


# Complejidad estructural en textos narrativos\*


LIGIA OCHOA SIERRA\*\*

Recepción: 18 de junio de 2021  
Aprobación: 3 de agosto de 2021

**Forma de citar este artículo:** Ochia, L. (2021). Complejidad estructural en textos narrativos. *Cuadernos de Lingüística Hispánica*, (38), e13002

 <https://10.19053/0121053X.n38.2021.13002>

\* Artículo de investigación.

\*\* Doctora en Gramática General y Comparada. Universidad Autónoma de Madrid. Profesora titular de la Universidad Nacional de Colombia. Colombia. Grupo de Investigación: Lingüística y Educación. Esta investigación fue financiada completamente por la Universidad Nacional de Colombia. lochoas@unal.edu.co  <https://orcid.org/0000-0002-8761-4819>

## Resumen

La narración es una secuencia textual que aparece de manera transversal en los distintos grados y ciclos escolares; sin embargo, no hay estudios que ofrezcan un conjunto de criterios que permitan determinar su complejidad, a fin de poder escalonar adecuadamente los textos donde prima dicha secuencia. En esta investigación, de corte cuantitativo con un diseño descriptivo-interpretativo, se indagó acerca de las variables que establecen la complejidad estructural inherente a 65 textos escolares narrativos. Con tal fin, se analizaron los textos con base en 20 variables pertenecientes a los niveles sintáctico, semántico y pragmático. Dichos textos fueron clasificados por expertos lingüistas en cuatro niveles de complejidad. Se aplicó un análisis estadístico para responder la pregunta central: ¿cuáles variables y criterios lingüísticos determinan la complejidad estructural del corpus objeto de análisis? Los resultados muestran que los factores de cantidad y variedad con variables predominantemente semánticas y pragmáticas posibilitan predecir la complejidad de los textos analizados. Estos resultados enriquecen la discusión teórica sobre los criterios que permiten medir la complejidad de un texto y pueden ser un insumo para futuras investigaciones tanto en el campo de la lingüística aplicada como en el de la educación.

**Palabras clave:** texto, narración, lingüística.

## Structural Complexity in Narrative Texts

### Abstract

Storytelling is a textual sequence that appears transversally in the different school grades and cycles; however, no studies providing a set of criteria to determine its complexity, in order to adequately stagger the texts where this sequence prevails. In this quantitative research with a descriptive-interpretive design, the variables that establish the structural complexity inherent in 65 narrative school texts were investigated. To this end, the texts were analyzed based on 20 variables belonging to the syntactic, semantic and pragmatic levels. These texts were classified by linguistic experts into four levels of complexity. A statistical analysis was applied to answer the central question: which linguistic variables and criteria determine the structural complexity of the corpus under analysis? The results show that the quantity and variety factors with predominantly semantic and pragmatic variables enable to predict the complexity of the analyzed texts. These results enrich the theoretical discussion on the criteria that allow measuring the complexity of a text and can be an input for future research both in the field of applied linguistics and in education.

**Keywords:** text, storytelling, linguistics.

## La complexité structurelle des textes narratifs

### Résumé

Le récit est une séquence textuelle qui apparaît de manière transversale dans les différents niveaux et cycles scolaires ; cependant, il n'existe pas d'études qui offrent un ensemble de critères permettant de déterminer sa complexité, afin de pouvoir échelonner de manière adéquate les textes où cette séquence prévaut. Dans cette recherche, de nature quantitative avec une approche descriptive-interprétative, nous avons étudié les variables qui établissent la complexité structurelle inhérente à 65 textes scolaires narratifs. A cette fin, les textes ont été analysés sur la base de 20 variables appartenant aux niveaux syntaxique, sémantique et pragmatique. Les textes ont été classés par des linguistes experts en quatre niveaux de complexité. Une analyse statistique a été appliquée pour répondre à la question centrale : quelles variables et quels critères linguistiques déterminent la complexité structurelle du corpus analysé ? Les résultats montrent que les facteurs de quantité et de variété avec des variables principalement sémantiques et pragmatiques permettent de prédire la complexité des textes analysés. Ces résultats enrichissent la discussion théorique sur les critères de mesure de la complexité d'un texte et peuvent constituer un apport pour de futures recherches dans le domaine de la linguistique appliquée ainsi que dans l'enseignement.

**Mots-clés:** texte, récit, linguistique.

## Complexidade estrutural em textos narrativos

### Resumo

A narração é uma sequência textual que aparece transversalmente nas diferentes séries e ciclos escolares; No entanto, não existem estudos que ofereçam um conjunto de critérios que permitam determinar a sua complexidade, de forma a poder escalonar adequadamente os textos onde prevalece a referida sequência. Nesta pesquisa quantitativa de desenho descritivo-interpretativo, foram investigadas as variáveis que estabelecem a complexidade estrutural inerente a 65 textos escolares narrativos. Para tanto, os textos foram analisados a partir de 20 variáveis pertencentes aos níveis sintático, semântico e pragmático. Esses textos foram classificados por especialistas em linguística em quatro níveis de complexidade. Uma análise estatística foi aplicada para responder à questão central: quais variáveis e critérios linguísticos determinam a complexidade estrutural do corpus em análise? Os resultados mostram que os fatores quantidade e variedade com variáveis predominantemente semânticas e pragmáticas permitem prever a complexidade dos textos analisados. Esses resultados enriquecem a discussão teórica sobre os critérios que permitem mensurar a complexidade de um texto e podem servir de subsídio para pesquisas futuras tanto no campo da linguística aplicada quanto na educação.

**Palavras-chave:** texto, narração, linguística.

## Introducción

Los conceptos de género, modo y secuencia textual resultan problemáticos, pues en ocasiones se asemejan entre sí y en otras se usan para clasificar textos sin el adecuado rigor académico. En esta investigación se empleará el concepto de secuencia textual, tal como ha sido definido por Adam (1992). Según este investigador, una secuencia es una estructura que tiene un propósito comunicativo determinado. A partir de esta consideración, Adam (1992) identifica cinco secuencias: narrativa, descriptiva, argumentativa, explicativa y dialogal.

Dentro de un texto es posible encontrar diversas secuencias con funciones comunicativas distintas, pero a la vez es posible identificar una secuencia textual dominante y otras secundarias. Con base en la secuencia textual predominante se puede hablar de textos narrativos, descriptivos, argumentativos, expositivos o dialogales. Un texto donde predomine la secuencia narrativa, como la autobiografía, puede considerarse un texto narrativo. Este fue el principio asumido en esta investigación.

Para Adam (1992), los textos narrativos son aquellos que informan sobre acciones y responden a la pregunta ¿qué pasó? En este tipo de textos abunda la predicación y los verbos de acción, y se destaca la estructura temporal. Esta secuencia discursiva se encuentra en cuentos, novelas, noticias, reportajes. Según Werlich (1975), las secuencias narrativas se evidencian en formas objetivas como el informe y en formas subjetivas como la narración oral.

Narrar es una de las secuencias más usadas por el ser humano. Se narran los sucesos que ocurren a diario, aquellos acaecidos en épocas pasadas y los que pueden venir en el futuro. Se narran tanto hechos reales como ficticios, y en estos últimos se crean universos de sentido alternos. En el ámbito educativo, los textos donde predominan las secuencias narrativas se leen y escriben a lo largo de todos los grados del periodo de escolaridad. Si bien puede afirmarse que es la secuencia textual más usada en el ámbito escolar, no se dispone de estudios consistentes que permitan medir la complejidad de los textos donde predomina dicha secuencia.

La necesidad de analizar la complejidad en secuencias discursivas particulares ha sido señalada en diversos trabajos, en cuanto distintas tipologías suponen medidas de complejidad diferentes (Frantz et al., 2015; Ortega, 2015; Ochoa, 2016); así, por ejemplo, criterios como la “conectividad” y la “subordinación”, entre otros, parecen variar entre secuencias (Aravena & Hugo, 2016).

En la revisión bibliográfica para el presente trabajo se encontraron algunas investigaciones que, al centrar su atención en textos discursivos particulares, dieron cuenta de la complejidad de los textos narrativos en comparación con textos donde prevalecen otras secuencias. Así, la argumentación (Yang et al., 2015; Polio & Yoon, 2018) y el discurso académico aparecen como los tipos textuales más complejos, en tanto que, en el polo opuesto, se ubica el discurso cotidiano; el texto narrativo se

hallaría en un punto de complejidad media (Nini, 2015). Según Meneses *et al.* (2012), a partir de un estudio longitudinal con estudiantes chilenos, los textos *narrativos orales* se oponen a los *expositivos escritos* en términos de longitud, diversidad léxica y relaciones intraclausales. Por su parte, Nippold *et al.* (2017) compararon textos narrativos y textos de pensamiento crítico que involucraban preguntas y respuestas con conversaciones y textos de pensamiento crítico producidos por adultos jóvenes, y concluyeron que los dos primeros textos tenían una complejidad sintáctica mayor que la tarea conversacional y que el primer texto era más complejo sintácticamente que el texto de pensamiento crítico.

Por otra parte, se han investigado, aunque en menor medida, los *niveles de complejidad* de distintos textos. Por ejemplo, Sheehan *et al.* (2015) indagaron si los libros de texto pensados para niños estadounidenses de primer grado y producidos por una editorial específica entre 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 tales textos han incluido una proporción creciente de pasajes informativos. Fundamentalmente, las modificaciones son las siguientes: 1) aumento en la proporción de palabras que tienden a aparecer con menos frecuencia en el texto impreso; 2) aumento en la proporción de palabras que es más característico del texto académico como opuesto a la ficción o la conversación; 3) niveles más bajos de cohesión referencial; 4) niveles más bajos de narratividad, y 5) menos instancias de un estilo interactivo/conversacional.

Ochoa (2021), por su parte, mediante un análisis factorial examinó textos descriptivos publicados en manuales escolares de distintos grados y los clasificó en los niveles: muy bajo, bajo, medio y alto. Esta investigadora encontró que, en ocasiones, el nivel de complejidad de los textos no aumenta a lo largo de la escolaridad y que un número importante de textos se ubica en los niveles muy bajo y bajo.

Como se ve, no hay investigaciones relacionadas con la complejidad de textos narrativos en sí mismos.

Ahora bien, la complejidad es definida por Rescher (1998) como una cuestión de cantidad y variedad de elementos, y de sus interrelaciones. Por su parte, Pallotti (2015) considera que el concepto de complejidad no tiene límites claros, pues hace referencia al menos a tres aspectos: 1) complejidad estructural (propiedad de los textos y de los sistemas lingüísticos); 2) complejidad cognitiva, relativa al costo de procesamiento de las estructuras lingüísticas; y 3) desarrollo de la complejidad de distintas estructuras. En un sentido similar, Miestamo (2017) distingue entre complejidad absoluta y complejidad relativa. La primera tiene un carácter inherente al sistema lingüístico o al texto y la segunda está relacionada con la dificultad de una determinada estructura para diversos usuarios.

A efectos de esta investigación se estudió la complejidad absoluta o estructural propia o inherente a los textos, definida por Pallotti (2015) como la característica

de los textos que tiene que ver con el número de constituyentes lingüísticos y sus esquemas relacionales. Este concepto se distancia del de lecturabilidad, la cual es entendida como el grado de facilidad o claridad que presenta un texto para ser comprendido por los lectores a los que está destinado. Tal facilidad se relaciona no solo con las características del texto, sino también con las competencias del lector y las características de la tarea asignada.

Respecto a las variables que determinan la complejidad, se consideran importantes las de “cantidad de elementos” (Kusters, 2008; Bulté & Housen, 2012; Lahuerta, 2017) y “variedad” (Rescher, 1998; Dahl, 2008; Miestamo, 2017). De acuerdo con estas variables, los criterios para determinar la complejidad se han centrado principalmente en indicadores sintácticos y léxicos (Vyatkina, 2013; Corriveau et al., 2016; Schilk & Schaub, 2016) relativos a la oración y la palabra. Hay algunos autores que han trabajado criterios morfológicos (Pallotti, 2015), semánticos (Plakans & Bilki, 2016; Maton & Doran, 2017) y pragmáticos, con unidades mayores que la oración (Yang et al., 2015).

Son *criterios morfológicos* la estructura interna (lexema y morfemas) y la categoría de las palabras (De Clercq & Housen, 2016). Entre los *criterios sintácticos* aparecen con alta frecuencia aquellos relacionados con la estructura del sintagma nominal y de la oración (Mancilla et al., 2017), así como los tipos y clases de oraciones (Peñaloza et al., 2017); entre ellas, la “subordinación”. En el nivel léxico se analiza la “cantidad de palabras”, la “diversidad léxica” (TTR) —entendida como la relación entre palabras diferentes versus palabras no repetidas—, y el grado de dificultad de las palabras. En términos semánticos y pragmáticos, se tienen en cuenta criterios como “cantidad de información”, “temas” y “subtemas” (progresión temática), “cohesión”, “marcadores discursivos” o “conectores”. En el nivel pragmático se estudia la “presuposición” y los “enunciadores” (Solnyshkina et al., 2017).

Algunos de los rasgos anteriores han sido cuestionados como rasgos de complejidad; tal es el caso del subjuntivo —relativizado por Norris y Ortega (2009), quienes señalan que su uso disminuye con la edad y que no es una constante en textos académicos—, y de los marcadores discursivos —Van Silfhout *et al.* (2015) afirman que la presencia de los marcadores conduce a un procesamiento más rápido—.

Teniendo en cuenta las consideraciones anteriores, en la presente investigación se planteó como objeto de estudio la complejidad lingüística en textos narrativos, y se centró la atención en la *complejidad estructural inherente* a los mismos textos (Pallotti, 2015); específicamente, como problema de investigación se indagó qué criterios sintácticos, semánticos y pragmáticos permiten predecir la complejidad de un texto narrativo.

## Metodología

La presente investigación es un estudio descriptivo-interpretativo con un diseño cuantitativo no experimental.

## El corpus

Para el corpus se seleccionaron textos narrativos escolares de dos editoriales colombianas de amplio uso en el medio educativo. La mayoría de los textos son del área de ciencias sociales, los otros textos pertenecen a las ciencias naturales, lenguaje y alguno de matemáticas. Se analizaron en total 65 textos escolares discriminados en cuatro grados: tercero, quinto, séptimo, noveno y once.

### Análisis cuantitativo

En términos estadísticos, el propósito de la investigación consistió en hallar los factores que determinan el nivel de *complejidad lingüística* en una muestra de 65 textos. Teniendo en cuenta el estado de la cuestión, la definición adoptada de complejidad y un análisis previo del corpus, se usaron las siguientes variables para analizar el corpus: “palabras totales (número total de palabras en el texto)”, “palabras repetidas”, “palabras diferentes” (no repetidas), “léxico especializado” (propio de la disciplina a la que pertenece el texto, es decir, de un ámbito específico), “presuposición” (información que no aparece explícita, sino que el lector debe construir), “complementos nominales” (todas aquellas estructuras que modifican al sustantivo, tales como adjetivos, apóstrofes, oraciones relativas, sintagmas preposicionales, deverbales), “sustantivos sin modificadores”, “complementos circunstanciales” (tales como complementos de tiempo, espacio, modo, instrumento, compañía, etc.), “oraciones subordinadas” (dependientes de un verbo principal), “total de oraciones” (número de oraciones que componen el texto), “variación de modo: indicativo, subjuntivo”, “variación de tiempo: pasado, presente, futuro”, “variación de voz: voz activa, voz pasiva”, “marcas de cohesión” (conectores y marcadores discursivos), “anticipaciones” (hipérbaton) y “progresión temática”<sup>1</sup> (relación entre temas y remas). En estas variables se tuvieron en cuenta los factores “cantidad de elementos” y “variedad”. En relación con este último factor se contó la cantidad de modos, tiempos, el tipo de progresión temática, para verificar cuáles variables resultaban productivas y si había o no variedad al usar varias de ellas. También se contó la ocurrencia de cada variable para aplicar el análisis estadístico. La mayoría de las variables se interpretaron de manera ascendente, es decir, si había una mayor entre ellas, más complejo era el texto. Solo las variables “palabras repetidas” y “sustantivos sin modificadores” se contaron en sentido inverso, esto es, cuanto más ocurrencias, menos complejidad. Con estas variables se invitó a un grupo de lingüistas expertos en evaluación de textos a establecer unos niveles que permitieran la clasificación de los textos. Estos establecieron cuatro niveles: 1. Bajo, 2. Medio, 3. Alto y 4. Superior, y generalizaron los rasgos de la siguiente manera:

*Bajo*: textos cortos, esquemas oracionales y nominales simples. Poca variación modal y temporal. Se manejan conceptos muy concretos y familiares. El léxico no requiere de aclaraciones ni de diccionarios. Nula presuposición.

1 Se sigue aquí la definición y clasificación de Combettes (1983).

*Medio*: sintagmas nominales con alguna complementación, oraciones simples y oraciones subordinadas. También pueden aparecer incisos entre los componentes de las oraciones. La longitud de los textos puede ser mayor. Se maneja ya cierto nivel de abstracción, aparece alguna que otra generalización. Baja presuposición.

*Alto*: predominio de oraciones subordinadas y sintagmas nominales complejos, textos con mayor número de palabras diferentes, el volumen de información es considerable. Aparecen conceptos abstractos y generales. Mediana presuposición.

*Superior*: estructuras complejas sintáctica y semánticamente (nominalizaciones, deverbales), términos abstractos y especializados, variedad temática, alta presuposición.

Una vez hecha esta jerarquización, los expertos clasificaron los textos de manera individual. Debido a que frente a algunos textos había discrepancia, se invitó a los expertos a un análisis en grupo sobre los textos donde la clasificación no era unívoca. Hecha la discusión, se llegó a una nueva clasificación por consenso (anexo 1).

Posteriormente se aplicó el modelo de odds proporcionales, que asume pendientes  $\beta$ 's diferentes para todas las variables explicativas, lo que requiere una gran cantidad de datos para estimar un coeficiente por cada predictor en cada logit acumulativo. Una forma de simplificar el modelo consiste en suponer que los coeficientes para cada predictor son idénticos para todas las categorías de respuesta de la variable dependiente; este supuesto de pendientes iguales se conoce como *supuesto de odds proporcionales* o también *supuesto de líneas paralelas*. En este caso, el modelo llega a ser

$$\text{logit}[P(Y \leq j|X)] = \alpha_j + \beta X = \alpha_j + \beta_1 X_1 + \dots + \beta_k X_k, j = 1, \dots, J - 1$$

donde la probabilidad de que la variable dependiente se encuentre por debajo o en la categoría  $j$  es

$$-P(Y \leq j|X) = \frac{e^{\alpha_j + \beta X}}{1 + e^{\alpha_j + \beta X}} = \frac{e^{\alpha_j + \beta_1 X_1 + \dots + \beta_k X_k}}{1 + e^{\alpha_j + \beta_1 X_1 + \dots + \beta_k X_k}}, j = 1, \dots, J - 1$$

y el odds de la variable dependiente por debajo o en la categoría  $j$

$$\frac{P(Y \leq j|X)}{P(Y > j|X)} = e^{\alpha_j + \beta X} = e^{\alpha_j + \beta_1 X_1 + \dots + \beta_k X_k}$$

## Resultados

Un total de 20 variables predictoras se midieron para los 65 textos que componen la muestra, lo que implica un total de 190 coeficientes de correlación (Tabla 1). Ahora bien, a excepción del número de verbos en tiempo futuro, que presenta muy bajos coeficientes de correlación, con el resto de las variables se observa, en general, una alta asociación lineal entre las variables: el 77.9 % de los coeficientes se



encuentra por encima de 0.4 y aproximadamente el 55 % es como mínimo 0.6. Tanto el gran número de variables como los altos coeficientes de correlación presuponen un inconveniente al momento de modelar el nivel de complejidad, como consecuencia del tamaño de la muestra y de la posibilidad de que el modelo estimado presente problemas de multicolinealidad.

**Tabla 1.** Matriz de correlación para las variables predictoras

Variable	Etiqueta	TOT PBR'S	PBR'S REP	PBR'S _DIF	LEX ESP	PRE SUP	COM NOM	SUST _SM	COM CIR	ORAC _SUB	TOT ORAC	MOD IND	MOD SUB	TMP PAS	TMP PRES	TMP FUT	VOZ ACT	VOZ PAS	MAR COH	ANTIC	PROG TEM
TOT PBR'S	Total palabras	1.00	0.96	0.99	0.77	0.68	0.91	0.92	0.88	0.89	0.93	0.94	0.49	0.84	0.64	0.11	0.87	0.60	0.63	0.75	0.58
PBR'S REP	Palabras repetidas		1.00	0.94	0.69	0.61	0.85	0.90	0.86	0.72	0.94	0.94	0.60	0.86	0.65	0.12	0.87	0.58	0.64	0.79	0.58
PBR'S DIF	Palabras diferentes			1.00	0.78	0.68	0.91	0.92	0.86	0.71	0.92	0.94	0.49	0.86	0.61	0.10	0.86	0.58	0.63	0.72	0.58
LEX ESP	Léxico especializado				1.00	0.95	0.75	0.76	0.58	0.38	0.68	0.64	0.17	0.64	0.27	-0.10	0.65	0.37	0.61	0.49	0.52
PRE SUP	Phras uposición					1.00	0.65	0.67	0.50	0.34	0.60	0.58	0.14	0.59	0.20	0.10	0.64	0.36	0.58	0.44	0.41
COM NOM	Complementos nominales						1.00	0.79	0.84	0.57	0.75	0.80	0.32	0.73	0.48	0.05	0.74	0.53	0.48	0.71	0.52
SUST SM	Sustantivos s in modificaciones							1.00	0.75	0.70	0.92	0.91	0.39	0.80	0.62	0.08	0.80	0.53	0.66	0.71	0.59
COM CIR	Complementos circunstanciales								1.00	0.58	0.83	0.84	0.42	0.75	0.54	0.13	0.77	0.40	0.54	0.73	0.47
ORAC SUB	Oraciones subordinadas									1.00	0.75	0.78	0.64	0.69	0.60	0.23	0.67	0.46	0.44	0.62	0.39
TOT ORAC	Total oraciones										1.00	0.96	0.55	0.88	0.64	0.17	0.85	0.62	0.67	0.75	0.59
MOD IND	Verbos en modo indicativo											1.00	0.59	0.90	0.88	0.19	0.88	0.60	0.61	0.78	0.55
MOD SUB	Verbos en modo subjuntivo												1.00	0.59	0.47	0.20	0.56	0.32	0.31	0.43	0.24
TMP PAS	Verbos en pasado													1.00	0.32	0.01	0.85	0.66	0.66	0.60	0.60
TMP PRES	Verbos en presente														1.00	0.39	0.51	0.39	0.39	0.61	0.22
TMP FUT	Verbos en futuro															1.00	0.11	0.05	0.15	0.21	-0.03
VOZ ACT	Verbos en voz activa																1.00	0.63	0.60	0.64	0.42
VOZ PAS	Verbos en voz pasiva																	1.00	0.31	0.38	0.29
MAR COH	Marcas de cohesión																		1.00	0.36	0.23
ANTIC	Anticipación																			1.00	0.63
PROG TEM	Progresión temática																				1.00

Este método consistió en reducir el número de variables mediante el análisis de componentes principales (ACP) y en usar las componentes extraídas como variables explicativas en la estimación del modelo de regresión logit acumulativo de odds proporcionales. Este camino tiene dos ventajas: reduce el número de variables y, dado que las componentes son no correlacionadas por construcción, el modelo ajustado no presenta problemas de colinealidad.

### Extracción de componentes principales

Se retuvieron tres componentes de acuerdo con el criterio del número valores propios mayores que 1, las cuales explican un 78.7 % de la variabilidad total en los datos. Con el fin de encontrar una estructura simple se aplicó un método de rotación oblicuo (el modelo factorial de rotación se presenta en la Tabla 2). Como se observa, del total de variables, 16 presentan cargas significativas sobre la primera componente, 3 sobre la segunda y solamente una (verbos en tiempo futuro) sobre la tercera componente; lo que era de esperarse, dadas las muy bajas correlaciones entre esta variable y las restantes. Es necesario anotar que todas las cargas son positivas.

Tabla 2. Modelo factorial de rotación

Variable	Componente 1	Componente 2	Componente 3
MOD_IND	0.90012		
PBRS_REP	0.88382		
TMP_PAS	0.87131		
ANTIC	0.87036		
TOT_ORAC	0.84119		
ORAC_SUB	0.83768		
TOT_PBRS	0.81033		
PBRS_DIF	0.80693		
COM_CIR	0.77818		
MOD_SUB	0.77291		
VOZ_ACT	0.75520		
PROG_TEM	0.74639		
SUST_SM	0.73504		
COM_NOM	0.72851		
VOZ_PAS	0.61222		
TMP_PRES	0.60339		
PRESUP		0.79065	
LEX_ESP		0.74711	
MAR_COH		0.69501	
TMP_FUT			0.80378

Una lectura de la componente 1 muestra que, en general, las cargas son muy altas; salvo verbos en voz pasiva y verbos en tiempo presente con cargas de 0.61 y 0.60, respectivamente, las restantes variables tienen cargas entre 0.73 y 0.90; por tanto, esta componente puede interpretarse como un **factor de tamaño**, que incluye atributos que aumentan el nivel de complejidad de un texto, así como algunos que lo reducen. Por otra parte, dadas las variables que cargan en la componente 2 (presuposición, léxico especializado y marcas de cohesión), esta componente puede interpretarse como un **factor de calidad de la comunicación**; el emisor espera que lo no explícito sea entendido, que las relaciones expuestas tengan validez lógica y que el léxico empleado sea accesible; del lector, por su parte, se espera capacidad de deducción, comprensión y análisis.

### Estimación del modelo

Se puede concluir que el modelo con intercepto y covariables se ajusta mejor que el modelo base (solo intercepto), dado que el valor de la estadística para cualquiera de los criterios tenidos en cuenta es mucho menor; lo que reafirma el rechazo de la hipótesis **global** de no regresión ( $H_0: \beta_1 = \beta_2 = \beta_3 = \beta_4 = 0$ ) a los niveles usuales de significancia y demuestra que el modelo adaptado presenta un buen ajuste. Las tres componentes resultan significativas, con coeficiente estimado positivo para las dos primeras y negativo para verbos en tiempo futuro. En lo que se refiere al desempeño predictivo, la estadística de D de Somer es, aproximadamente, 0.86 significativamente menor que para el modelo seleccionado con el método 1 (0.93).

**Tabla 3.** Estimación de modelo logit acumulativo de odds proporcionales usando componentes principales como predictores

Prueba Score para supuesto de odds proporcionales						
Chi-Square	DF	Pr > ChiSq				
1.2240	3	0.7473				
Estadísticas de Ajuste del Modelo			Prueba Global para Hipótesis Nula: BETA=0			
Criterion	Intercept Only	Intercept and Covariates	Test	Chi-Square	DF	Pr > ChiSq
AIC	129.278	74.977	Likelihood Ratio	60.3007	3	<.0001
SC	133.626	85.849	Score	30.0491	3	<.0001
-2 Log L	125.278	64.977	Wald	21.2850	3	<.0001
Análisis de Estimaciones Máximo Verosímil						
Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq	Exp(Est)
Intercept Alto-Superior	1	1.6306	0.6330	6.6365	0.0100	5.107
Intercept Medio	1	5.0343	1.0019	25.2492	<.0001	153.586
Componente 1	1	2.0347	0.8622	5.5689	0.0183	7.650
Componente 2	1	3.5668	0.9680	13.5764	0.0002	35.402
Componente 3	1	-1.1624	0.5904	3.8767	0.0490	0.313
Asociación de Respuestas Observadas y Valores Predichos						
Percent Concordant	92.7	Somers' D	0.855			
Percent Discordant	7.3	Gamma	0.855			
Percent Tied	0	Tau-a	0.503			
Pairs	1224	c	0.927			

### Modelo seleccionado

$$\frac{P(Y \leq 3|X)}{P(Y > 3|X)} = e^{1.63+2.03*CP1+3.57*CP2-1.16*CP3}$$

y

$$\frac{P(Y \leq 2|X)}{P(Y > 2|X)} = e^{5.03+2.03*CP1+3.57*CP2-1.16*CP3}$$

Con CP1, CP2 y CP3 cada una de las componentes principales.

### Interpretación

**Tabla 4.** Estimaciones de odds ratio para el modelo que usa componentes principales como predictores

Efecto	Estimación
Componente 1	7.650
Componente 2	35.402
Componente 3	0.313

De la tabla puede concluirse que:

- El odds (posibilidad ponderada) de un texto de tener mayor nivel de complejidad es 7.65 veces mayor al aumentar en una unidad la componente principal 1.

- La posibilidad ponderada de un texto de tener mayor nivel de complejidad es 35.4 veces mayor si se aumenta la componente principal 2 en una unidad.
- La posibilidad ponderada de un texto de tener mayor nivel de complejidad se reduce en un 68.7 % al aumentar en una unidad el número de verbos en tiempo futuro.

## Discusión y conclusiones

El hecho de que la mayoría de los criterios se ubiquen en un solo factor presenta inconvenientes a la hora de explicarlos teóricamente. No obstante, es claro que en el primer componente se agrupan criterios morfosintácticos y semánticos relacionados con la variable “cantidad”, lo cual concuerda con trabajos como los de Kusters (2008), Bulté y Housen (2012), y Lahuerta (2017). En el segundo componente los criterios son claramente semántico-pragmáticos, en cuanto el léxico especializado y la presuposición exigen un conocimiento que se ha de construir con la lectura o un conocimiento previo por parte del lector. Igualmente, las marcas de cohesión guían la interpretación del lector, ya que muestran el tipo de relación semántica que se establece entre las proposiciones del texto.

Es interesante que el factor que mayor incrementa la complejidad es el que está en el componente 2 que, como se ha señalado, tiene una estrecha conexión con los niveles semántico y pragmático. Este hallazgo muestra que los criterios sintácticos, tal como los empleados por Vyatkina (2013), Corriveau *et al.* (2016), y Schilk y Schaub (2016), no son suficientes para dar cuenta de la complejidad estructural. Por el contrario, la conclusión es coherente con los pocos estudios reportados en la bibliografía en los que la semántica ocupa un lugar principal (Plakans & Bilki, 2016) al igual que la pragmática (Solnyshkina *et al.*, 2017).

Los resultados coinciden también con la primera revisión del corpus y con la intuición de los expertos: mayor presuposición, más y variadas marcas de cohesión y mayor presencia de léxico especializado tornan más complejo un texto.

Otras variables asociadas al factor “cantidad” como “palabras totales”, “total de oraciones”, “marcas de cohesión”, “complementos circunstanciales” resultaron significativas, pero en menor medida. Esto nos lleva a inferir que la longitud de un texto no implica necesariamente mayor complejidad, por cuanto un texto puede tener gran longitud, pero no ser complejo y viceversa: textos cortos pueden tener un alto grado de complejidad. De igual manera, criterios morfológicos y sintácticos relacionados con tiempos y modos verbales, voz, complementación nominal, tipos de oraciones, inciden, pero en una proporción menor. Llama la atención que el léxico diferente no aparezca en el componente dos, sino en el uno. Puede hipotetizarse que la cantidad y no la calidad (variedad) fue decisiva para su ubicación. Así mismo puede suceder que las palabras sean diferentes pero que sean de uso frecuente, que el

término no entorpezca la comprensión global o que su significado pueda ser inferido del contexto intratextual.

El tiempo futuro aparece con el signo negativo, posiblemente para señalar el alejamiento en la escala de la prototipicidad del tiempo privilegiado en las narraciones: el pasado. No obstante, es curioso que aparezca él solo en un componente independiente.

Los resultados generales de esta investigación muestran que, contrario a lo señalado por Frantz *et al.* (2015), Ortega (2015), y Aravena y Hugo (2016), los criterios que miden la complejidad no dependen de un tipo o género textual específico, ya que es evidente que las tres variables principales que explican la mayor complejidad, a saber, presuposición, conectores y léxico especializado, se presentan en cualquier tipo discursivo. Sin embargo, se puede afirmar que prototípicamente los textos narrativos escolares cuentan un suceso de forma cronológica, muestran acciones que se desarrollan principalmente en tiempo pasado, se construyen en voz activa, modo indicativo. Una mayor variación de estos indicadores podría aumentar la complejidad y con ello retar cognitivamente a los estudiantes.

Como se señaló, los rasgos semánticos son decisivos a la hora de evaluar la complejidad textual. Esta información es importante para profesores, coordinadores académicos, editores y programadores del currículum: contar con un conjunto de criterios lingüísticos apoya la tarea que tienen educadores y coordinadores escolares de seleccionar adecuadamente los textos para cada grado escolar, desarrollar la competencia comunicativa en forma progresiva y continua, y evaluar la comprensión lectora.

También puede ser útil para personas y organismos encargados de la evaluación de la comprensión lectora, pues ofrece unos indicadores acerca de qué evaluar y cuándo. La atención debe centrarse entonces en qué tanta información ofrece un texto, qué tan variada es, de qué calidad, qué tan novedosa es para los estudiantes, qué información presupone o está implícita y qué tipo de conexiones o inferencias implica.

Para terminar, es importante señalar que esta es una investigación exploratoria tanto en relación con los factores y variables que dan cuenta de la complejidad como con el tipo de texto estudiado. En este sentido, los resultados son preliminares. Es preciso indagar con un corpus más amplio, donde se incluyan textos narrativos no escolares y con textos donde primen otras secuencias textuales para validar los resultados obtenidos. Así mismo es importante contrastar el análisis con procesos de comprensión de textos en poblaciones estudiantiles de diversos grados y ciclos escolares (complejidad relativa). También es conveniente llevar a cabo análisis cualitativos que posibiliten un estudio más profundo de las variables y los componentes que permiten describir la complejidad textual.

## Referencias

- Adam, J. M. (1992). *Les textes. Types et prototypes*. Nathan.
- Aravena, S. & Hugo, E. (2016). Desarrollo de la complejidad sintáctica en textos narrativos y explicativos escritos por estudiantes secundarios. *Lenguas Modernas*, 47, 9-40. <https://revistas.uchile.cl/index.php/LM/article/view/45181/47238>
- Bulté, B., & Housen, A. (2012). Defining and Operationalising L2 Complexity. In A. Housen, F. Kuiken y Vedder, I. (eds.), *Dimensions of L2 Performance and Proficiency: Investigating Complexity, Accuracy and Fluency in SLA* (pp. 21-46). John Benjamins Publishing Company.
- Combettes, B. (1983). *Pour une grammaire textuelle. La progression thématique*. DeBoeck-Duculot.
- Corriveau, K. H., Kurkul, K., & Arunachalam, S. (2016). Preschoolers' Preference for Syntactic Complexity Varies by Socioeconomic Status. *Child Development*, 87(5), 1529-1537. <https://dx.doi.org/10.1111%2Fcddev.12553>
- Dahl, Ö. (2008). Grammatical Resources and Linguistic Complexity: Sirionó as a Language without NP Coordination. In M. Miestamo, K. Sinnemäki & F. Karlsson (eds.), *Language Complexity: Typology, contact, change* (pp. 153-164). John Benjamins Publishing Company. <https://doi.org/10.1075/slcs.94.10dah>
- De Clercq, B., & Housen, A. (2016). The Development of Morphological Complexity: A Cross-Linguistic Study of L2 French and English. *Second Language Research*, 35(1), 71-97. <https://doi.org/10.1177%2F0267658316674506>
- Frantz, R., Starr, L., & Bailey, A. (2015). Syntactic Complexity as an Aspect of Text Complexity. *Educational Researcher*, 44(7), 387-393. <https://doi.org/10.3102%2F0013189X15603980>
- Kusters, W. (2008). Complexity in Linguistic Theory, Language Learning and Language Change. In M. Miestamo, K. Sinnemäki, & F. Karlsson. (Eds.), *Studies in Language Companion Series*, n.º 94 (pp. 3-22). John Benjamins Publishing Company.
- Lahuerta, A. (2017). Syntactic Complexity in Secondary-Level English Writing: Differences Among Writers Enrolled on Bilingual and Non-Bilingual Programmes. *Porta Linguarum*, 28, 67-80. [http://www.ugr.es/~portalin/articulos/PL\\_numero28/5\\_Lahuerta.pdf](http://www.ugr.es/~portalin/articulos/PL_numero28/5_Lahuerta.pdf)
- Mancilla, R., Polat, N., & 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. <https://doi.org/10.1093/applin/amv012>
- Maton, K., & Doran, J. (2017). Condensation: A Translation Device for Revealing Complexity of Knowledge Practices in Discourse, Part 2—Clausing and Sequencing. *Onomázein*, (número especial LSF y TCL sobre educación y conocimiento), 77-110. <https://doi.org/10.7764/onomazein.sfl.04>
- Meneses, A., Ow, M., & Benítez, R. (2012). Complejidad sintáctica: ¿modalidad comunicativa o tipo textual? Estudio de casos de producciones textuales de estudiantes de 5.º básico. *Onomázein*, 25, 65-93. <https://repositorio.uc.cl/handle/11534/8063>
- Miestamo, M. (2017). Linguistic Diversity and Complexity. *Lingue e Linguaggio*, 16(2), 227-253. [https://www.mv.helsinki.fi/home/matmies/publications/03\\_Miestamo\\_2017\\_2.pdf](https://www.mv.helsinki.fi/home/matmies/publications/03_Miestamo_2017_2.pdf)
- Nini, A. (2015). The Multidimensional Analysis Tagger [MAT, v. 1.3]. A Replication of Douglas Biber's *Variation Across Speech and Writing* Tagger. <http://sites.google.com/site/multidimensionaltagger> |
- Nippold, M., Frantz-Kaspar, M., & Vigeland, L. (2017). Spoken Language Production in Young Adults: Examining Syntactic Complexity. *Journal of Speech, Language, and Hearing Research*, 60(5), 1339-1347. [https://doi.org/10.1044/2016\\_JSLHR-L-16-0124](https://doi.org/10.1044/2016_JSLHR-L-16-0124)
- Norris, J., & Ortega, L. (2009). Towards an Organic Approach to Investigating CAF in Instructed SLA: The Case of Complexity. *Applied Linguistics*, 30(4), 555-578. <https://doi.org/10.1093/applin/amp044>
- Ochoa, L. (2016). *Estudio sobre los niveles de dificultad de los textos que conforman las pruebas Saber 3°, 5° y 9° en el área de Lenguaje*. ICFES. (Inédito).
- Ochoa, L. (2021). *Complejidad estructural en textos escolares descriptivo-expositivos: estudio de corpus*. *Folios*, 54, 91-110. <https://doi.org/10.17227/folios.54-11977>
- Ortega, L. (2015). Syntactic Complexity in L2 Writing: Progress and Expansion. *Journal of Second Language Writing*, 29(1), 82-94. <http://dx.doi.org/10.1016/j.jslw.2015.06.008>
- Pallotti, G. (2015). A Simple View of Linguistic Complexity. *Second Language Research*, 31(1), 117-134. <https://doi.org/10.1177/0267658314536435>
- Peñalosa, C., Araya, C. & Coloma, C. (2017). Desarrollo de la complejidad sintáctica en recuentos narrativos de niños preescolares y escolares. *Logos*, 27(2), 334-349. <https://revistas.userena.cl/index.php/logos/article/view/949/1106>

- Plakans, L., & Bilki, Z. (2016). Cohesion Features in ESL Reading: Comparing Beginning, Intermediate and Advanced Textbooks. *Reading in a Foreign Language*, 28(1), 79-100. <http://dx.doi.org/10.15443/RL2726>
- Polio, C., & Yoon, H. (2018). The Reliability and Validity of Automated Tools for Examining Variation in Syntactic Complexity Across Genres. *International Journal of Applied Linguistics*, 28(1), 165-188. <https://doi.org/10.1111/ijal.12200>
- Rescher, N. (1998). *Complexity: A Philosophical Overview*. Transaction Publishers.
- Schilk, M., & Schaub, S. (2016). Noun Phrase Complexity Across Varieties of English: Focus on Syntactic Function and Text Type. *English World-Wide*, 37(1), 58-85. <http://doi.org/10.1075/eww.37.1.03sch>
- Sheehan, K., Flor, M., Napolitano, D. & Ramineni, Ch. (2015). Using *TextEvaluator* to Quantify Sources of Linguistic Complexity in Textbooks Targeted at First-Grade Readers Over the Past Half Century. *ETS Research Report Series*, 2, 1-17.
- Solnyshkina, M., Zamaletdinov, R., Gorodetskaya, L., & Gabitov, A. (2017). Evaluating Text Complexity and Flesch-Kincaid Grade Level. *Journal of Social Studies Education Research*, 8(3), 238-248. <https://files.eric.ed.gov/fulltext/EJ1162266.pdf>
- Text Ease and Readability Assessor [TERA]. <http://129.219.222.70:8084/Coh-Matrix.aspx>
- TextEvaluator (2017). *About the TextEvaluator® Technology*. <https://textevaluator.ets.org/TextEvaluator/Docs/AboutTextEvaluator.pdf>
- VanSilfhout, G., Evers-Vermeul, J., & Sanders, T. (2015). Connectives as Processing Signals: How Students Benefit in Processing Narrative and Expository Texts. *Discourse Processes*, 52(1), 47-76. <https://doi.org/10.1080/0163853X.2014.905237>
- Vyatkina, N. (2013). Specific Syntactic Complexity: Developmental Profiling of Individuals Based on an Annotated Learner Corpus. *The Modern Language Journal*, 97(S1), 11-30. <https://doi.org/10.1111/j.1540-4781.2012.01421.x>
- Werlich, E. (1975). *Typologie der Texte*. Quelle & Meyer.
- Yang, W., Lu, X., & 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. <https://doi.org/10.1016/j.jslw.2015.02.002>



## Anexo 1

### Clasificación de 65 textos realizada por panel de lingüistas expertos

#### Clasificación de los 65 textos según su grado de complejidad

Bajo	Medio	Alto	Superior
1, 3, 6, 8, 9, 10, 11, 12 (tercero), 23 (quinto),	2, 4, 5, 7, 13 (tercero), 14, 16,17, 18, 19, 20, 22, 24, 25, 26 (quinto), 30, 35, 36, 37, 38 (séptimo),	15, 21 (quinto), 27,28,29, 31, 32, 33, 34, 39 (séptimo), 41, 44, 47, 48, 50, 51 (novenos), 54, 59, 60, 61, 63 (once)	40, 42, 43, 45, 46, 49, 52 (novenos), 53, 55, 56, 57, 58, 62, 64, 65 (once)