de complejidad lingüística

Fuente: elaboración propia.

Si se fija el área en la tabla 7, se observa que, en general, el número de textos disminuye a medida que crece el nivel de complejidad; así, por ejemplo, en matemáticas las cifras ascienden a 13 (65 %), 6 (30 %) y 1 (5 %) para niveles de complejidad muy bajo, bajo y medio, respectivamente; el mismo patrón se presenta en el área de sociales. En lenguaje, la diferencia obedece a que esta área presenta los únicos textos con un nivel de complejidad alto; mientras que, en ciencias, el patrón varía ligeramente pues el número de textos es mayor en el nivel de complejidad bajo comparado con el nivel muy bajo. Es importante aclarar que los textos de lenguaje ubicados en el nivel alto son textos que describen movimientos literarios con bastante información relacionada con autores, fechas y lugares geográficos, es decir, densos semánticamente y de una longitud considerable. En el corpus analizado, los textos de matemáticas suelen ser cortos, lo que implica menos cantidad y variedad de información y de estructuras.

En la tabla 8 se presenta la distribución por grados y nivel de complejidad. Se advierte, para el nivel muy bajo, que a medida que aumenta el grado disminuye de manera sistemática el número de textos, pasando de 14 (36,8 %) en el grado 3.º a 1 (2,6 %) en el grado 11.º; esta tendencia se invierte en el nivel de complejidad bajo, donde el número de textos aumenta de 2 (7,1 %) en el grado 3.º a 9 (32,1 %) en el grado 11.º. Textos de nivel medio de complejidad se registran en los grados 7.º, 9.º y 11.º, mientras que los de complejidad alta solo aparecen en los grados 9.º y 11.º. Textos de nivel medio de complejidad se registran en los grados 7.º y 9.º con una participación de 18,8 %, y, 11.º con una participación de 25,0 %; mientras que los de complejidad alta solo aparecen en los grados 9.º y 11.º, con idéntica participación, 12,5 %.

Nivel de complejidad	Total textos	%	Grado														
			3			5			7			9			11		
			Total	% HOR.	% VER.												
Total	80	100,0	16	20,0	100,0	16	20,0	20,0	100,0	16	20,0	20,0	100,0	16	20,0	20,0	100,0
Muy bajo	38	47,5	14	36,8	87,5	12	31,6	75,0	37,5	6	15,8	13,2	31,3	5	13,2	2,6	6,3
Bajo	28	35,0	2	7,1	12,5	4	14,3	25,0	43,8	7	25,0	21,4	37,5	6	21,4	32,1	56,3
Medio	10	12,5				3	30,0	18,8	18,8			30,0	18,8	3	30,0	40,0	25,0
Alto	4	5,0												2	50,0	50,0	12,5

Nota. % HOR indica participaciones por fila (nivel de complejidad), mientras que %VER hace referencia a participaciones por columna (Grado)

Fuente: elaboración propia.

Para finalizar, se observa, en la lectura vertical, que en los grados 7.º, 9.º y 11.º, los textos con mayor participación corresponden al nivel de complejidad bajo: 43, 8 %, 37, 5 % y 56, 3 %, respectivamente; en tanto que para los grados de básica primaria la mayor participación la tienen los textos clasificados en el nivel de complejidad muy bajo: 87, 5 % y 75 %.

Discusión y conclusiones

Los resultados de la investigación muestran que la cantidad (Bulté y Housen, 2012; Kusters, 2008; Rescher, 1998), tanto de formas (Kurdi, 2017; Lahuerta, 2017) como de contenidos (Plakans y Bilki, 2016), es un factor importante para determinar la complejidad lingüística.

El factor *variedad*, reportado por Vyatkina (2013) y Ochoa (2016) resultó productivo para explicar algunos de los rasgos pero no un factor como totalidad; la *propotipicidad* es más significativa ya que permite explicar la mayoría de rasgos asociados con el factor 2, lo mismo se puede decir de la variable que tentativamente podría denominarse *informatividad*, asociada con el factor 3.

En relación con los rasgos, se encontró que los aspectos sintácticos por sí solos no son suficientes para medir la complejidad. Esto ya había sido señalado por autores como Bunch et ál., (2014) y Ochoa (2016).

Como se señaló en la metodología, se excluyeron algunos rasgos en el primer análisis del corpus y otros fueron excluidos por el análisis factorial. Esto no implica que no sean válidos para medir la complejidad, sino que no resultaron productivos para los textos descriptivos-expositivos escolares o al menos para el corpus elegido. Habría que evaluar otros tipos de textos, idea que ya fue presentada por Frantz, Starr y Bailey (2015) y Ortega (2015), entre otros autores.

En esta investigación se intentó medir la complejidad en unidades superiores a la palabra, frase u oración. Medir la complejidad únicamente a partir de estas últimas unidades (Lu, 2010; Mancilla, Polat y Akcay, 2017) resulta limitado, ya que es evidente

que un texto es un entramado complejo que involucra unidades de orden mayor. En ese sentido, rasgos como *diversidad léxica*, *TTR*, *marcas de cohesión*, etc., que involucran al texto como totalidad, deben ser considerados.

El análisis de los datos nos muestra que hay unos criterios generales que son independientes del tipo de texto, como: cantidad de palabras; cantidad de renglones; longitud promedio de la palabra; vocabulario abstracto. Sin embargo, la aplicación de los criterios sí permite caracterizar tipos de texto y sus niveles de complejidad. Así, se puede concluir que los textos descriptivos-explicativos escolares complejos típicamente presentan un número importante de verbos y variedad de ellos (no solo verbos como *ser* o *estar*, *tener*, *haber* o *consistir*), complementos verbales variados, sustantivos acompañados de modificadores, información nueva expresada a través de sustantivos, adjetivos y verbos.

En relación con la complejidad de los textos analizados, se observa una coincidencia con hallazgos como los de Berendes et al. (2018), por cuanto se halló un gran número de textos que se ubican en los dos primeros niveles y la mayoría de textos de los grados séptimo y once se ubican en el nivel bajo. Este hallazgo es preocupante si se considera que la lectura es la principal fuente de información de los aprendices y un instrumento ideal para modificar sus conocimientos previos e incrementar su capital cultural.

Aunque en la bibliografía se ha evidenciado que a mayor complejidad textual mayor tiempo de lectura y viceversa (Graesser et ál., 2004), y que textos que contienen pocas palabras, léxico frecuente, estructuras sintácticas simples, alta cohesión y conexiones explícitas, son más fáciles de procesar que los que tienen baja cohesión (McNamara y Shapiro, 2005), se sabe también que textos que no retan cognitivamente al estudiante (sin ser excesivos) resultan adversos a los procesos de aprendizaje. Por tanto, no es conveniente asignar textos con una complejidad baja en grados y ciclos escolares superiores e incluso en grados medios. Y ello porque, siguiendo el concepto de zona de desarrollo próxima de Vygotsky

(1979), es necesario lograr con la ayuda del profesor y los pares que los estudiantes salgan del nivel real de desarrollo y alcancen su nivel potencial. Esto se logra en relación con la lectura, evaluando y seleccionado con cuidado los textos que se asignan para cada grado escolar, implementando de forma consciente y decidida una estrategia didáctica que fomente el uso de estrategias cognitivas y metacognitivas, y retroalimentando de manera constructiva los procesos de comprensión lectora.

Frente a las áreas, la ubicación de los textos de matemáticas y lenguaje resulta inesperada en tanto los primeros se conciben generalmente como textos más complejos que los segundos. Al respecto, es importante señalar que un criterio que no se tuvo en cuenta en esta investigación es el relacionado con la información implícita y, por ende, sobreentendida que caracteriza los textos que resultaron ubicados en el nivel alto. Posiblemente los textos de matemáticas sean sencillos en términos morfosintácticos y semánticos, pero no a nivel pragmático. La presuposición parece ser un rasgo que amerita ser tenido en cuenta para futuras investigaciones.

Es preciso indagar más sobre los criterios y sobre las razones por las que un análisis factorial los agrupa en uno u otro factor. Las hipótesis aquí presentadas requieren ser investigadas a profundidad. Igualmente, los rasgos deben aún validarse y en ese sentido las conclusiones son aún objeto de revisión. El valor de esta investigación consiste en haber hecho un análisis estadístico que puede convertirse en un punto de partida para otros estudios.

Referencias

- Adam, J. M. (1992). *Les textes: types et prototypes*. Nathan.
- Andrews, S. J. y Lamb, M. E. (2017). The structural linguistic complexity of lawyers' questions and children's responses in Scottish criminal courts. *Child Abuse and Neglect*, 65, 182-193.
- Aravena, S. y Hugo, E. (2016). Desarrollo de la complejidad sintáctica en textos narrativos y explicativos escritos por estudiantes secundarios. *Lenguas Modernas*, 47, 9-40.

- Berasain, M. J. (s. f.). *La descripción: análisis y producción de textos descriptivos*. Universidad Complutense de Madrid.
- Berendes, K., Vajjala, S., Meurers, D., Bryant, D., Wagner, W., Chinkina, M. y Trautwein, U. (2018). Reading demands in secondary school: Does the linguistic complexity of textbooks increase with grade level and the academic orientation of the school track? *Journal of Educational Psychology*, 110(4), 518-543. doi.org/10.1037/edu0000225
- Bulté, B. y Housen, A. (2014). Conceptualizing and measuring short-term changes in L2 writing complexity. *Journal of Second Language Writing*, 26, 42-65.
- Bunch, G., Walqui, A. y Pearson, P. (2014). Complex text and new common standards in The United States: Pedagogical implications for English learners. *Tesol Quarterly*, 48(3), 533-559.
- Campos, D., Contreras, P., Riffo, B., Véliz, M. y Reyes, A. (2014). Complejidad textual, lecturabilidad y rendimiento lector en una prueba de comprensión en escolares adolescentes. *Universitas Psychologica*, 13(3), 1135-1146. [https:// dx.doi.org/10.11144/Javeriana.UPSY13-3.ctrl](https://dx.doi.org/10.11144/Javeriana.UPSY13-3.ctrl)
- Coh-Metrix. (s. f.). *Coh-Metrix version 3.0 Indices*. http://cohmetrix.memphis.edu/cohmetrixhome/documentation_indices.html#:~:text=Coh%2DMetrix%20version%203.0%20indices&text=Coh%2DMetrix%20is%20a%20computational,mental%20representation%20of%20the%20text
- Cunningham, J., Hiebert, E. y Mesmer, H. (2012). Toward a theoretical model of text complexity for the early grades: learning from the past, anticipating the future. *Reading Research Quarterly*, 47(3), 235-258.
- De Marneffe, M. C., Manning, C. D. y Potts, C. (2012). Did it happen? The pragmatic complexity of veridicality assessment. *Computational Linguistics*, 38(2), 301-333.
- Díaz, L. G. y Morales, M. A., (2016). *Análisis estadístico de datos multivariados*. Bogotá: Universidad Nacional de Colombia, Facultad de Ciencias.
- Disselhoff, K. (2007). Einzelanalyse der textsorte beschreibung. En G. Augst et al., *Text-sorten-kompetenz. Eine echte longitudinalstudie zur entwicklung der textkompetenz im grudschulalter* (pp. 167-197). Peter Lang.
- Feilke, H. (2017). Beschreiben. En J. Baurmann, C. Kammler y A. Müller (eds.), *Handbuch deutschun-*

- terricht. Theorie und praxis des lehrens und lernens* (pp. 71-75). Klett-Kallmeyer.
- Frantz, R., Starr, L. y Bailey, A. (2015). Syntactic complexity as an aspect of text complexity. *Educational Researcher*, 44(7), 387-393.
- Graesser, A. y McNamara, D. S. (2011). Computational analyses of multilevel discourse comprehension. *Topics in Cognitive Science*, 3, 371-398.
- Graesser, A., McNamara, D., Lowerse M. y Cai, Z. (2004). Coh-Metrix: analysis of text on cohesion and language. *Behavior Research Methods, Instruments, & Computers*, 36(2), 193-202.
- Hair, J., Anderson, R., Tatham, R. y Black, W. (1999). *Análisis multivariante*. Pearson, Prentice Hall.
- Hoek, J., Zufferey, S., Evers-Vermeul, J. y Sanders, T. (2017). Cognitive complexity and the linguistic marking of coherence relations: a parallel corpus study. *Journal of Pragmatics*, 121, 113-131.
- Instituto Cervantes. (2006). *Plan Curricular del Instituto Cervantes. Niveles de referencia para el español*. Tomos i, ii y iii. Madrid: Instituto Cervantes.
- Juola, P. (2008). Assessing linguistic complexity. En M. Miestamo, K. Sinnemäki y F. Karlsson (eds.), *Language complexity: typology, contact, change* (pp. 89-108). Amsterdam: John Benjamins Publishing Company. doi: <https://doi.org/10.1075/slcs.94.07juo>
- Kaiser, H. (1974). An index of factorial simplicity. *Psychometrika*, 39, 31-36.
- KoRpus (2017). *Package 'koRpus'*. Recuperado de <https://cran.r-project.org/web/packages/koRpus/koRpus.pdf>
- Kurdi, M. (2017). Lexical and syntactic features selection for an adaptive reading recommendation system based on text complexity. En *ICISDM '17 Proceedings of the 2017. International Conference on Information System and Data Mining* (pp. 66-69). Nueva York: Association for Computing Machinery.
- Kusters, W. (2008). Complexity in linguistic theory, language learning and language change. En M. Miestamo, K. Sinnemäki y F. Karlsson (eds.), *Studies in language companion. Series 94* (pp. 3-22). John Benjamins Publishing Company.
- Lahuerta, A. C. (2017). Syntactic complexity in secondary-level English writing: differences among writers enrolled on bilingual and non-bilingual programmes. *Porta Linguarum*, 28, 67-80.
- Lexile. (s. f.). *Lexile. Framework for reading*. <https://lexile.com/>
- Lu, X. (2010). Automatic analysis of syntactic complexity in second language writing. *International Journal of Corpus Linguistics*, 15(2), 474-496.
- Lu, X. y Ai, H. (2015). Syntactic complexity in college-level English writing: differences among writers with diverse L1 backgrounds. *Journal of Second Language Writing*, 29, 16-27.
- Mancilla, R., Polat, N. y Akcay, A. (2017). An investigation of native and nonnative English speakers' levels of written syntactic complexity in asynchronous online discussions. *Applied Linguistics*, 38(1), 112-134.
- McNamara, D. (2004). Aprender del texto: efectos de la estructura textual y las estrategias del lector. *Revista Signos*, 37(55), 1-12.
- McNamara, D. y Shapiro, A. (2005). Multimedia and hipermedia solutions for promoting metacognitive engagement, coherence, and learning. *Journal Educational Computing Research*, 33(1), 1-29.
- Meneses, A. y Ow, M. (2016). Syntactic development in Chilean students' narrative and expository discourse throughout schooling: different measures of syntactic complexity. *Estudios de Psicología*, 37(1), 135-161.
- Meneses, A., Ow, M. y Benítez, R. (2012). Complejidad sintáctica: ¿Modalidad o tipo de texto? Estudio de casos de producciones textuales de estudiantes de 5° básico. *Onomázein*, 25, 65-93.
- Nini, A. (2015). *MAT - Multidimensional Analysis Tagger (versión 1.3)*. <http://sites.google.com/site/multidimensionaltagger>.
- Nippold, M., Frantz-Kaspar, M. y Vigeland, L. (2017). Spoken language production in young adults: examining syntactic complexity. *Journal of Speech, Language, and Hearing Research*, 60(5), 1339-1347.
- Norris, J. y Ortega, L. (2009). Towards an organic approach to investigating caf in instructed sla: the case of complexity. *Applied Linguistics*, 30(4), 555-578.
- Ochoa, L. (2016). *Estudio sobre los niveles de dificultad de los textos que conforman las pruebas Saber 3.º, 5.º y 9.º en el área de Lenguaje*. [Sin publicar]. Icfes.
- Ortega, L. (2015). Syntactic complexity in L2 writing: progress and expansion. *Journal of Second Language Writing*, 29, 82-94.
- Pallotti, G. (2014). A simple view of linguistic complexity. *Second Language Research*, 31(1), 117-134.

- Pallotti, G. (2015). A simple view of linguistic complexity. *Second Language Research*, 31, 117-134.
- Peña, D. (2002). *Análisis de datos multivariantes*. McGraw-Hill.
- Plakans, L. y Bilki, Z. (2016). Cohesion features in esl reading: comparing beginning, intermediate and advanced textbooks. *Reading in a Foreign Language*, 28(1), 79-100.
- Polio, C. y Yoon, H. (2018). The reliability and validity of automated tools for examining variation in syntactic complexity across genres. *International Journal of Applied Linguistics*, 28, 165-188.
- Rescher, N. (1998). *Complexity: a philosophical overview*. Transaction Publishers.
- Schilk, M. y Schaub, S. (2016). Noun phrase complexity across varieties of English: focus on syntactic function and text type. *English World-Wide*, 37(1), 58-85.
- Sheehan, K. M., Flor, M., Napolitano, D. y Ramineni, C. (2015). *Using TextEvaluator® to quantify sources of linguistic complexity in textbooks targeted at first-grade readers over the past half century* (Research Report No. RR-15-38). Educational Testing Service. <http://dx.doi.org/10.1002/ets2.12085>
- Solnyshkina, M., Zamaletdinov, R., Gorodetskaya, L. y Gabitov, A. (2017). Evaluating text complexity and Flesch-Kincaid grade level. *Journal of Social Studies Education Research*, 8(3), 238-248.
- Van Silfhout, G., Evers-Vermeul, J. y Sanders, T. (2015). Connectives as processing signals: how students benefit in processing narrative and expository texts. *Discourse Processes*, 52(1), 47-76.
- Vyatkina, N. (2013). Specific syntactic complexity: developmental profiling of individuals based on an annotated learner corpus. *The Modern Language Journal*, 97(S1), 11-30.
- Vygotsky, L. S. (1979). *El desarrollo de las funciones psicológicas superiores*. Grijalbo.
- Werlich, E. (1976). *Typologie der texte*. Quelle & Meyer.
- Yang, W., Lu, X. y Weigle, S. (2015). Different topics, different discourse: relationships among writing topic, measures of syntactic complexity, and judgments of writing quality. *Journal of Second Language Writing*, 28, 53-67.
- Yoon, H. J. (2017). Linguistic complexity in L2 writing revisited: issues of topic, proficiency, and construct multidimensionality. *System*, 66, 130-141.
- Youn, S. (2014). Measuring syntactic complexity in L2 pragmatic production: Investigating relationships among pragmatics, grammar, and proficiency. *System*, 42, 270-287.
- Zunino, G. (2017). Processing causality and counter-causality: interactions between syntactic structure and world knowledge during the comprehension of semantic relations. *Revista Signos*, 50(95), 472-491.