

# MEDIDAS DE MADUREZ SINTÁCTICA

## APLICADAS A LECTURAS DE ELE

IRENE CHECA GARCÍA  
Universidad de Almería

*RESUMEN.* Se analiza un corpus de lecturas de ELE (integrado por 30 fragmentos) para determinar si los niveles de dificultad asignados, además de exhibir una riqueza creciente de vocabulario, presentan una progresiva complejidad sintáctica medida por los índices de madurez sintáctica de K.W. Hunt. Los resultados indican un crecimiento continuado de los tres índices primarios, sólo consistentemente significativo para la longitud de la unidad terminal. A partir de estos y otros resultados se indican líneas y modos futuros posibles de investigación.

*PALABRAS CLAVE.* Aprendizaje de L2, comprensión lectora, complejidad sintáctica, madurez sintáctica.

*ABSTRACT.* We analyzed a 30 fragment set of Spanish as a second language readings in order to determine if their assigned difficulty consists of both lexical variety and syntactic complexity as measured by K. W. Hunt's proposed syntactic maturity indices. Results indicate a steady increase of the indices, significant within the T-unit length. These and other supporting results suggest new directions for research.

*KEY WORDS.* L2 acquisition, reading comprehension, syntactic complexity, syntactic maturity.

### 1. MEDICIÓN DE LA MADUREZ SINTÁCTICA

La preocupación por mejorar la calidad de expresión de los escolares, así como por detectar posibles deficiencias, y la necesidad de trazar para todo ello las líneas generales de su desarrollo, ha conducido a investigadores de varios campos a buscar medidas del desarrollo de distintas competencias lingüísticas y comunicativas, entre ellas del desarrollo sintáctico. A tal fin, K.W. HUNT (1964), basándose en unidades anteriores diseñadas por otros investigadores (L. LA BRANT 1933, W. LOBAN 1963), propone unos índices de madurez sintáctica de naturaleza cuantitativa, lo que permite el estudio de su variación en función de factores diversos, tales como edad y curso –para comprobar que efectivamente se trata de medidas válidas del desarrollo sintáctico- o variables sociales de lo más diversas (clase social, tipo de centro, coeficiente intelectual, rendimiento escolar, etc.). Entre estos índices destacan, por ser los más frecuentemente aplicados, los índices primarios: longitud de la unidad terminal (LUT), longitud de la cláusula (LCL) e índice de subordinación (IS). Dichos índices se basan en dos unidades fundamentales de análisis: la unidad terminal (UT) y la

cláusula (CL), que definimos a continuación para pasar después a especificar en qué consiste cada uno de los índices primarios.

UT: Cláusula principal más todas las subordinadas que se añadan, resultando un fragmento terminable con un punto sin que quede ningún resto. Se incluyen los periodos interordinados.<sup>1</sup>

CL: Una oración simple o una proposición –principal o subordinada– (con o sin sujeto expreso). Se incluyen formas no personales del verbo (salvo que funcionen en perífrasis verbales).

LUT (Longitud de la Unidad Terminal): total de palabras / total de unidades terminales.

IS (Índice de Subordinación): total de cláusulas / total de unidades terminales.

LCL (Longitud de la CLáusula): total de palabras / total de cláusulas.

Dadas estas definiciones se demuestra fácilmente (I. CHECA GARCÍA 2004: 91-93) que  $LUT = IS \times LCL$ . La relación matemática se explica lingüísticamente porque LUT puede aumentarse añadiendo más cláusulas –aumentando así el IS– o añadiendo más complementos dentro de una misma CL –aumentando así la LCL. En ambos casos, se amplía la información –las proposiciones básicas– en una única UT.

## 2. APLICACIÓN DE LOS ÍNDICES DE MADUREZ SINTÁCTICA AL APRENDIZAJE DE SEGUNDAS LENGUAS

Una vez que resultados experimentales refrendaron los índices primarios de HUNT, pronto se quiso comprobar su eficacia para medir diferencias a lo largo del aprendizaje de segundas lenguas. Los índices de madurez sintáctica presentaban el atractivo de ser medidas «positivas» –esto es, de logros y no de carencias– de la competencia sintáctica, al contrario que las medidas usuales de proporción de errores. Sin embargo, la aplicación de los índices primarios a la medición de dichas diferencias para el aprendizaje de L2 se topa con resultados que sugieren la necesidad de incorporar de algún modo también la medición de los errores –esto es, la medición «negativa» de la competencia aprendida. En concreto, con respecto a la medida resumen, la LUT, se observa que: (i) es menos adecuada para los niveles iniciales, en los que aún hay poca madurez sintáctica y las diferencias son mínimas, (ii) no distingue bien entre grupos próximos entre sí y, (iii) no refleja si la LUT aumenta a costa de un notable número de errores, y dado que la LUT no se ha visto excesivamente correlacionada con el número de errores, parece necesario tener en cuenta ambos rasgos de la competencia sintáctica. Por todas estas razones, se incorpora la medición de errores a los índices de madurez sintáctica, en la forma de una modificación de la unidad cuya longitud se tiene en cuenta, pasando a considerarse la denominada Unidad Terminal Libre de Errores (error-free T-unit)<sup>2</sup>, que sin embargo no tiene una definición unánime debido a la variabilidad de la noción de error.

<sup>1</sup> La definición de UT y de CL se basan en la adaptación a partir de K.W. HUNT (1964: 27 para la primera definición de UT, 1967:8-9 para la justificación de su denominación, 1970<sup>a</sup>: 4; 1970<sup>b</sup>: 4 para la definición más precisa de CL) propuesta en I. CHECA GARCÍA (2004: 409 para CL y 416 para UT) a partir, a su vez, de la adaptación para el español de M. VÉLIZ DE VOS (1986: 53 para UT y 50 para CL). Los detalles de segmentación pueden verse allí (capítulo V.7.1.2. para UT y V.7.1.3. para CL).

<sup>2</sup> M. SCOTT & G.R. TUCKER (1974), D. LARSEN-FREEMAN & V. STROM (1977) entre otros.

En este trabajo, sin embargo, nos ocupa la aplicación de estos índices a la medición de la complejidad sintáctica no ya de eductos, sino de aductos, en concreto lecturas de ELE, clasificadas en distintos niveles de dificultad. Conviene aquí hacer una distinción importante entre complejidad y madurez sintácticas. Los índices de madurez sintáctica se llaman «de madurez» y no «de complejidad» porque no buscan caracterizar el discurso en función de su complejidad sintáctica, que podría definirse de maneras varias, sino en función del avance por cursos. Es decir, son índices válidos para la medición de la variación en función de la edad, para medir el progresivo desarrollo –o su estancamiento si fuera el caso. A pesar de ello, se han utilizado en algunas lenguas para estudiar la elaboración sintáctica de aductos para alumnos de una L2. Al respecto conviene señalar que las dificultades que una mayor complejidad sintáctica supone podrían ser distintas en una L2 y una L1 por estar debidas a razones diferentes. Mientras que en una L1 la impericia sintáctica y muchas veces un deficiente desarrollo conceptual pueden explicar muchos de los errores, en L2, al menos en el caso de los adultos, el desarrollo conceptual necesario se supone alcanzado y los errores se cometen por transferencia de una L1 o desconocimiento del mecanismo sintáctico en cuestión más que por dificultades para densificar la información incluyendo más constituyentes básicos en una misma UT o en una misma CL y, por tanto, no por una cuestión de madurez sintáctica en sí, aunque se refleje en las medidas de HUNT.

### 3. OBJETIVOS

Al clasificar en niveles de dificultad las lecturas para L2, el criterio más común es la diversidad de vocabulario, medida usualmente en número de palabras distintas que contiene la lectura. Sin embargo, no se hace alusión nunca –aunque acaso algunas editoriales lo indiquen a los encargados de elaborar la lectura– a la complejidad sintáctica que exhiban dichas lecturas. Los índices de madurez sintáctica, y en particular la LUT, han sido utilizados, no obstante, en varios estudios para comprobar la complejidad del discurso de los profesores en el aula de lengua extranjera, y también en alguna ocasión la complejidad de los textos empleados como aducto para una L2.

Se analizarán, pues, textos de lectura de ELE en función de estos índices para dilucidar si, conforme cabría esperar, la elaboración sintáctica de dichas lecturas es creciente según se avanza en el nivel de dificultad asignada. Una ausencia de correlación entre nivel de dificultad y madurez sintáctica indicaría tal vez una falta de atención hacia la complejidad sintáctica en estas lecturas, que aunque sería menos grave probablemente que en lecturas de español como L1, seguirían siendo objeto de posible corrección para la mejora de la ideonidad de estas lecturas.

### 4. METODOLOGÍA

#### 4.1. CORPUS, SEGMENTACIÓN Y VARIABLES INDEPENDIENTES (FACTORES)

Se ha dispuesto un corpus consistente en 30 fragmentos, clasificados en función de:

-nivel de dificultad: en un principio, siguiendo la clasificación de la biblioteca, se tuvieron en cuenta 6 niveles en función del número de palabras distintas en el texto, para posteriormente agruparlos, por necesidades estadísticas, en 3 grupos, de forma que en lugar de 5 observaciones por nivel dispusiésemos de

10. Estos tres niveles pueden denominarse: inicial (1), medio (2) y avanzado (3). El nivel inicial aglutina textos de entre 200 a 900 palabras, el medio de 900 a 1800 y el avanzado de 1800 a 3000.

-tipo de texto: original o adaptación de una obra literaria;

-colección a la que pertenece (pero no se ha dispuesto de volúmenes suficientes de cada colección para ambos tipos de textos y cada uno de los 6 niveles de dificultad) por lo que no se ha utilizado esta variable para el análisis estadístico.

Cada uno de los fragmentos se compone de todas aquellas unidades terminales que se habían de incluir para tener en cuenta al menos 200 palabras. Todos ellos son de tipo de narrativo<sup>3</sup>. Han sido después segmentados en palabras, unidades terminales y cláusulas, conforme a los criterios de segmentación especificados en I. CHECA GARCÍA (2004: V.7.).

Para detectar la significación estadística de las diferencias entre las medias se han empleado los test de diferencia de medias no paramétricos (Kruskal-Wallis), o el paramétrico ANOVA cuando ha sido posible, lo cual se indica en su caso.

#### 4.2. VARIABLES DEPENDIENTES (RESPUESTA)

Las variables dependientes o respuesta tenidas en cuenta han sido los tres índices primarios de K. W. HUNT definidos en 1.

### 5. RESULTADOS

#### 5.1. RELACIÓN ENTRE NIVEL DE DIFICULTAD E ÍNDICES DE MADUREZ SINTÁCTICA

La primera clasificación en 6 niveles de dificultad indica que hay una tendencia de los tres índices primarios a aumentar conforme lo hace el nivel de dificultad de las lecturas, con dos muy notables excepciones: el nivel 6 y sobre todo el 4 (véase Figura 1).

Los mismos resultados agrupados en tres niveles —según se hace necesario para algunas pruebas estadísticas— muestran mayor regularidad (Figura 2) y las medias y medianas de la tabla 1 señalan un crecimiento continuado.

Tanto en la tabla 1 como en las figuras 3 y 4 —más intuitivamente— se observa cómo la tendencia al crecimiento de los índices corre aproximadamente paralela al aumento de dificultad asignada a las lecturas analizadas, tanto en un tipo de textos como en otro. Sin embargo, hay algunas notables excepciones que sirven además para distinguir el comportamiento de textos originales de aquellos adaptados. Los textos originales presentan un decrecimiento del IS entre los niveles 2 a 3 que afecta también a la media de la LUT. En los textos adaptados, en cambio, se observa un decrecimiento en las medias de todos los índices desde el nivel 1 al 2, de modo que el descenso en la LUT entre estos niveles de dificultad podría deberse tanto al descenso en IS como en LCL.

<sup>3</sup> La influencia del tipo de texto (narrativo, descriptivo, argumentativo) ha sido ampliamente estudiada y mayoritariamente refrendada para el español y para el inglés. Inició las pesquisas al respecto C.P.M. SAN JOSE (1972) para el inglés y M. VÉLIZ DE VOS ET ALII (1991) para el español. Puede verse un resumen de resultados y tendencias en I. CHECA GARCÍA (2004: 234 y ss. Capítulo III: 3.3.).

FIGURA 1: Evolución de las medias de los 3 índices primarios a lo largo de los 6 niveles de dificultad de las lecturas (de menor a mayor dificultad)

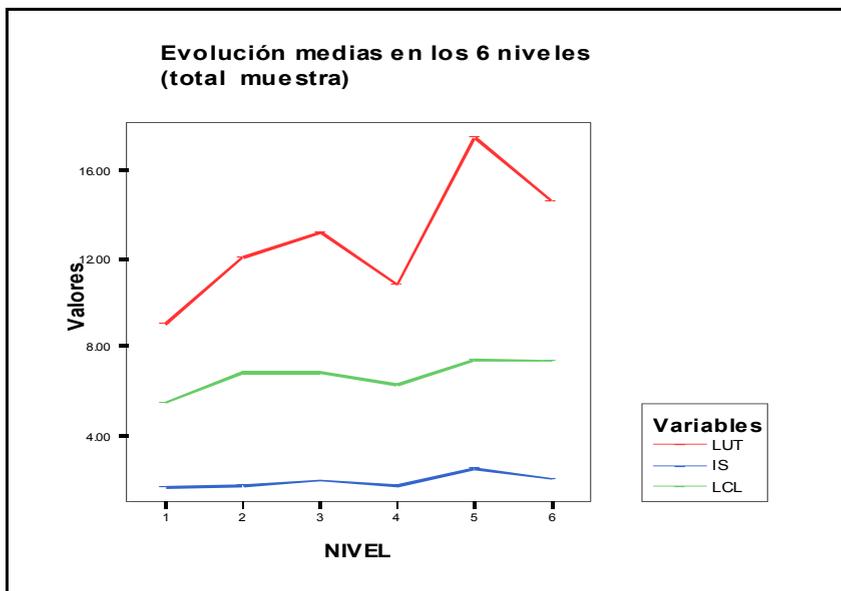


FIGURA 2: Evolución de las medias de los 3 índices primarios a lo largo de 3 niveles de dificultad de las lecturas (de menor a mayor dificultad)

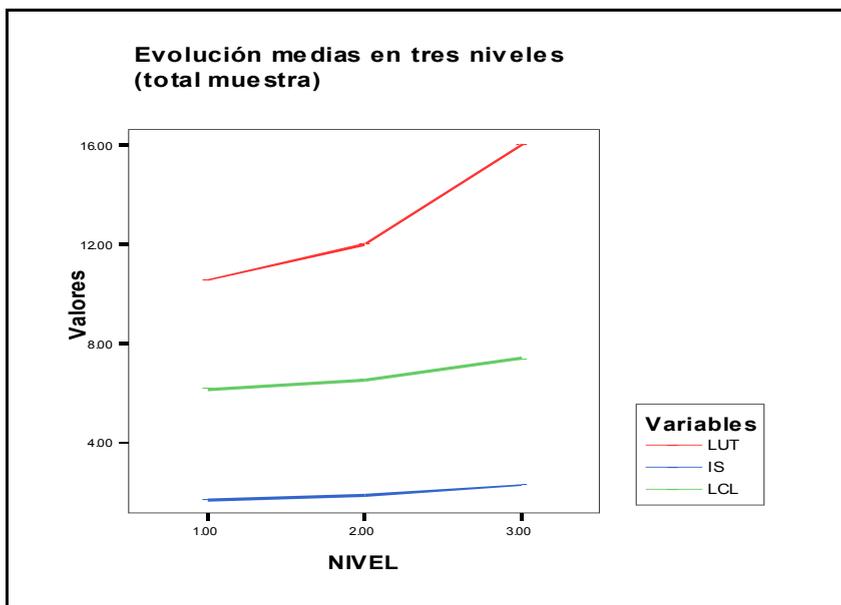


TABLA 1: Estadísticos descriptivos de tendencia central y dispersión en cada nivel (por tipos y totales)

Índice	Estadístico	TIPO TEXTO	NIVEL DE DIFICULTAD (en tres niveles)			
			Nivel 1	Nivel 2	Nivel 3	Total
LUT	Media	adaptación	13.358	11.620	18.676	14.551
		original	8.710	12.260	14.323	11.764
		Total	10.569	12.004	16.064	12.879
	Mediana	adaptación	12.847	11.035	19.432	13.156
		original	7.214	11.859	14.321	10.862
		Total	10.531	11.088	14.379	12.660
	Desv. tip.	adaptación	1.410	1.8652	4.1537	4.003
		original	4.049	2.702	3.8212	4.117
		Total	3.941	2.308	4.349	4.236
IS	Media	adaptación	1.964	1.823	2.937	2.241
		original	1.509	1.852	1.798	1.720
		Total	1.691	1.841	2.254	1.928
	Mediana	adaptación	1.980	1.800	2.811	2.011
		original	1.312	1.798	1.714	1.661
		Total	1.621	1.800	2.108	1.800
	Desv. tip.	adaptación	.325	.084	.848	.703
		original	.487	.351	.318	.400
		Total	.471	.266	.801	.591
LCL	Media	adaptación	6.946	6.397	6.587	6.643
		original	5.652	6.649	7.944	6.748
		Total	6.169	6.548	7.401	6.706
	Mediana	adaptación	6.985	5.914	6.608	6.355
		original	5.585	6.461	7.663	6.461
		Total	5.988	6.360	7.201	6.461
	Desv. tip.	adaptación	1.331	1.201	1.656	1.297
		original	.863	1.080	1.356	1.427
		Total	1.204	1.070	1.558	1.354

FIGURA 3: Medias de los índices primarios en los textos originales organizados en tres niveles de dificultad (de menor a mayor)

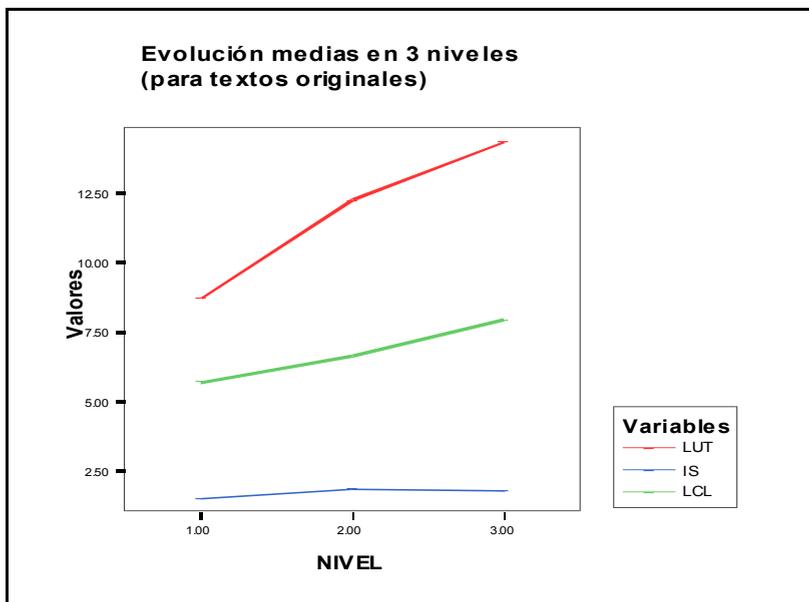
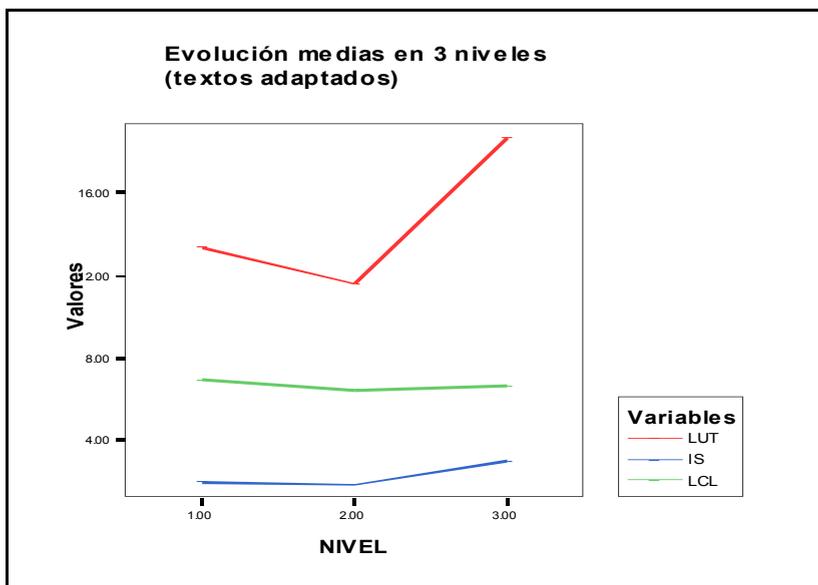


FIGURA 4: Medias de los índices primarios en los textos adaptados organizados en tres niveles de dificultad (de menor a mayor)



Para saber si estas diferencias entre niveles son significativas hemos de observar los resultados de las pruebas estadísticas de diferencia de medias empleadas y cuyos resultados se anotan en la tabla 2.

TABLA 2: Niveles de significación del nivel de dificultad como factor sobre los índices primarios de madurez sintáctica.

Índices	Niveles de significación			
	6 niveles (ambos tipos)	3 niveles (ambos tipos)	3 niveles txts originales	3 niveles txts adaptados
LUT	.047	.026	.054	.049
IS	.563	.210*	.161	.031
LCL	.157	.111*	.012	.779

\*Test paramétrico aplicado (ANOVA)

El nivel de dificultad incide sobre los índices de madurez sintáctica de distinta manera. Incide siempre sobre la LUT, de manera positiva además, de forma que cuando mayor es el nivel de dificultad mayor es también la LUT. Esta incidencia es más significativa cuando en lugar de 6 niveles, o en lugar de tener en cuenta sólo uno de los dos tipos de fragmentos, se tienen en cuenta todos los fragmentos y se clasifican en 3 niveles. Esto puede querer decir que la LUT discrimina elaboraciones sintáctica, pero de forma poca sutil, de modo que necesita más salto de nivel para poder discriminarlos, o, alternativamente, que los niveles se diferencian poco entre sí en función de la LUT. En cuanto al IS, su variación sólo es significativa en función del nivel de dificultad de los textos en el caso de los adaptados clasificados en 3 niveles. Al contrario ocurre a la LCL, que varía significativamente según el nivel de dificultad de los textos cuando sólo tenemos en cuenta los originales.

## 5.2. RELACIÓN ENTRE TIPOS DE TEXTO E ÍNDICES DE MADUREZ SINTÁCTICA

Las medias de los textos procedentes de adaptaciones son siempre mayores que las de textos originales y más distanciadas de la poblacional, salvo en el caso de la LCL, para la que las adaptaciones presentan una media inferior a la poblacional y a la de los textos originales, aunque la diferencia aquí es escasa. Esto, además de comprobarse en la tabla 3, puede observarse de forma muy clara en los gráficos de dispersión de las figuras 5 a 7.

TABLA 3: Estadísticos descriptivos en cada tipo y en el total

ÍNDICE	TIPO	Media	Mediana	Desv. típ.	Rango	Mínimo	Máximo
LUT	adaptación	14.5512	13.1563	4.00330	12.68	10.10	22.78
	original	11.7641	10.8622	4.11741	15.24	6.03	21.27
	Total	12.8789	12.6597	4.23622	16.75	6.03	22.78
IS	adaptación	2.2413	2.0107	.70318	2.44	1.56	4.00
	original	1.7197	1.6607	.39984	1.31	1.15	2.46
	Total	1.9283	1.7995	.59061	2.85	1.15	4.00

ÍNDICE	TIPO	Media	Mediana	Desv. típ.	Rango	Mínimo	Máximo
LC	adaptación	6.6432	6.3548	1.29653	3.95	4.59	8.54
	original	6.7481	6.4607	1.42665	5.77	4.40	10.17
	Total	6.7061	6.4607	1.35406	5.77	4.40	10.17

TABLA 4: Niveles de significación del tipo de texto sobre los índices primarios en función del nivel de dificultad

Índices	Niveles de significación			
	Todos los niveles	En nivel 1	En nivel 2	En nivel 3
LUT	.077*	.114	.914	.257
IS	.015*	.171	.914	.019
LCL	.839*	.171	.476	.171

\*Test paramétrico aplicado (ANOVA)

La significatividad estadística de las diferencias en la tabla 3 se muestra en la tabla 4. El tipo de texto sólo tiene una influencia significativa sobre el IS, y eso sólo en el total de niveles y en el nivel 3. Los textos adaptados tienen un IS mucho más elevado, que parece estar detrás de la diferencia en la LUT también a favor de las adaptaciones.

FIGURA 5: Dispersión de la LUT y medias por tipos de texto y total

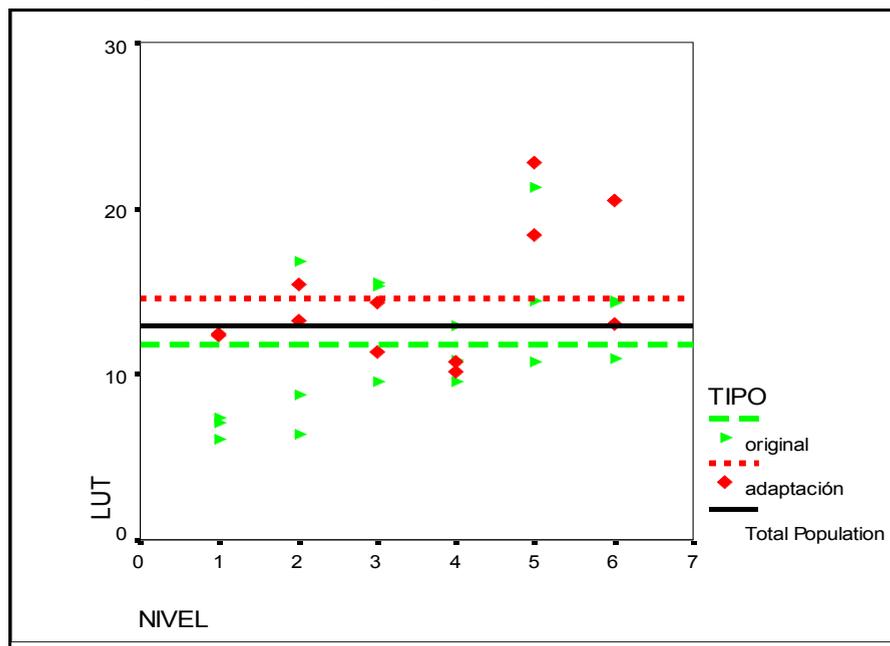


FIGURA 6: Dispersión del IS y medias por tipos de texto y total

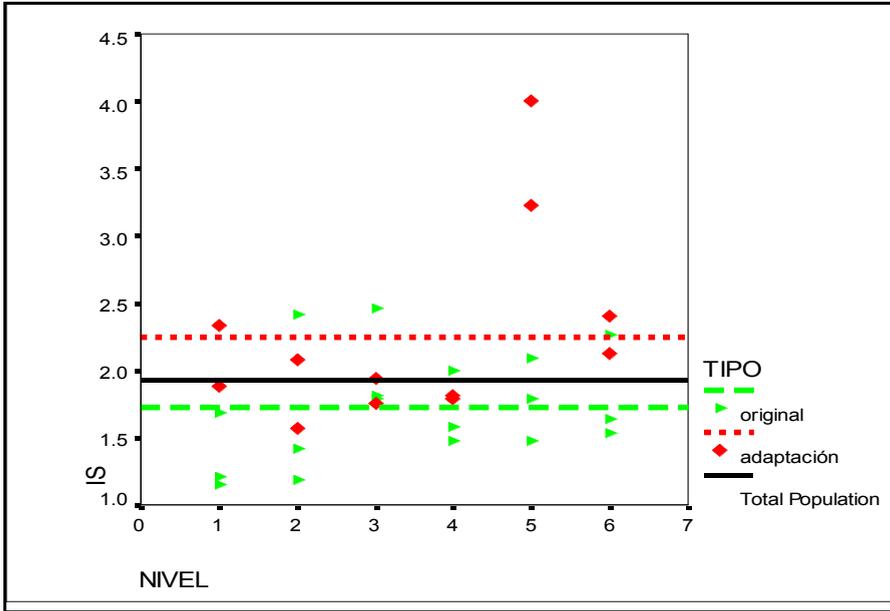
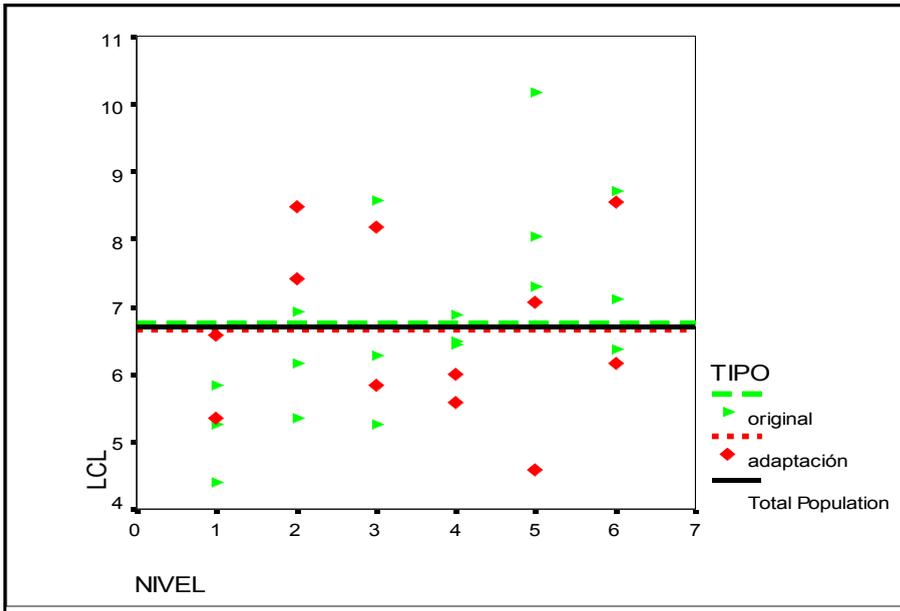


FIGURA 7: Dispersión de la LCL y medias por tipos de texto y total



Como breve apunte sobre la dispersión, y como muestran los gráficos anteriores, la dispersión es en general alta, pero homogénea entre niveles, salvo el nivel 3, en general más homogéneo, y el nivel 5, el más disperso. También se observa que los patrones de dispersión de la LUT son más similares a los del IS, siendo más disímiles de la dispersión de la LCL.

### 5.3. CORRELACIONES ENTRE LOS ÍNDICES

TABLA 5: Correlaciones de los 3 índices primarios en con los fragmentos según el tipo

Fragmentos	Índices	LUT	IS	LCL
TOTAL	LUT	1	.754 **	.582 **
	IS		1	-.068
	LCL			1
Originales	LUT	1	.797**	.807 **
	IS		1	.296
	LCL			1
Adaptaciones	LUT	1	.709 **	.321
	IS		1	-.405
	LCL			1

\*\*Significatividad al nivel 0.001

En consonancia con las medias de las figuras 6 y 7 y con las significaciones en la tabla 4, el IS influye decisivamente sobre la LUT en todos los casos. Sin embargo, la LCL sólo se correlaciona con la UT, esto es, aporta decisivamente a su longitud, en el caso del total y de los originales, pero probablemente en el total por esa significación tan alta de su aportación en los textos originales, ya que en los adaptados no es significativa.

## 6. CONCLUSIONES

### 6.2. RESULTADOS DE LOS ÍNDICES PRIMARIOS EN EL ESTUDIO PILOTO

El crecimiento de los estadísticos de tendencia central refleja un incremento general de los índices de madurez sintáctica conforme se avanza en el nivel de dificultad de los textos. Sin embargo, este crecimiento no siempre es continuado ni estadísticamente significativo. Tal como muestra la figura 1, el crecimiento en 6 niveles no es sostenido, sino que presenta dos puntos de ruptura, en el nivel 4 y en el 6. En el nivel 4 esta ruptura es muy abrupta, y se produce un descenso muy pronunciado. Este descenso puede deberse en parte a los resultados de los textos adaptados, que son los más bajos de todos los de su categoría. Cuando contemplamos la evolución de las medias agrupando los 6 niveles en 3 el crecimiento es continuado sin interrupciones, aunque algo más lento entre 1 y 2 que entre 2 y 3. Aquí parece pesar el descenso que experimentan los índices entre los niveles 1 a 2 en los textos adaptados, que se compensa por la el aumento algo más pronunciado en esa primera etapa de los índices primarios en los textos originales.

Sin embargo, los crecimientos señalados no son siempre significativos. La LUT es el único índice cuyo crecimiento es significativo a través de los niveles –tanto si consideramos 6 como si consideramos 3, aunque bastante más en este último caso– y a través de los tipos de textos, pues tanto si consideramos sólo los textos originales como si consideramos sólo los adaptados, su variación tiene una influencia significativa sobre la de la LUT, aunque rozando el límite en ambos casos. En cuanto a los otros dos índices, el IS sólo varía significativamente en los textos adaptados. Por el contrario, la LCL sólo lo hace en los textos originales. Esta diferencia de significación entre IS y LCL se explica también por que la LCL no se correlaciona con la LUT en los textos adaptados y sí lo hace, y más que el IS, en los originales. Es decir, en los textos adaptados lo que hace aumentar la LUT es el incremento de la LCL y puede decirse, por tanto, que principalmente se añade información en el nivel subclausal, esto es, se elaboran más los sintagmas. En cambio, en los textos originales no se elaboran apenas los sintagmas (como revela la nula correlación de LUT y LCL) aunque sí que lo hacen las oraciones gracias a mecanismos de subordinación. Las diferencias entre adaptaciones y originales, no obstante, sólo son significativas en el caso del IS, a favor de las adaptaciones. Igualmente, y aunque no significativas estadísticamente, las diferencias de LCL, aunque mucho menores, son favorables a los originales.

## 6.2. FUTURAS INDAGACIONES

Los resultados indican que de los índices de madurez sintáctica sólo la LUT refleja un aumento significativo de la elaboración sintáctica de los textos de lectura para ELE paralelo a su nivel de dificultad. Sólo un estudio con más textos podría descartar que no se trata de un problema de representatividad estadística, pero al margen de esta posible causa –que provoca que valores extremos puntuales incidan mucho no ya sólo sobre la media, sino sobre la mediana–, podría haber otras. Resulta llamativo, empero, que la LUT sí se correlacione en todos los casos estudiados (6 niveles o 3 y todos los textos o por tipos) con el nivel de la lectura. Más aún, cuantas más observaciones integran un grupo, mayor es la significación de su correlación. Otro motivo no del todo descartable es la fiabilidad de los índices de madurez sintáctica para medir el tipo de elaboración creciente sintáctica de los textos. El otro, obvio, es que los textos no sean consistentemente más complejos (lo que una ojeada al corpus hará más evidente aún).

Sea como fuere, debe preguntarse si lo que miden estos índices, con la segmentación propuesta –inicialmente para escolares adolescentes en su L1–, es lo que nos interesa medir, o, por el contrario, convendría ajustar detalles de la segmentación a las dificultades particulares que una L2 vs. una L1 y una competencia pasiva como la lectura, frente a la escritura, reviste, de modo que lo que es más difícil en un caso no tiene por qué serlo en otro. Un ejemplo puede ser la consideración de las perífrasis de infinitivo, pues si bien son de menor dificultad para un hablante nativo, pueden serlo y bastante para un aprendiz de L2.

Así pues, se impone realizar un estudio con un mayor número de observaciones. Realizado éste, si IS y LCL siguen sin aumentar significativamente, cabrá plantearse una segmentación diferente y testar de nuevo los resultados. Los nuevos criterios de segmentación deben estar inspirados en los intereses de una investigación sobre L2.