LECTURAS ADAPTADAS A LA ENSEÑANZA DEL ESPAÑOL COMO L2: VARIABLES LINGÜÍSTICAS PARA LA DETERMINACIÓN DEL NIVEL DE LEGIBILIDAD[1]

ALBERTO ANULA
Universidad Autónoma de Madrid

Como otras muchas herramientas metodológicas para el aprendizaje de español como lengua extranjera o segunda lengua[2], las lecturas graduadas o adaptadas deben ser escalonadas conforme a los presupuestos del *Marco Común Europeo de Referencia para las lenguas* (MCER) y del *Plan Curricular del Instituto Cervantes* (PCIC). El presente trabajo aborda la cuestión de cómo determinar el nivel de legibilidad de estos materiales para adecuarlo a los distintos niveles contemplados por el *MCER* y el *PCIC*. Sin embargo, el tema elegido es demasiado extenso como para abarcarlo aquí en su conjunto, por lo que nos limitaremos a tratarlo parcialmente. En concreto, centraremos nuestra atención en los textos literarios de carácter narrativo-descriptivo y propondremos que la determinación del nivel de legibilidad de las lecturas para el aprendizaje de L2 debe realizarse analizando la *complejidad lingüística*[3] de los textos a través de un conjunto reducido de variables centradas en las propiedades léxicas y sintácticas de la microestructura textual.

1. Principales factores lingüísticos que dificultan la comprensión lectora en aprendices de L2

El protagonista de *El curioso incidente del perro a medianoche* (Haddon 2004), Christopher, afirma que no le gustan las novelas porque en ellas se dicen cosas como:

(1) Estoy veteado de hierro, de plata y del barro más burdo. No puedo contraerme en ese puño firme que aprietan aquellos que no dependen de estímulos. (Haddon, 2004: 14)

Y se pregunta a continuación ¿qué significa eso?, pues ni él, ni su padre, ni sus profesores saben responderle. ¿Podría responderle un estudiante de español como L2 de nivel intermedio o avanzado? ¿Podría responder este hipotético estudiante a la misma pregunta en relación con el siguiente texto?:

(2) Se vuelven a tomar fibras de pluma de mallard de un largo suficiente de modo que un ala saliendo desde el punto donde quedó el hilo alcance a llegar hasta al menos la mitad de la cola, pero no sobrepasarla. Una vez más se aplican estas fibras intentando mantener su curvatura hacia arriba de modo de formar un ala arqueada[4].

Podemos suponer que no y la respuesta probablemente no sería diferente si se la formuláramos a un lector de L1. Los textos de (1) y (2) plantean problemas de distinto orden relacionados con los significados no literales de (1) o con el conocimiento del mundo previo que nos permite descifrar la acción descrita en (2)[5]. Sin embargo, aquí no trataremos estas cuestiones, ni otras que también influyen en la comprensión lectora (tales como las diferencias interculturales, la interferencia de la L1 sobre la L2, los aspectos cognitivos relacionados con el lector, etc., etc.). En su lugar prestaremos atención a dos aspectos formales que dificultan la comprensión lectora de los aprendices de L2: el desconocimiento del léxico y la dificultad para el reconocimiento de estructuras sintácticas (véase Acquaroni 2004: 951) o la presencia en los textos de estructuras demasiado complejas (Gracia 2006: 150).

La dificultad del conocimiento del léxico puede quedar ilustrada suficientemente con el ejemplo siguiente:

[1] El presente trabajo cuenta con una ayuda concedida por la Comunidad de Madrid y la Universidad Autónoma de Madrid al proyecto *La complejidad lingüística en los discursos escritos y la comprensión lectora* (CCG6-UAM_HUM-0104).
[2] En adelante uniremos ambos términos bajo la noción de "L2".
[3] Naturalmente, hay otras propiedades de los textos que podrían igualmente evaluarse. Sin embargo, creemos que la *complejidad lingüística* es el factor más relevante a tener en cuenta o, si se prefiere, el primer factor a considerar. Quizá por ello, en la presentación de la sección *Lecturas paso a paso* del Instituto Cervantes, se apela a la complejidad lingüística de los textos para justificar los niveles en que organizan las lecturas: "hemos organizado en tres niveles —inicial, intermedio y avanzado— según la complejidad de los textos que incluyen" (Instituto Cervantes: http://213.4.108.140/aula/lecturas/).
[4] Tomado de http://www.riosysenderos.com/baul/relledmuddler.htm
[5] El texto describe un arte de la pesca como es la técnica del atado del azuelo.

Centro Virtual Cervantes

(3) El sufí trasiega la manigueta / Silban falenas aleatorias (Quintero 1991: 300)

Salvo el verbo *silbar*, ninguna otra palabra esta comprendida en el inventario léxico del *PCIC* y ésta está indicada para el nivel de referencia B2. Parece bastante evidente que el éxito o fracaso de la comprensión lectora dependerá del conocimiento previo que tenga el lector del significado de *sufí, trasegar, manigueta, y falena*; todas ellas voces de baja frecuencia. Respecto del dominio previo de las estructuras sintácticas que conforman el discurso, la situación no puede ser tan nítida como en el caso del léxico, donde podemos diferenciar entre palabras conocidas y desconocidas para el lector. No obstante, un texto como el ilustrado en (4) resulta claramente inadecuado para aprendices de español como L2 de los niveles A y B del *PCIC* y ello, entre otras razones, porque la secuencia de (4) constituye una única unidad textual desde el punto de vista de su trabazón ortográfica[6]:

(4) Por cuantos datos descubrí eclipsados entre los hechos reales y las fábulas verosímiles que Leo Mistral tramó casi siempre con una asombrosa habilidad para contar mentiras que parecían verdades irrefutables, reafirmaré mi convicción de que un broker del negocio petrolero a quien nunca conocí en persona, Patricio Crown, por sobrenombre amistoso el Coronel Lawrence, logró de todos nosotros durante el tiempo que le fue necesario cuanto se propuso para romper el maleficio que había condenado a Ada, la mujer hechizada a la que convenció para que lo acompañara hasta Madrid, a vagar por las playas insulares de Mafasca, sin que ninguno de nosotros advirtiéramos con certera exactitud cuándo habíamos sucumbido bajo la férula de su caprichosa naturaleza (Armas Marcelo 1994: 9).

Ahora bien, el factor gramatical señalado anteriormente (la complejidad estructural y/o el conocimiento previo de las estructuras sintácticas utilizadas en el discurso escrito) constituye un problema teórico de primer orden de cara a proponer herramientas o aplicaciones de carácter pedagógico que guíen el proceso de adquisición de las habilidades de comprensión lectora en los aprendices de español como L2. El problema no puede resolverse a partir de rigurosas exclusiones del inventario de conocimientos. Pongamos un ejemplo de esta estrategia inadecuada. El *PCIC,* al establecer el inventario de estructuras oracionales graduado por niveles, determina que, en el nivel A2, la subordinación sustantiva queda limitada a la modalidad de oraciones de infinitivo en función de sujeto de verbos psicológicos inacusativos (del tipo de *encantar*) y en función de complemento de régimen con verbos de dos argumentos (del tipo de *aprender a*). ¿Quiere esto decir que quedan excluidas entonces las construcciones completivas flexionadas y en función de objeto en los textos propuestos como lecturas para los niveles iniciales? Unos pocos ejemplos tomados de textos del nivel inicial del apartado de *Lecturas paso a paso* del Instituto Cervantes[7] niegan tal posibilidad:

(5) Oí que Margarita, la secretaria, cogía el teléfono. [...] Creo que sí. [...] (Miquel y Sans 1995. [Nivel inicial. Grado de dificultad 5])

(6) Todos contestan que sí con la cabeza. [...] Quiero decir que volar en un avión no es divertido [...] Además, ahora sabe que siente algo muy especial por ese chico. (Acquaroni Muñoz 1994 [Nivel inicial. Grado de dificultad 3])

(7) Sabe que son cincuenta y cinco pesetas. [...] Ahora tengo que irme [...] Pepe piensa que Susi es insoportablemente lista y que, además, es la persona que mejor lo conoce en el mundo. (De Miguel y Santos 1987 [Nivel inicial. Grado de dificultad 1])

Como probablemente no podría ser de otra forma, en los textos de "nivel inicial" seleccionados como lecturas por el Instituto Cervantes, las construcciones completivas flexionadas en función de objeto están suficientemente representadas. Si nos fijáramos en otras estructuras, igualmente básicas, el resultado no sería distinto. Por ello, la aproximación a la determinación del nivel de dificultad de los textos en virtud de sus características sintácticas no debe centrarse en la presencia o ausencia de ciertos tipos de estructuras o elementos.

2. Variables lingüísticas para la determinación del nivel de legibilidad de las lecturas adaptadas para la enseñanza del español como L2

Se han realizado distintos esfuerzos por analizar la complejidad lingüística del discurso. La mayor parte de las investigaciones en este ámbito han partido tradicionalmente de comparar las propiedades del lenguaje oral y del lenguaje escrito para establecer una lista más o menos extensa de propiedades caracterizadoras de lo que es simple o complejo en el discurso. En la tabla 1 se listan algunas de estas propiedades:

[6] Véase más adelante la noción de "segmento discursivo".
[7] Acceso disponible en línea en http://cvc.cervantes.es/aula/lecturas/.

Centro Virtual Cervantes

Caracterizadoras de la simplicidad	Propiedades léxicas y sintácticas	Caracterizadoras de la complejidad
Menor número de palabras diferentes	Redundancia léxica	Mayor número de palabras diferentes
Predominio de oraciones simples	Naturaleza oracional	Más oraciones coordinadas y subordinadas
Predominio de la yuxtaposición y la coordinación	Enlaces sintácticos entre oraciones	Predominio de la subordinación
Limitada a la estructura argumental	Desarrollo clausal	Abundan los sintagmas no argumentales
El verbo como pieza fundamental	Estilo verbal o nominal	Mayor desarrollo de las estructuras nominales
Unidades discursivas breves	Longitud de las unidades discursivas	Unidades discursivas extensas
Menor número de estructuras incrustadas	Incrustación, recursividad	Mayor número de estructuras incrustadas

Tabla 1. Propiedades léxicas y sintácticas para la caracterización de la complejidad del discurso.

Con independencia de estas propiedades (y de otras que puedan añadirse a la tabla 1), la cuestión más relevante para nuestro propósito es la de cómo evaluar la complejidad de los textos adaptados y graduados para establecer (o ratificar en su caso) el nivel de lectura de un texto conforme a las especificaciones del *MCER* y del *PCIC*.

En un trabajo reciente, Checa García (2006) ha aplicado las medidas de "madurez sintáctica" de Hunt (1964, 1967, 1970) al análisis de un corpus de 30 fragmentos textuales de 200 palabras aproximadamente correspondientes a discursos narrativos tomados de lecturas graduadas en seis niveles de dificultad. El objetivo de la autora consistía en determinar si los niveles de dificultad asignados a los textos correlacionaban con una progresiva complejidad sintáctica medida mediante los índices primarios de Hunt[8]. Los resultados de su estudio no resultaron concluyentes pues, si bien el "índice de longitud de la unidad terminal" mostró que el incremento de la dificultad a través de los distintos niveles propuestos fue significativo, el resto de los índices utilizados no ofrecieron los resultados esperados. Más allá de las razones que aduce su autora para explicar los resultados[9], lo que nos interesa señalar en esta ocasión es que un modelo como el de Hunt solo contempla una parte de las propiedades que caracterizan la simplicidad o complejidad de los textos escritos y, por tal motivo, resulta inapropiado. En nuestra opinión, se requiere un modelo de análisis que tenga en cuenta una parte significativa de las propiedades listadas en la tabla 1, al que habrá que añadir alguna propiedad especialmente relevante en el ámbito de la enseñanza de L2, como es la frecuencia de uso de las palabras y de las estructuras léxicas.

En Anula (2007a y 2007b) hemos propuesto que la determinación del nivel de legibilidad de un texto adaptado para la enseñanza del español como L2 debe realizarse mediante la fijación del nivel de complejidad lingüística que tal texto posee en su conjunto, teniendo en cuenta fundamentalmente tres dimensiones de análisis de la microestructura del texto y seis variables o índices de medida:

(8) Dimensión léxica; medida por: (a) frecuencia y (b) densidad o variación léxica
(9) Dimensión textual/segmental; medida por: (a) longitud y (b) naturaleza de los segmentos discursivos
(10) Dimensión oracional/clausal; medida por: (a) subordinación e (c) incrustación/recursividad oracional

A continuación, enunciaremos e ilustraremos brevemente las dimensiones y variables enumeradas. Después procederemos a aplicar sus presupuestos a los modelos prototípicos de la macrofunción narrativa recogidos en el *PCIC* para mostrar su capacidad descriptiva[10].

[8] Los índices primarios son: a) Longitud de la unidad terminal (LUT): mide el promedio de palabras por UT de un texto, siendo la UT un fragmento de texto que contiene a la oración principal y todas las oraciones subordinadas dependientes de ésta y se caracteriza en el discurso escrito por empezar en mayúsculas y finalizar en punto o interrogación de cierre. b) Índice de subordinación (IS): mide el porcentaje de cláusulas (tanto principales como subordinadas) en relación con las UU.TT. de un texto. c) Longitud de la cláusula. Mide el promedio de palabras por claúsula.

[9] La autora señala tres problemas como posibles causas de estos resultados: *a)* posibles problemas de representatividad estadística de la muestra, *b)* problemas de fiabilidad de los índices de madurez sintáctica de Hunt para "medir el tipo de elaboración creciente sintáctica de los textos" y *c)* el hecho de que los textos seleccionados por la autora "no sean consistentemente más complejos". (Checa García 2006:12).

[10] La fiabilidad de los modelos del *PCIC* no es objeto de análisis propiamente dicho.

Centro Virtual Cervantes

2.1. *Frecuencia de uso y densidad léxica*

Es de sobra conocido que la frecuencia de uso de las palabras condiciona las posibilidades de comprensión lectora de un texto[11]. Tal supuesto puede articularse metodológicamente en un indicador de evaluación que mida el número de palabras poco usadas por unidad textual, de forma que se establezca que cuantas más palabras poco usadas contenga un texto más difícil será su comprensión para un aprendiz de español como L2[12]. Al mismo tiempo, la variación léxica que muestre un texto es un indicador de la complejidad del mismo y podemos suponer que a mayor densidad léxica (mayor número de palabras diferentes por texto) mayor dificultad para la comprensión lectora. Este supuesto es admitido en la práctica por los promotores de textos graduados, pues rara es la propuesta de lecturas adaptadas que no determine la graduación de las lecturas en función del número de palabras diferentes que contiene cada uno (a menudo, éste es el único criterio que determina los niveles de lectura). De acuerdo con lo anterior, la frecuencia de uso y la densidad (o carga) léxica son variables que deben contribuir eficazmente a determinar el grado de legibilidad de un texto. En anula (2007a) hemos propuesto las siguientes medidas para evaluar estas propiedades:

Índices	Fórmula
• **IPBF** (Índice de palabras de baja frecuencia)	• Nº total de palabras léxicas de baja frecuencia · 100 / Nº total de palabras léxicas.
• **IDL** (Índice de densidad léxica)	• Nº de palabras léxicas o de contenido (nombre, verbo, adjetivo y adverbio) distintas / Nº total de segmentos discursivos.
• **ICL** (Índice de complejidad léxica)	• IPBF + IDL / 2

Tabla 2. Índices de evaluación de la complejidad léxica.

2.2. *Longitud y naturaleza de los segmentos discursivos*

Si pasamos del estudio del léxico al estudio de las construcciones sintácticas conviene proponer una unidad de análisis que sea representativa del texto como unidad discursiva. En este sentido, conviene superar el nivel de la cláusula u oración como unidad básica. En Anula (1998 y 2000), basándonos en un análisis de la estructura sintáctica de la unidad "discurso" propuesto por Garrido (1997: 186), establecimos como unidad de análisis del discurso oral el concepto de *segmento discursivo*[13]. Recientemente, en Anula (2007a), hemos extendido dicha definición para referirnos a los textos escritos, de forma que, por *segmento discursivo* entendemos "una oración, una frase, un grupo de oraciones o frases sintácticamente trabado que está comprendido entre dos pausas ortográficas que confieren al segmento independencia sintáctica" (Anula 2007a). Según esta definición, el texto de (1) consta de dos segmentos discursivos, tal y como representamos en (11) mediante la utilización de una barra vertical y una letra identificadora de cada segmento:

(11) *a*. Estoy veteado de hierro, de plata y del barro más burdo. | *b*. No puedo contraerme en ese puño firme que aprietan aquellos que no dependen de estímulos. |

En ambos casos la frontera entre los dos segmentos está determinada por un punto que funciona como una pausa ortográfica[14]. Por otra parte, la comparación entre los dos segmentos discursivos de (11) nos muestra que ambas unidades son diferentes desde el punto de vista de su constitución formal: (11a) es un segmento unitario y (11b) es un segmento múltiple[15].

La longitud de la unidad discursiva y su naturaleza simple o sencilla (o en otros términos, unitaria o múltiple) son aspectos tenidos en cuenta en la mayoría de los manuales de estilo que se encargan de

[11] Véase en Belinchón, Igoa y Rivière (1992) un resumen de los efectos experimentales registrados en la investigación psicolingüística relacionados con la frecuencia de uso de las palabras. En relación con el papel que desempeña la frecuencia de uso de las palabras en la adquisición del léxico por aprendices de español como L2 pueden verse los trabajos de Baralo (1996) y Gómez Molina (1997). Un trabajo que sigue de cerca los resultados de las investigaciones psicolingüísticas sobre el reconocimiento léxico aplicado a la enseñanza del español como L2 para inmigrantes analfabetos es el de Villalba Martínez y Hernández García (2000).

[12] Naturalmente, la cuestión clave para hacer operativa una variable de este tipo es contar con indicadores de frecuencia léxica fiables. En la bibliografía sobre este tema contamos con algunas referencias para el español, aunque no todos miden de igual manera la frecuencia. Véanse, entre otros, Alameda y Cuetos (1995), Justicia Justicia (1995), Terrádez Gurrea (2001) y Almela, Sánchez y Sarmiento (2005). Para caracterizar la frecuencia de uso del léxico utilizado en los textos adaptados a la enseñanza de español como L2, utilizaremos el sistema de valores recogido en el *Gran Diccionario de uso del español actual* (*GDUEA*), previo contraste con los valores equivalentes en el corpus *CREA* de la Real Academia Española. El *GDUEA* estable que el umbral de baja frecuencia de una palabra es aquel que, respecto un corpus determinado (en su caso el Corpus Cumbre), la voz tiene menos de cinco ocurrencias por millón de palabras (lo que representa un porcentaje del 0,25 por millón (véase *GDUEA*, pág. 10).

[13] Entonces definimos este concepto como "una oración, una frase, un grupo de oraciones o frases sintácticamente trabado que está comprendido entre dos pausas fonológicas marcadas, posee un tonema propio y goza de autonomía sintáctica" (Anula 1998: 219).

[14] El punto y seguido no es la única manifestación de una pausa ortográfica que da fin a un segmento discursivo. Más adelante, en la nota 20, especificamos los indicadores ortográficos que dan lugar al cierre de segmentos discursivos.

[15] Son segmentos unitarios aquellos que están constituidos por una única cláusula u oración, como en (11a); y son segmentos múltiples aquellos formados por más de una cláusula u oración, como en (11b).

Centro Virtual Cervantes

establecer los principios de una escritura eficaz comunicativamente. Así, por ejemplo, los libros de estilo periodístico señalan de modo generalizado que las oraciones no deben exceder de las 15/20 palabras por oración para que transmitan eficazmente el mensaje. De igual manera, los manuales generales de escritura en español suelen recoger este criterio como un mecanismo para que la escritura sea "fluida" y "clara". Así lo hace, por ejemplo, el manual del Instituto Cervantes *Saber escribir* cuando señala la "preferencia por las *oraciones breves*, sencillas y simples" (op. cit., pág. 328). Por todo ello, el aligeramiento de la prosa mediante la limitación del número de palabras por segmento discursivo y la utilización preferente de segmentos unitarios frente a múltiples resulta un recurso útil para favorecer la comprensión lectora de los textos en general y es particularmente necesario en los materiales adaptados o graduados para la adquisición de la competencia lectora en la clase de español como L2 en sus niveles inicial e intermedio. Teniendo en cuenta estas consideraciones, en Anula (2007a) hemos propuesto los siguientes índices de evaluación de la complejidad de los segmentos del texto[16]:

Índice	Fórmula
• **ILS** (Índice de longitud segmental)[17]	• Nº total de palabras / Nº total de segmentos discursivos
• **ISM** (Índice de segmentos múltiples)	• Nº total de segmentos discursivos múltiples · 100 / Nº total de segmentos discursivos
• **ICS** (Índice de complejidad segmental)	• ILS + ISM / 2

Tabla 3. Índices de evaluación de la complejidad de los segmentos del texto.

2.3. ***Complejidad y recursividad oracional***

La tercera de las dimensiones propuestas para el análisis de la complejidad sintáctica se refiere fundamentalmente a la subordinación sintáctica, pues este procedimiento suele identificarse como un rasgo de la complejidad formal de un texto. Nuestra propuesta se plantea dos variables. Por un lado, analizar la carga de subordinación presente en el texto mediante el recuento de cláusulas oracionales subordinadas a otros constituyentes (ya sean oracionales o no) y, por otro, analizar el nivel de incrustación y/o recursividad existente en la estructural oracional. Estas variables se evalúan a través de los siguientes índices:

Índice	Fórmula
• **IPSub** (Índice de predicados verbales en oraciones subordinadas)	• Nº total de predicados verbales de oraciones subordinadas · 100 / Nº total de predicados verbales.
• **IRO** (Índice de recursividad oracional)	• Nº total de predicados verbales con doble incrustación · 100 / Nº total de predicados verbales.
• **ICE** (Índice de complejidad estructural)	• IPsub + IRO / 2

Tabla 4. Índices de evaluación de la complejidad oracional.

El cálculo del IPSub se establece para determinar la presencia porcentual de estructuras subordinadas frente a oraciones simples (ya estén aisladas o formen parte de estructuras coordinadas o yuxtapuestas), fragmentos no oracionales, etc. El cálculo del IRO analiza las estructuras que manifiestan propiedades recursivas (si bien ceñidas al ámbito oracional; esto es, no se tiene en cuenta el fenómeno en el ámbito nominal) que muestran un nivel de incrustación o profundidad de grado secundario. Las estructuras que entran bajo su análisis son de dos tipos. Por un lado, se computan las construcciones sintácticas subordinadas cuya cláusula o elemento subordinante está subordinado a su vez a otra estructura sintáctica. Es decir, construcciones sintácticas que se ajustan al esquema de (12):

(12) [Oppal [O sub^{1} [Osub^{2}]]]

Dicho esquema valora por tanto que el nivel de incrustación de una estructura subordinada sea de orden secundario. Por otro, el IRO analiza un tipo de construcciones complejas en las que también está presente la cuestión de la doble incrustación, si bien participa en su realización las construcciones coordinadas y yuxtapuestas, y no sólo las subordinadas. Los esquemas que dan cuenta de este tipo de estructuras son los siguientes:

[16] El ILS y el ISM se apoyan en los supuestos de que cuanto más elevado es el número de palabras de que consta un segmento y cuanto más segmentos discursivos sean múltiples, más complejo resultará el texto y más difícil será su legibilidad.

[17] El ILS guarda relación con medidas tradicionales de la investigación de la forma lingüística de los textos, como el índice conocido como "longitud media de los enunciados" o el índice de "longitud de la unidad terminal mínima" propuesto por Hunt.

Centro Virtual Cervantes

(13) {OCpta <1º/2º miembro [O ppal. [OSub]]>}
(14) {OCpta <1º/2º miembro { <1º miembro> <2º miembro>}>}
(15) [Oppal {<1º miembro> <2º miembro>}]

(13) recoge un tipo de unidades en las que un miembro de una oración compuesta (con independencia de que sea yuxtapuesta o coordinada) es una oración compleja; (14) recoge unidades en las que un miembro de una oración compuesta por coordinación o yuxtaposición está constituido a su vez por una oración compuesta (es decir, hay una estructura coordinada o yuxtapuesta dentro de otra estructura coordinada o yuxtapuesta); finalmente, (15) computa unidades en los que una oración compleja está formada por una oración principal cuya cláusula subordinada está constituida por una estructura coordinada. En los enunciados siguientes enumeramos algunos ejemplos *ad hoc* de cada tipo:

(16) [Juan quiere que [María recite el poema que [escribió ayer]]]
(17) { < [Juan quiere que [María recite el poema]] > pero < no se lo ha dicho > }
(18) { < María recitó el poema > y { < Juan se mostró contento > pero < no lo manifestó > } }
(19) [María esperaba que { < Juan escuchara el poema atentamente > y < se mostrara complacido }]

3. Aplicación del modelo de evaluación a los prototipos de macrofunción narrativa recogidos por el *PCIC*

Para ilustrar la aplicación de los índices propuestos a la valoración de la complejidad lingüística de los textos narrativos, tomaremos en consideración los textos de (20) a (22), propuestos en el *PCIC* como textos prototípicos de la macrofunción narrativa para los niveles A2, B2 y C2 respectivamente. Presentamos los textos segmentados de acuerdo con la definición de segmento discursivo expuesta en §2.2[18]:

(20) *a.* Vivo en Barcelona | *b.* El lunes pasado fui a la playa, está cerca de mi casa. | *c.* Era muy temprano, así que nadé, tome un poco el sol y me dormí. | *d.* Cuando desperté, había una mujer a mi lado. | *e.* La señora era muy guapa. | *f.* Era rubia, tenía el pelo largo y los ojos verdes. | *g.* Yo me senté en la toalla, ella se sentó en una silla más pequeña que mi bolso. | *h.* Me dijo: | *i.* «¿Eres española?». | *j.* Yo dije: | *k.* «No, soy inglesa»
(21) *a.* De niña solía ir siempre en verano a la playa, todavía me acuerdo de un día que hacía más calor que nunca, un calor horrible, y me fui a nadar un rato. | *b.* Cansada me tumbé en la arena y me dormí. | *c.* Al despertar, había una señora a mi lado. | *d.* Era una persona muy especial, tenía una elegancia natural. | *e.* Estaba sentada en una hamaca de la playa y no miraba al agua, sino al infinito.
(22) *a.* Cuando era pequeño solía subirme a un árbol del patio; | *b.* me divertía mirar a la gente sin que nadie me viera. | *c.* Aun a riesgo de caerme, me acomodaba entre las ramas más fuertes y desde allí podía ver toda la calle, al margen de lo que pasaba dentro de las casas. | *d.* Una vez vi algo realmente extraordinario, a veces creo que lo soñé, pero luego miro la foto que hice. | *e.* No, aquello sucedió realmente. | *f.* Un chico, muy seguro él, se acercó a mi árbol y sacó una llave del bolsillo, la introdujo en algún lugar que yo no veía y abrió el árbol como si fuera un armario.

El análisis de estos textos mediante los índices de evaluación de la complejidad lingüística propuestos nos ofrece los resultados recogidos en la tabla 5:

[18] Como puede deducirse de la segmentación expuesta en (20), interpretamos como segmentos discursivos diferentes cada una de las partes estructurales de las expresiones de estilo directo "(20*h*) Me dijo: / (20*i*) «¿Eres española?»" y "(20*j*) Yo dije: / (20*k*) «No, soy inglesa»", y eso a pesar de que estas secuencias se podrían interpretar, siguiendo a Concepción Maldonado (1991), como estructuras yuxtapuestas y a pesar de la aparente contradicción que presenta el hecho de interpretar como un único segmento discursivo las construcciones yuxtapuestas de (21*b*) o (22*d*), entre otras. Sin embargo, lo que deshace la contradicción, de acuerdo con la definición propuesta del concepto "segmento discursivo", es la pausa ortográfica que confiere independencia sintáctica a la expresión introductoria con verbo de lengua flexionado. En este sentido, conviene tener en cuenta que consideramos pausas ortográficas que conforman nuevos segmentos discursivos las siguientes: el punto y seguido, el punto y aparte, el punto y coma (si separa proposiciones yuxtapuestas –véase *Ortografía* (RAE 1999: §5.4.2)–, los dos puntos (si preceden a citas textuales, están usados en fórmulas de saludo o cortesía, o funcionan como mecanismos de enlace entre oraciones –véase *Ortografía* (RAE 1999: §5.3.2, §5.3.3 y §5.3.4)–, los puntos suspensivos (siempre que el enunciado que siga comience en mayúsculas y constituya una oración) y la exclamación e interrogación de cierre.

Centro Virtual Cervantes

Dimensiones de análisis	Índices de Evaluación	Texto nivel A2	Texto nivel B2	Texto nivel C2
Léxica	IPBF, expresado en porcentaje	0	2,7	0
	IDL, expresado en promedio	3	7	6,67
	ICL, expresado en promedio	1,5	4,85	3,34
Segmental	ILS, expresado en promedio	7,09	14,80	18
	ISM, expresado en porcentaje	45,45	100	83,33
	ICS, expresado en promedio	26,27	57,4	50,67
Oracional	IPSub, expresado en porcentaje	16,67	30,77	42,86
	IRO, expresado en porcentaje	11,11	7,69	38,10
	ICE, expresado en promedio	13,89	19,23	40,48
Las cifras se expresan en puntuaciones brutas				

Tabla 5.

Si consideramos ahora los indicadores de complejidad léxica (ICL), segmental (ICS) y estructural (ICE) (cada uno de ellos aglutina las variables de sus dimensiones respectivas) y los unimos en un único indicador que evalúe globalmente el texto, como el que proponemos en la tabla 6, obtendríamos unos resultados que expresarían la complejidad lingüística de los textos sujetos a evaluación. Estos son los resultados que se presentan en la tabla 7 para los textos del *PCIC* examinados.

Índice	**Fórmula**
IGC: Índice global de complejidad textual	ICL + ICS + ICE / 3

Tabla 6.

Índice	Texto Nivel A2	Texto Nivel B2	Texto Nivel C2
IGC	13,89	27,16	31,50

Tabla 7. Valores de complejidad lingüística de los textos.

En relación con la noción de *complejidad lingüística*, lo que manifiesta la tabla 7 es que los textos están graduados en su complejidad formal de manera que el texto de Nivel A2 tiene el menor grado de complejidad (13,89) y el texto de nivel C2 posee el de mayor grado de complejidad (31,50), tal y como cabría esperar que fuera si los textos propuestos por el Instituto Cervantes representan textos prototípicos de la macrofunción narrativa graduados en su nivel de complejidad de acuerdo con los presupuestos del *MCER* y del *PCIC*[19].

4. Conclusión y limitaciones del estudio

Como principal conclusión del trabajo queremos destacar, con independencia de los resultados anteriormente señalados, la capacidad descriptiva del modelo que se intuye tras el escueto análisis realizado. Como hemos podido apreciar, el sistema de valoración de la complejidad lingüística basado en el análisis conjunto de seis indicadores diferentes que recogen aspectos básicos de tres dimensiones significativas de la microestructura del texto es suficientemente discriminativo y nos permite categorizar los textos según el grado o nivel de dificultad que presente su forma lingüística. Por otra parte, conviene señalar que la validez del modelo expuesto no se basa en una o en otra medida aislada, sino en el conjunto de todas ellas, convenientemente recogido en el indicador que hemos denominado *Índice global de complejidad textual*. Solo si utilizamos un modelo de evaluación de la complejidad lingüística que evalúe conjuntamente propiedades léxicas y propiedades sintáctico-discursivas como las aquí analizadas podremos garantizar un sistema de determinación de los niveles de dificultad de los textos acorde con las dos claves más importantes de la adquisición de la competencia lectora por aprendices de L2: el conocimiento del vocabulario y la familiarización con las diferentes estructuras sintácticas propias de la gramática del español.

[19] El hecho de que los valores resultantes no sean homogéneos, pues la distancia que separa los valores obtenidos por los textos B2 y C2 es escasa (4,34 puntos), si la comparamos con la diferencia existente entre las puntuaciones obtenidas por los textos A2 y B2 (13,27), sugiere la necesidad de determinar una escala empírica real de la complejidad lingüística de los textos y géneros discursivos que fije unos valores de referencia para cada nivel de competencia comunicativo-lingüística de los aprendices de español como L2 que pueda luego aplicarse a la enseñanza-aprendizaje del español como L2. En la actualidad estamos trabajando en la elaboración de una escala de complejidad formal basada en el análisis de textos narrativos reales a partir de las medidas de evaluación expuestas aquí. Los resultados provisionales de la investigación sugieren que la escala empírica recoge una horquilla de variación más amplia que la obtenida de los textos aquí examinados.

Centro Virtual Cervantes

Además de esto, huelga decir que un modelo como el aquí expuesto debe ser refrendado por una investigación empírica de doble alcance. Por un lado, el modelo debe probar su capacidad descriptiva con un abanico amplio de textos (no solo narrativo-descriptivos) reales (entendiendo por tal, textos no manipulados con fines específicos). Por otro, el modelo debe justificar experimentalmente que los diferentes niveles de clasificación de los textos según su complejidad lingüística correlacionan más o menos directamente con diferentes niveles de comprensión lectora, tanto en aprendices de español como L2, como en otros grupos de sujetos (sean personas con problemas de lectura o sean niños y adolescentes en proceso de adquisición de las destrezas lectoras). Hasta que tengamos esta investigación empírica convenientemente desarrollada, el modelo expuesto solo tiene un valor provisional de carácter teórico y propedéutico.

Bibliografía

ACADEMIA ESPAÑOLA, REAL (1999), *Ortografía de la lengua española*, Madrid, Espasa Calpe.

ACADEMIA ESPAÑOLA, REAL, Banco de Datos CREA [en línea], Corpus de referencia del español actual, <http://www.rae.es>

ACQUARONI MUÑOZ, Rosana (2004): «La comprensión lectora», *Vademécum para la formación de profesores. Enseñar español como segunda lengua (L2)/lengua extranjera (LE)*, Madrid: SGEL, 943-964.

ACQUARONI MUÑOZ, Rosana (1994): *El sueño de Otto*, Madrid: Santillana.

ALAMEDA, José Ramón y Fernando CUETOS (1995): *Diccionario de frecuencias de las unidades lingüísticas del Castellano*, Oviedo: Departamento de Psicología: Universidad de Oviedo.

ALMELA Pérez, Ramón, Aquilino SÁNCHEZ y Ramón SARMIENTO (2005): *Frecuencias del español: diccionarios y estudios léxicos y morfológicos*, Madrid, Editorial Universitas.

ANULA, Alberto (1998): *La capacidad lingüística en la esquizofrenia: estudio sintáctico sobre la producción de oraciones en el discurso esquizofrénico*. Tesis doctoral no publicada, Universidad Autónoma de Madrid.

ANULA, Alberto (2000): «La evaluación de la complejidad sintáctica en la producción de discursos orales: marco teórico y requisitos metodológicos para el análisis de la capacidad lingüística», en J.J. de Bustos Tovar, P. Charaedeau, J.L. Girón Alconchel, S. Iglesias y C. López Alonso (eds.), *Lengua, discurso y texto (I Simposio Internacional de Análisis del Discurso)*, Madrid: Visor, vol. I, pp. 727-739.

ANULA, Alberto (2007a), «Tipos de textos, complejidad lingüística y facilitación de la lectura», *Actas del IV Congreso de la Asociación Asiática de Hispanistas*, Manila, en prensa.

ANULA, Alberto (2007b), «La facilitación de la lectura y su aplicación a la enseñanza del español como L2», *Actas del V Congreso Nacional de Investigaciones Lingüístico-Filológicas. La enseñanza de lengua(s) en el tercer milenio*, Lima: Ed. en CD-Rom.

ARMAS MARCELO, Juan José (1994): *Madrid, distrito federal*, Barcelona: Seix Barral.

BARALO OTTONELO, Marta (1996): «Algunos aspectos de la adquisición de la morfología léxica del español como lengua extranjera», En S. Montgesa y P Gomis (eds.). *Tendencias actuales en la enseñanza del español como lengua extranjera. Actas del V Congreso Internacional de ASELE*, Málaga: Universidad de Málaga, tomo I, 143-150. Disponible en línea http://cvc.cervantes.es/obref/antologia_didactica/morfologia/baralo.htm#notat.

BELINCHÓN, Mercedes, José Manuel IGOA y Ángel RIVIÈRE (1992): *Psicología del lenguaje. Investigación y teoría*, Madrid: Editorial Trotta.

CERVANTES, Instituto (2002): *Marco común europeo de referencia para las lenguas: aprendizaje, enseñanza, evaluación*, Madrid: Secretaría General Técnica del MECD y Grupo Anaya [Traducción de *Common European Framework for Languages: Learning, Teaching, Assessment*, Strasbourg: Council for Cultural Cooperation Education Committe, 2001]

CERVANTES, Instituto (2006): *Saber escribir*, Madrid: Aguilar.

CERVANTES, Instituto (2007): *Plan Curricular del Instituto Cervantes. Niveles de referencia para el español*, Vol I, II y III, Madrid: Editorial Biblioteca Nueva.

CERVANTES, Instituto (España): *Lecturas paso a paso* [En línea] http://cvc.cervantes.es/aula/lecturas/.

CHECA García, Irene (2006): «Medidas de madurez sintáctica aplicadas a lecturas de ELE», *Interlingüística*, 16, 1-15 [disponible en línea: http://www.ajl.uma.es/Pdf/31.pdf].

DE MIGUEL, L. y A. SANTOS (1987): *Doce a las doce*, Madrid: Edelsa.

GARRIDO, Joaquín (1997): *Estilo y texto en la lengua*, Madrid: Gredos.

Centro Virtual Cervantes

GÓMEZ MOLINA, José Ramón (1997): «El léxico y su didáctica: una propuesta metodológica», *REALE*, vol. 8, 69-93 [disponible en línea: http://cvc.cervantes.es/obref/antologia_didactica/morfologia/gomez.htm#notat].

GRACIA, Ángela (2006), "La lectura – una destreza imprescindible para la adquisición de Español como lengua extranjera", en *Didáctica (Lengua y Literatura)*, vol. 18, pp. 147-161.

HADDON, Mark (2004): *El curioso incidente del perro a Medianoche*, Barcelona: Ed. Salamandra [Traducción de *The Curious Incident of the Dog in the Night-Time*, Random House Group Ltd.].

HUNT, Kellog W. (1964): *Differences in grammatical structures written at three grade levels, the structures to be analized by transformational methods*. Tallahassee [Fl]: Florida State University.

HUNT, Kellog W. (1967): *Sentence structures used by superior students in grades four and twelve, and by superior adults*. Tallahassee, [Fl]: Florida State University.

HUNT, Kellog W. (1970): «Recent measures in syntactic development», *Reading in applied transformational grammar*, Mark Lester (ed.), New York: Holt, Rinehart and Wiston, 193-201.

JUSTICIA JUSTICIA, Fernando (1995): *El desarrollo del vocabulario. Diccionario de frecuencias*, Granada: Serv. de Pub. de la Universidad de Granada.

MALDONADO, Concepción (1991): *Discurso directo y discurso indirecto*, Madrid: Taurus.

MIQUEL, Lourdes y Neus SANS (1995): *Lejos de casa*, Barcelona: Difusión.

QUINTERO, Ednolio (1991): *La danza del jaguar*, Caracas: Monte Ávila.

SÁNCHEZ, Aquilino (dir.); Alberto Anula, Pascual Cantos, Lourdes Cerezo, Juan Camilo Conde, Maria Dueñas, Fernando Huerta, Francisco Iborra y Mar Vilar (2001). *Gran diccionario de uso del español actual*, Madrid, SGEL.

TERRÁDEZ GURREA, Marcial (2001): *Frecuencias léxicas del español coloquial: análisis cuantitativo y cualitativo*, Valencia: Servicio de publicaciones de la Universidad de Valencia.

VILLALBA MARTÍNEZ, Félix y M.ª Teresa HERNÁNDEZ GARCÍA (2000): ¿Se puede aprender una lengua sin saber leer? Alfabetización y aprendizaje de una L2, *Carabela 48*, 2000, 85-110 [Disponible en línea http://cvc.cervantes.es/obref/antologia_didactica/inmigracion/villalba.htm].

Centro Virtual Cervantes