

La complejidad lingüística en los niveles de competencia del MCER: el caso de la variedad verbal en la expresión escrita en ELE

Linguistic complexity across CEFR proficiency levels: the case of variety of verb forms in written spanish

Nuria de la Torre García

Instituto Cervantes de Estocolmo.

nuria.delatorre@cervantes.es

de la Torre García, N. (2016). La complejidad lingüística en los niveles de competencia del MCER: el caso de la variedad verbal en la expresión escrita en ELE. *Revista Nebrija de Lingüística Aplicada* (2016) 20.

RESUMEN

Este trabajo es una contribución a la descripción empírica de los niveles de competencia definidos por el *Marco común europeo de referencia para las lenguas* (MCER) y certificados por el examen multinivel de español académico de la Universidad Nebrija. Los datos del estudio consistieron en 124 textos de respuesta a dos tareas de la prueba de expresión e interacción escrita de una versión experimental del examen mencionado. Los textos fueron clasificados por expertos en los niveles A-C1 del MCER utilizando una escala holística y una escala de alcance y control gramatical. En ese contexto se examinó la relación entre los niveles asignados y mediciones de variedad verbal. Los resultados mostraron que hay correlaciones débiles ($p < 0,05$) entre las calificaciones asignadas por los evaluadores y las mediciones objetivas de variedad verbal. Además, se encontró que algunos rasgos lingüísticos analizados fueron capaces de discriminar de forma limitada entre los niveles del MCER ($p < 0,05$).

Palabras clave: complejidad sintáctica, variedad verbal, evaluación de la expresión escrita, Marco común europeo de referencia para las lenguas

ABSTRACT

This study is a contribution to the empirical description of the proficiency levels defined by the Common European Framework of Reference for Languages (CEFR) and certified by the multilevel exam of Spanish for academic purposes of Nebrija University. The data consisted of 124 texts written in response to two tasks of the writing paper of an experimental form of the above mentioned exam. Expert raters using two descriptor scales, one holistic and one focusing on grammatical range and accuracy, classified the texts in the A-C1 CEFR levels. The study examined the relationship between the assigned CEFR levels and measures of variety of verb forms. The results showed that there were weak correlations ($p < 0,05$) between assigned CEFR levels as assessed by the raters using the two descriptor scales measures of syntactic variety. Furthermore, it was found that some of the analyzed linguistic features could distinguish across the CEFR levels ($p < 0,05$).

Keywords: syntactic complexity, variety of verb forms, writing assessment Common European Framework of Reference for Languages

Fecha de recepción: 10 octubre de 2015

Fecha de aceptación: 27 noviembre de 2015

1. INTRODUCCIÓN

En el año 2001 el Consejo de Europa publica el *Marco común europeo de referencia para las lenguas: aprendizaje, enseñanza, evaluación* (MCER), el documento que probablemente más ha influido sobre la enseñanza y evaluación de las lenguas extranjeras en Europa en lo que llevamos de siglo. El MCER se concibió como un documento abierto que propone interrogantes e invita a seguir investigando y decidiendo sobre muchos aspectos que un trabajo de alcance tan amplio deja necesariamente sin definir (Figueras, 2008). Uno de ellos es el de los correlatos lingüísticos con los que se sustentan en cada lengua en particular los niveles funcionales descritos en el MCER de forma general y para todas las lenguas en las que

“No se considera posible elaborar una escala de la progresión relativa a la estructura gramatical que sea aplicable a todas las lenguas” (Consejo de Europa, 2002:111). Queda, por tanto, en manos de los usuarios del MCER, determinar “qué elementos, categorías, clases, estructuras, procesos y relaciones gramaticales tendrá que dominar el alumno” (Consejo de Europa, 2002:111). Ese trabajo se puede hacer de forma apriorística dejando en manos de expertos la decisión de qué rasgos lingüísticos deben manejar los aprendices en cada nivel o se pueden discernir los rasgos propios de cada nivel a partir del análisis de muestras de lengua previamente clasificadas en los niveles funcionales del MCER. El estudio del que damos cuenta parcialmente en este artículo se encuadra en esa última línea de trabajo y se justifica porque, quince años después de la publicación del MCER y, a pesar de que ya contamos con estudios empíricos sobre bastantes lenguas europeas que responden a esa problemática, el español está prácticamente ausente de ese cuerpo de investigación que intenta describir, a partir de datos empíricos, los rasgos formales que se corresponden y singularizan los distintos niveles del MCER.

El objetivo principal del estudio aludido y del que nace la redacción de este artículo era analizar rasgos de complejidad gramatical en la actuación escrita de aprendices de español y comprobar si esos rasgos se correlacionaban y eran capaces de discriminar entre los diferentes niveles del MCER. Para ello se analizaron textos evaluados previamente por expertos en el marco del pilotaje de un examen multinivel de español académico de la Universidad Nebrija vinculado al MCER. En este artículo informamos sobre el comportamiento de algunas de las variables de complejidad gramatical utilizadas en ese trabajo.

1.1 LA COMPLEJIDAD LINGÜÍSTICA: DEFINICIÓN Y MEDICIÓN

En el ámbito de las segundas lenguas (L2) la complejidad lingüística se define de forma general como la habilidad que muestran los aprendices para usar un repertorio amplio y variado de estructuras y vocabulario de la lengua meta2. Algunos investigadores se refieren también en sus definiciones al carácter más o menos sofisticado del léxico y las construcciones que aparecen en la actuación de los usuarios de la lengua (Ortega, 2015; Housen, Kuiken y Vedder, 2012).

La definición del concepto de complejidad no está exenta de polémica y Bulté y Housen (2012) señalan que los trabajos empíricos en el campo de la complejidad producen en ocasiones resultados inconsistentes porque el constructo se ha venido definiendo de forma vaga y circular, lo que ha dificultado su operacionalización coherente. También Norris y Ortega (2009) advierten de los problemas de interpretación de los resultados obtenidos en los estudios empíricos sobre la complejidad, debido, en su opinión, a problemas de redundancia, reduccionismo e incongruencia al tratar la noción de complejidad en muchos trabajos de investigación. Tras una profunda revisión de las variables con las que se había operacionalizado y medido la complejidad sintáctica hasta el momento de la redacción de su artículo, encuentran que, por lo general, estaban basadas en la subordinación como principal y a veces única fuente de complejización. Frente a ese protagonismo casi absoluto de la subordinación, Norris y Ortega (2009) abogan por un modo más orgánico de medir un concepto multidimensional como la complejidad y proponen la medición de al menos cinco subconstructos: la complejidad global, la complejidad vía subordinación, la complejidad subclausular o sintagmática, la complejidad vía coordinación y, por último, la complejidad basada en la variedad, sofisticación y ritmos de adquisición de las formas gramaticales que aparecen en la producción. En este artículo informamos sobre variables relacionadas principalmente con ese último subconstructo, por lo que nos detendremos brevemente en él.

Según Norris y Ortega (2009) las variables que miden ese subconstructo son los índices de desarrollo que se utilizan mucho en la investigación en L1 y menos en la de L2. Esos índices suelen consistir en la medición de frecuencias de un conjunto ponderado de estructuras y construcciones que se seleccionan a partir de datos empíricos de desarrollo y/o jerarquías de dificultad basadas en producción o en juicios de gramaticalidad. En la investigación de L2 es más común utilizar indicadores menos sofisticados como son el conteo de determinada morfología verbal, clases de verbos, tipos de subordinadas, etc. Aunque son las variables con las que menos se ha trabajado, tanto Bulté y Housen (2012) como Norris y Ortega (2009) las consideran muy prometedoras y, de hecho, como se verá más adelante, se están utilizando con resultados interesantes en el ámbito europeo para la validación empírica de los niveles del MCER.

En el caso concreto de este artículo informaremos sobre el comportamiento de una variable que hemos denominado variedad verbal (VV) y que consiste en el cómputo del número total de paradigmas verbales diferentes que aparecen en los textos analizados. En el marco de la investigación de la enseñanza de lenguas extranjeras mediante tareas, Yuan y Ellis (2003) y Ellis y Yuan (2004) incluyen ese tipo de variable dentro de su conjunto de variables para medir el grado de complejidad de textos generados bajo condiciones de planificación diferentes. Según esos autores, y en consonancia con Norris y Ortega (2009), esa variable representaría una de las dimensiones de la complejidad gramatical, en concreto la de la variedad sintáctica.

1.2 LOS ESTUDIOS DE COMPLEJIDAD GRAMATICAL Y LOS NIVELES DEL MCER

De acuerdo con Ortega (2012), las mediciones de complejidad lingüística se han utilizado en la investigación para describir la actuación de los aprendices, medir su grado de dominio de la lengua meta y establecer puntos de referencia en su desarrollo lingüístico. En el fondo de esa práctica late la idea de que la habilidad de producir textos orales o escritos más complejos desde el punto de vista lingüístico refleja la maduración de la interlengua, razón por la que las mediciones de complejidad podrían proporcionar indicadores objetivos y prácticos de los niveles de dominio y desarrollo.

Sin embargo, la literatura también muestra acuerdo general en que las mediciones de complejidad, por exhaustivas que puedan llegar a ser, resultan claramente insuficientes para dar cuenta de constructos tan amplios como el dominio o el desarrollo lingüístico. De hecho, en la mayoría de las investigaciones, las mediciones se extienden a las otras dos dimensiones clásicas de la actuación lingüística: la corrección y la fluidez.

Por otro lado, el crecimiento lineal de la complejidad a medida que se desarrolla el sistema interlingüístico y se avanza en los niveles de dominio tampoco se da por garantizado en la bibliografía. Se considera que hay muchos factores que modulan ese posible crecimiento y sus formas de manifestarse en la producción, desde cuestiones relacionadas con las características de los aprendices (L1, nivel relativo de dominio, contexto de aprendizaje) hasta los mismos textos sobre los que se realizan los análisis (requerimientos funcionales y cognitivos de las tareas de elicitación, modalidad, género), pasando por las características propias de la naturaleza del desarrollo lingüístico, que es un proceso altamente individual (Larsen-Freeman, 2009; Pallotti, 2009; Ortega, 2003, 2015).

Siendo así, tampoco parece haber muchas dudas de que el constructo de complejidad, interpretado correctamente, puede dar cuenta de aspectos importantes de la actuación lingüística aunque no constituya por sí mismo un indicador directo del dominio y del desarrollo lingüístico subyacente. Por esa razón, el cuerpo de investigación al que aludíamos en la Introducción de este artículo utiliza en muchas ocasiones mediciones de complejidad gramatical para investigar los perfiles lingüísticos que caracterizan los niveles funcionales del MCER. A continuación damos cuenta de algunos de esos estudios³.

Forsberg y Bartning (2010) analizan 83 muestras escritas por 42 estudiantes universitarios suecos. Los textos clasificados mediante escalas basadas en descriptores del MCER se sometieron al análisis y medición de indicadores relacionados con rasgos morfosintácticos, discursivos y léxicos. Los rasgos morfosintácticos analizados, específicos del francés, se seleccionaron a priori a partir de fenómenos lingüísticos que ya se habían demostrado cruciales en estudios empíricos sobre los órdenes de adquisición de esa lengua extranjera. Se analiza la emergencia de estructuras y la frecuencia de error en aspectos como la morfología verbal, la nominal, la estructura de los enunciados, la concordancia, etc. El cálculo de errores en esos fenómenos señala diferencias significativas entre las producciones de los niveles A2, B1 y B2 (no se hizo estadística inferencial con los niveles A1 y C1). También es interesante observar un desarrollo lineal a lo largo del espectro de niveles del MCER de ciertos rasgos discursivos y gramaticales, como algunos conectores y conjunciones subordinadas, el pluscuamperfecto y el gerundio. La conclusión es que hay algún tipo de relación entre el desarrollo comunicativo que miden las escalas y el desarrollo de la interlengua que miden los cómputos en torno a la emergencia y corrección de determinados fenómenos lingüísticos.

Muy diferente, sin embargo, es el panorama que presentan Prodeau, Lopez y Véronique (2012). En este caso se busca establecer correlatos lingüísticos entre los estadios de interlengua que analiza el programa informático *Direkt Profil* (un analizador de la interlengua francesa basado en los mismos estadios de desarrollo que sirvieron de base para el trabajo de Forsberg y Bartning, 2010). Se estudiaron textos producidos por 40 participantes con diferentes L1, clasificados previamente mediante un examen de francés anclado en el MCER (*Test de Connaissance du Français*) en el nivel B1 (n=20) y en el nivel B2 (n=20). Con todos los textos se procedió de la misma forma: primero se aislaron los recursos gramaticales que emergían en ambos niveles y a continuación se analizaron desde el punto de vista de la corrección en su uso. Ni utilizando el criterio de emergencia ni el de corrección se encontraron diferencias claras entre los textos producidos por participantes de uno y otro nivel, por lo que fue imposible hallar un perfil gramatical que discriminara entre la expresión escrita de aprendices de los niveles B1 y B2. Aunque basados en el mismo marco teórico, los estudios de Forsberg y Bartning (2010) y Prodeau, Lopez y Véronique (2012) tienen diferencias metodológicas importantes que pueden explicar la diferencia de resultados.

Martin et al. (2010) hacen un estudio exploratorio sobre el finés como lengua segunda. Se analizan 669 muestras escritas y clasificadas en los seis niveles del MCER mediante escalas funcionales, sin alusiones explícitas a las formas lingüísticas. Este trabajo presenta un interesante y original modelo de análisis que permite rastrear el desarrollo de determinadas estructuras desde su emergencia hasta su dominio, midiendo parámetros relacionados con su frecuencia, distribución y corrección. Ponemos como ejemplo de su análisis el desarrollo de los casos locativos en finés. Advierten que las expresiones

espaciales concretas y estáticas se dominan en el nivel B1 como muy tarde, mientras que los usos metafóricos de esas mismas expresiones se desarrollan mucho más lentamente. También documentan interacciones interesantes entre la emergencia, la frecuencia, la complejidad y la corrección en el uso de recursos gramaticales complejos: un mismo rasgo puede emerger ya en el nivel A1, su frecuencia de uso puede no ser significativa hasta el nivel B1 y un salto significativo en términos de corrección puede no producirse hasta más tarde, por ejemplo en el nivel B2. Esto significaría que cuestiones como la emergencia, la frecuencia de uso y la corrección de una misma estructura pueden ser indicadores significativos y capaces de discriminar entre los niveles de dominio descritos en el MCER.

Un poco en la misma línea de este último estudio se halla el de Toropainen y Lahtinen (2014), que trabajan con el finés y el sueco como lenguas meta. El objetivo del estudio es rastrear, mediante la frecuencia de uso de tres tipos de cláusulas interrogativas (totales, parciales e indirectas), el desarrollo lingüístico a lo largo de los niveles del MCER. Los participantes en el estudio son escolares del sistema educativo finlandés. Se recogieron 329 muestras escritas, 208 en finés y 121 en sueco, que se clasificaron en los niveles A1-B2 por medio de juicios de profesionales que utilizaron escalas derivadas del MCER. Los resultados sugieren que la complejidad crece a medida que avanza el nivel en ambas lenguas y que hay interacción entre las estructuras gramaticales y sus funciones pragmáticas cuando aumentan los niveles de dominio. Este trabajo también resulta interesante porque es un ejemplo de cómo se puede estudiar un fenómeno lingüístico concreto en dos lenguas tipológicamente muy diferentes.

El trabajo de Kuiken, Vedder y Gilbert (2010) sobre el holandés, el italiano y el español, a diferencia de los estudios que acabamos de ver, utiliza mediciones de complejidad sintáctica basadas en la densidad de subordinación. Los datos analizados proceden de 206 textos breves producidos por 34 aprendices universitarios de holandés, 42 de italiano y 27 de español. Los textos se evaluaron utilizando dos escalas: una juzgaba la adecuación comunicativa y la otra, la complejidad lingüística. Ese fue el procedimiento que se utilizó para clasificar las muestras en cinco niveles del MCER, de A2 a C1. Los investigadores encuentran correlaciones con sus variables de diversidad léxica y corrección pero no, en ningún caso y en ninguna de las tres lenguas estudiadas, con las variables de complejidad sintáctica elegidas para este estudio.

Salamoura y Saville (2010) describen el programa *English Profile*, un proyecto interdisciplinar que busca identificar rasgos lingüísticos capaces de discriminar la actuación de usuarios de los niveles del MCER a la luz de modelos psicolingüísticos de adquisición y utilizando avanzadas herramientas de la lingüística de corpus y computacional. El objetivo final de todo el proyecto es desarrollar descripciones lingüísticas de los niveles de referencia basadas en los datos empíricos de los aprendices que se recogen en el *Cambridge Learner Corpus*. Un hallazgo interesante de ese proyecto es el papel que tiene la frecuencia de un determinado rasgo en la lengua meta para determinar su orden de adquisición. Se observa que la aparición de ciertas estructuras en la interlengua está marcada por su frecuencia en el habla nativa y por el nivel que haya alcanzado el aprendiz. Por ejemplo, el patrón sintáctico SN-V-SN (*He loved her*) ya aparece en el A2, pero el que se corresponde con una oración como *They admitted to the authorities that they had entered illegally* (SN-V-SP-Clausula subordinada) no se advierte hasta B1. Otro factor que, según estos investigadores, parece influir en el orden de adquisición es lo que denominan complejidad semántica y estructural. Investigan si las jerarquías de complejidad establecidas previamente en la lingüística del inglés para algunas estructuras se corresponden con el uso que hacen de ellas los aprendices en el corpus. Por ejemplo, estudian el orden de adquisición de las oraciones de relativo, desde las que se consideran a priori más simples (relativos con función de sujeto) hasta las más complejas (relativos con función de genitivo). Y encuentran que existe correlación entre las jerarquías de complejidad establecidas previamente para esas estructuras y los niveles del MCER ya que, por ejemplo, las relativas más complejas no aparecen en el corpus hasta el nivel C1. Lo mismo ocurre respecto a cuestiones semánticas, documentándose, por ejemplo, usos figurados más tarde que usos concretos de ciertos fenómenos lingüísticos, tal y como también detectan Martin et al. (2010) para el finés. Por último, hay otro principio que estos autores piensan que conforma el desarrollo lingüístico: se trata de la transferencia de la L1. Según lo que observan en su corpus, los elementos que ya están en la L1, si no están afectados por las propiedades anteriores (baja frecuencia y/o alta complejidad) se adquirirán antes y con menos esfuerzo. Así parece ocurrir con el uso del artículo, donde las tasas de error en la interlengua de aprendices cuya L1 tiene artículos son inferiores, en todos los niveles, a las de los hablantes de lenguas sin artículo. Por ejemplo, la tasa de error por ausencia del artículo indeterminado en los hispanohablantes de nivel A2 es 4.52 y en el C2, de 3.58. En el caso de los hablantes de turco, una lengua sin artículo, la tasa de ese mismo error en el A2 es de 24.29 y en el C2 de 11.86. Es evidente la progresión en ambos casos hacia las formas canónicas de la lengua meta pero también es obvia la diferencia entre las dos interlenguas.

Gyllstad et al. (2014) recogen datos en el sistema educativo y universitario sueco de 120 participantes: 54 aprendices de inglés como L2, 38 de francés como L3 y 28 de italiano como L4. Los participantes escribieron dos textos que fueron evaluados por expertos utilizando una escala funcional derivada del MCER. Los niveles alcanzados en todos los casos fueron del A1 al B2. A continuación los textos se codificaron y analizaron en torno a tres subconstructos de la complejidad sintáctica:

la global, la sintagmática y la densidad de subordinación. Encontraron correlaciones moderadas entre las mediciones y los niveles asignados por los evaluadores. Sus resultados parecen indicar también que el nivel A es homogéneo en las tres lenguas estudiadas pero no así el nivel B en el que, según estos investigadores, surge verdaderamente la complejidad sintáctica.

Finalmente, el trabajo de Kang (2013) analiza un total de 120 muestras orales obtenidas en los exámenes de Cambridge anclados en cuatro niveles del MCER: 32 de B1 (*PET*), 32 de B2 (*FCT*), 34 de C1 (*CAE*) y 22 de C2 (*CPE*). El estudio se enmarca en un proceso de validación de las pruebas orales de esos exámenes con el objetivo de explorar los rasgos lingüísticos más prominentes y capaces de distinguir unos niveles de otros, niveles que se establecen con escalas que contienen criterios basados en el manejo del discurso, los recursos léxicos y gramaticales y la pronunciación. Para ello se analiza el primer minuto de respuesta a la tarea de monólogo sostenido de las pruebas. En general, según los resultados, parece evidente que el uso de construcciones complejas aumenta en los niveles más altos, encontrándose también interacciones interesantes entre complejidad y corrección precisamente en esos niveles. Algunos fenómenos relacionados con la complejidad sintáctica prueban su capacidad de discriminar incluso entre niveles adyacentes.

Como vemos, el estado de la cuestión es complejo ya que encontramos metodologías muy variadas y resultados contradictorios. Algunos de los estudios utilizan medidas generales de complejidad sintáctica, básicamente las mediciones clásicas de longitud oracional y densidad de subordinación (Kuiken, Vedder y Gilabert 2010; Gyllstad et al., 2014), mientras que la mayoría miden también, o exclusivamente, la emergencia y corrección de construcciones específicas de las lenguas meta estudiadas (Forsberg y Bartning, 2010; Prodeau, Lopez y Véronique, 2012; Martin et al., 2010; Toropainen y Lahtinen, 2014; Salamoura y Saville, 2010). Los resultados son divergentes: en la mayoría de las ocasiones, las mediciones del desarrollo de la complejidad gramatical son significativas y discriminan entre niveles de desarrollo (Forsberg y Bartning, 2010; Salamoura y Saville, 2010; Gyllstad et al., 2014) pero también hay casos en que no lo son (Prodeau, Lopez y Véronique, 2012; Kuiken, Vedder, y Gilabert, 2010).

En general, se observa que parece haber variables intervinientes en todos esos resultados: las lenguas maternas de los participantes, los contextos de instrucción y la propia selección de variables ya que, probablemente, esté aun por decidir cuáles son las que mejor representan el constructo de la complejidad sintáctica o gramatical. Parece evidente que la multidimensionalidad del propio constructo de complejidad aconseja operacionalizarlo con variables que intenten capturar la mayor cantidad posible de subconstructos (véase, por ejemplo, la falta de resultados de Kuiken, Vedder, y Gilabert, 2010, que solo utilizan indicadores de subordinación). Aunque una buena combinación de variables generales tiene resultados bastante satisfactorios (véase Gyllstad et al., 2014), parece que las más prometedoras son las más específicas y dependientes de la lengua meta. Otra cuestión interesante respecto a esto es el punto de vista desde el que se computan esas variables: a veces la complejidad gramatical se valora desde la mera emergencia y en otras ocasiones parece que es determinante tener en cuenta la frecuencia y el grado de corrección con que se usan determinadas estructuras.

Por último, siendo lo ideal encontrar indicadores que progresen linealmente y que sean capaces de discriminar estadísticamente entre niveles, muchos de los estudios revisados descartan la idea de que las lenguas segundas se adquieran de forma lineal y homogénea por sus aprendices. Las diferencias individuales, el dinamismo de los procesos de cambio y desarrollo y la propia naturaleza compleja de los sistemas lingüísticos dificultan, si no contradicen, la búsqueda de tales indicadores.

Este complicado panorama no impide, de todas maneras, que la mayoría de las investigaciones tengan resultados relativamente satisfactorios y muestren que los indicadores de complejidad gramatical pueden dar cuenta, aunque sea parcialmente, del avance a lo largo de los niveles de competencia comunicativa que propone el MCER.

1.3 PREGUNTAS DE INVESTIGACIÓN

El principal objetivo de este trabajo es estudiar el comportamiento de la medición objetiva de variedad y frecuencia de aparición de paradigmas verbales a lo largo del espectro de niveles del MCER certificados por el examen multinivel de español académico de la Universidad Nebrija (A, B1, B2 y C1).

En consonancia con ello, nos formulamos las siguientes preguntas de investigación:

1. Las mediciones objetivas de variedad y frecuencia de aparición de paradigmas verbales, ¿correlacionan de forma significativa con las calificaciones de los textos obtenidas a partir de escalas de evaluación?
2. Respecto a las mediciones objetivas de variedad y frecuencia de aparición de paradigmas verbales, ¿discriminan de forma significativa entre los diferentes niveles asignados mediante los juicios de los calificadores basados en escalas?

2. MÉTODO

2.1 PARTICIPANTES Y FORMA DE RECOGIDA DE LOS DATOS

Los datos que se analizan en este trabajo proceden de 124 textos escritos por 62 estudiantes de la Universidad Nebrija de Madrid⁴. Se trata de hablantes no nativos de español con diferentes lenguas maternas que, en el momento de la recogida de datos, estudiaban español en cursos de cuatro niveles diferentes: A2 (n=12), B1 (n=24), B2 (n=22) y C1 (n=5). Lamentablemente, no disponemos de más información sobre los participantes aunque, por la lectura de los textos, podemos inferir que se trata de personas jóvenes de procedencia tan diversa como EEUU, China, Alemania o Inglaterra que realizaban estudios en Madrid temporalmente.

Los textos se recogieron en el marco del pilotaje de un examen multinivel de certificación de español con fines académicos, administrado por la Universidad Nebrija. El objetivo del examen era determinar si los candidatos demostraban con su actuación un nivel de dominio lingüístico B1, B2 o C1 de español, según el MCER, en el ámbito universitario/académico. Los textos que analizamos en este trabajo proceden de la prueba de expresión e interacción escrita del examen, que tenía una duración estimada de 60 minutos. Los objetivos específicos y las escalas de calificación de la prueba se determinaron a partir de los descriptores del MCER relacionados con el uso de la lengua en el ámbito académico en los niveles B1, B2 y C1.

Cada uno de los participantes respondió de forma manuscrita a las dos tareas que constituían la prueba. La tarea 1 (T1) consistía en escribir un correo electrónico a un profesor. Las instrucciones de la tarea planteaban una situación en la que el estudiante estaba en desacuerdo con una nota recibida, por lo que debía expresar su opinión y formular al profesor la solicitud que considerase oportuna. En la tarea 2 (T2) se trataba de describir y explicar una tabla con datos estadísticos sobre el emplazamiento segregado de alumnos con necesidades especiales en distintos países de Europa. A continuación el estudiante debía desarrollar un texto argumentando su postura personal sobre el tema.

Aunque ambas tareas están dentro del ámbito académico y requieren, más claramente en el caso de la T2, un registro formal, también presentan diferencias significativas. La más obvia es la de la extensión de los textos: las instrucciones de la T1 requerían un mínimo de 125 palabras y las de la T2 un mínimo de 200 y un máximo de 250. Además, la T1 es de carácter transaccional y consiste en redactar un texto epistolar, con cierto grado de formalidad, en el que se formulen de manera adecuada una opinión y una petición que entraña cierto conflicto pragmático que deberá solventar el estudiante. Por su lado, la T2 tiene una primera fase de transferencia de información que da lugar a un texto expositivo que, a su vez, deberá ir seguido de un texto argumentativo. Hay, por tanto, entre ambas tareas diferencias evidentes desde el punto de vista de la tipología textual y de sus requerimientos funcionales y pragmáticos.

Todos los participantes realizaron las tareas descritas anteriormente y cada texto de respuesta fue evaluado independientemente por dos calificadores que utilizaron dos escalas de evaluación diferentes⁵. Uno de ellos utilizó una escala de calificación holística y el otro una escala de calificación analítica. En ambos casos se trataba de escalas tipo Likert de cuatro bandas que, basándose en las descripciones de niveles del MCER en el ámbito académico, incluían descriptores de calificación multinivel. La banda 1 describía la actuación escrita de participantes que no alcanzaban el nivel B1 del MCER. La banda 2 se destinaba a describir la producción de nivel B1, la banda 3 se correspondía con el nivel B2 y la banda 4 con el nivel C1. Utilizando esas escalas, nuestros textos quedaron clasificados, a juicio de los calificadores, en los tres niveles del MCER mencionados o por debajo de todos ellos, es decir, en el nivel A. Para nuestro análisis se utilizó la clasificación de los textos derivada de la aplicación de la escala holística así como la derivada de la escala de alcance y control gramatical, el componente de la escala analítica más cercano al constructo de complejidad sintáctica.

En la tabla I mostramos la distribución resultante del proceso de evaluación de los textos analizados según las escalas, niveles y tareas que se investigan en el trabajo.

	Escalas	Nivel A	Nivel B1	Nivel B2	Nivel C1	n=124
Tarea 1	Holística	10	25	22	5	n=62
	Alcance y control gramatical	20	30	11	1	
Tarea 2	Holística	12	24	25	1	n=62
	Alcance y control gramatical	15	30	16	1	

Tabla I: Distribución de los textos del corpus por niveles, tareas y escalas

2.2 VARIABLES

Como ya hemos señalado y siguiendo a Norris y Ortega (2009), nuestra variable de variedad verbal (VV) pretende representar una de las dimensiones del constructo de complejidad gramatical, la de la variedad y sofisticación de las formas usadas en los textos. Nosotros restringimos el rastreo de esos fenómenos al análisis de las formas verbales ya que la adquisición del tiempo, el modo y el aspecto presenta en español idiosincrasias suficientemente contrastadas. A ello hay que añadir que la escala de calificación de alcance y control gramatical hace mención específica y bastante detallada de las formas verbales que se espera ver aparecer en los diferentes niveles. Obtenemos el valor de esta variable contando el número total de paradigmas verbales diferentes que aparecen en cada texto.

Además, analizamos la frecuencia de aparición en los textos de 14 paradigmas: presente de indicativo y subjuntivo, pretérito perfecto compuesto de indicativo y subjuntivo, pretérito perfecto simple de indicativo, pretérito imperfecto de indicativo y subjuntivo, pretérito pluscuamperfecto de indicativo y subjuntivo, futuro simple y compuesto de indicativo, condicional simple y compuesto e imperativo. Se incluyen también dos formas impersonales del verbo: el gerundio y el infinitivo. La frecuencia de las 16 formas verbales rastreadas se computa dividiendo el número de apariciones de cada una de ellas por el número total de verbos de cada texto. Se obtiene así el porcentaje con el que cada forma contribuye a la variedad verbal de las muestras.

2.3 PROCEDIMIENTOS DE ANÁLISIS

Los textos se transcribieron en el formato CHAT (*Human Analysis of Transcripts*) para ser luego analizados con CLAN (*Computerized Language Analysis*). Ambos programas forman parte del proyecto *Child Language Data Exchange System* o CHILDES (MacWhinney, 2000). La codificación y etiquetado de los elementos correspondientes a la variable de variedad verbal se realizó manualmente con la ayuda del programa *Machinese Phrase Tagger*, un analizador-etiquetador producido por el grupo finlandés *Connexor*.

Por último, los análisis estadísticos se generaron en *Microsoft Excel* mejorado con la extensión *Real Statistics Resource Pack* (*Release 4.3 Copyright 2013 – 2015 Charles Zaiontz*), un complemento que proporciona herramientas avanzadas que permiten llevar a cabo fácilmente los procedimientos de análisis paramétricos y no paramétricos más habituales en las investigaciones de lingüística aplicada.

3. RESULTADOS

3.1. PREGUNTA DE INVESTIGACIÓN 1

La primera pregunta de investigación cuestiona las relaciones entre las variables de medición de la variedad verbal (VV) y las puntuaciones que se derivan de la aplicación de las escalas por parte de los examinadores.

Para investigar las correlaciones en este trabajo se utiliza el coeficiente *tau* de Kendall, ya que se considera adecuado para conjuntos pequeños de datos, variables ordinales y gran número de rangos empatados como ocurre con nuestros datos. Se informa, además, del coeficiente de determinación (Coef. Det. en las distintas tablas), que es el coeficiente de correlación elevado al cuadrado y que da cuenta del grado de solapamiento de las variables (Herrera, Martínez y Amengual, 2011).

En la tabla II se muestran las correlaciones obtenidas entre las mediciones y las puntuaciones de los textos.

VV	Tarea 1 (n=62)				Tarea 2 (n=62)			
	Escala holística		Escala de alcance y control gramatical		Escala holística		Escala de alcance y control gramatical	
	Tau de Kendall	Coef Det	Tau de Kendall	Coef Det	Tau de Kendall	Coef Det	Tau de Kendall	Coef Det
	0,279**	0,078	0,134	0,018	0,301**	0,091	0,257*	0,066

**= $p < 0,01$ *= $p < 0,05$

Tabla II: Correlaciones y coeficientes de determinación obtenidos entre las mediciones de variedad verbal y las calificaciones de los textos.

Como se puede observar, la variedad verbal se correlaciona positivamente con las puntuaciones, excepto con las que se obtienen en los textos de la tarea 1 a partir de la escala de alcance y control gramatical, donde el coeficiente no es estadísticamente significativo. Las correlaciones son débiles y los coeficientes de determinación indican que la variable explica pequeños porcentajes del rango de calificaciones. No es sorprendente este resultado que refleja el comportamiento de una sola variable, ya que es lógico pensar que sea necesaria una combinación de rasgos para explicar mayores porcentajes de variación. Sin embargo, las correlaciones encontradas indican que la variedad verbal crece a medida que los evaluadores clasifican los textos en niveles más altos y parece, por tanto, que la riqueza y diversidad de paradigmas verbales tiene algún papel a la hora de juzgar la calidad de los textos.

Tarea 1						
	Escala holística			Escala gramatical		
	A/B1	B1/B2	A/B2	A/B1	B1/B2	A/B2
Nº de paradigmas verbales diferentes						✓
Pretérito perfecto compuesto de indicativo						✓
Pretérito perfecto simple						✓
Condicional simple				✓		
Tarea 2						
	Escala holística			Escala gramatical		
	A/B1	B1/B2	A/B2	A/B1	B1/B2	A/B2
Nº de paradigmas verbales diferentes		✓				✓
Presente de indicativo						✓
Presente de subjuntivo						✓

Tabla III: Tiempos verbales que aparecen en los textos de la tarea 1

Como ya se ha indicado, además de la variedad verbal en general, teníamos interés en observar más de cerca el comportamiento en nuestros textos de los diferentes paradigmas verbales analizando su frecuencia de aparición. Como era de esperar, las frecuencias difieren mucho de unas formas a otras. Dos de los paradigmas rastreados no aparecen en ningún texto de la tarea 1: el condicional compuesto y el futuro compuesto. Esta última forma, junto con el imperativo y el pretérito pluscuamperfecto de indicativo, tampoco aparece en ningún texto de la tarea 2. En cuanto a los tiempos verbales que sí aparecen en los textos, hemos considerado tres grupos:

- Tiempos verbales de frecuencia baja: se registran en menos del 30% de los textos y/o tienen menos de 30 apariciones.
- Tiempos verbales de frecuencia media: se registran en el 50% - 75% de los textos y/o tienen entre 30 y 100 apariciones.
- Tiempos verbales de frecuencia alta: se registran en todos los textos y/o tienen más de 100 apariciones.

Las tablas III y IV muestran respectivamente las formas que aparecen en los textos de las tareas 1 y 2.

TAREA 2			
	Formas verbales	Nº de apariciones	Nº de textos con presencia
Frecuencia baja	Condicional compuesto	1	1
	Pretérito pluscuamperfecto de subjuntivo	1	1
	Pretérito perfecto compuesto de subjuntivo	2	2
	Pretérito imperfecto de subjuntivo	8	8
	Pretérito perfecto compuesto de indicativo	20	13
	Futuro simple	20	13
	Pretérito perfecto simple	23	13
Frecuencia media	Gerundio	20	15
	Condicional simple	60	17
	Pretérito imperfecto de indicativo	33	18
Frecuencia alta	Presente de subjuntivo	89	38
	Infinitivo	630	62
	Presente de indicativo	1.257	62

Tabla IV: Tiempos verbales que aparecen en los textos de la tarea 2

Se observa la relativa baja incidencia de los tiempos compuestos que, a excepción del pretérito perfecto de indicativo, no aparecen en el corpus o aparecen con muy baja frecuencia. Recordemos que, en principio, las formas verbales compuestas son inherentemente más complejas que las simples (Bulté y Housen, 2012).

TAREA 1 (n=62)				
ESCALAS	Pretérito perfecto simple		Pretérito perfecto compuesto de indicativo	
	Tau de Kendall	Coef. Det.	Tau de Kendall	Coef. Det.
Holística	-0,109	0,012	0,324**	0,105
Gramatical	-0,247*	0,061	0,285**	0,081
ESCALAS	Condicional simple			
	Tau de Kendall	Coef. Det.		
Holística	0,240*	0,058		
Gramatical	0,230*	0,053		
TAREA 2 (n=62)				
ESCALAS	Condicional simple		Pretérito perfecto compuesto de indicativo	
	Tau de Kendall	Coef. Det.	Tau de Kendall	Coef. Det.
Holística	0,229*	0,052	0,337**	0,114
Gramatical	0,122	0,015	0,213	0,046
ESCALAS	Presente de subjuntivo		Presente de indicativo	
	Tau de Kendall	Coef. Det.	Tau de Kendall	Coef. Det.
Holística	0,175	0,031	-0,220*	0,048
Gramatical	0,282**	0,080	-0,226*	0,051

**= $p < 0,01$ *= $p < 0,05$

Tabla V: Correlaciones y coeficientes de determinación obtenidos entre las mediciones de frecuencia de aparición de paradigmas verbales y las calificaciones de los textos.

De todas las formas analizadas, solo las que aparecen en la tabla V presentan una frecuencia de aparición que correlaciona significativamente con las puntuaciones de los calificadores.

Se trata, de nuevo, de correlaciones muy débiles que, sin embargo, muestran significación estadística, aunque no siempre en ambas escalas de calificación. En los textos de ambas tareas la frecuencia de aparición del pretérito perfecto compuesto de indicativo y del condicional simple crece a medida que se avanza de nivel. Igual ocurre con el presente de subjuntivo en los textos de la tarea 2. Por el contrario, la frecuencia de uso del pretérito perfecto simple en los textos de la tarea 1 y del presente de indicativo en los de la tarea 2 decrece cuando mejora la calificación. Recordemos a propósito de este resultado, el trabajo de Salamoura y Saville (2010) donde se discute el papel que tiene la frecuencia de un determinado rasgo en la lengua meta para determinar su orden de adquisición en la interlengua. Observan que los rasgos más frecuentes son objeto de abuso por parte de los aprendices, que a medida que avanzan de nivel, los van sustituyendo por otras formas al diversificarse y consolidarse su gramática o su vocabulario. Esa pudiera ser la razón por la que en nuestros datos la

frecuencia de aparición del presente de indicativo disminuye a medida que crece el nivel.

En resumen y como respuesta a la pregunta de investigación 1, hemos observado correlaciones débiles pero estadísticamente significativas entre las variables de medición objetiva y las puntuaciones subjetivas de los examinadores basadas en escalas de calificación vinculadas al MCER. Esto indica que los juicios de los calificadores se corresponden, aunque sea parcialmente, con la dimensión del constructo de complejidad gramatical que nuestras mediciones de variedad y frecuencia verbal quieren representar.

3.2. PREGUNTA DE INVESTIGACIÓN 2

La tabla VI muestra la estadística descriptiva de los resultados del análisis de la variedad verbal en los textos clasificados en los niveles A, B1 y B2 tras la aplicación de las escalas holística y de alcance y control gramatical. El nivel C se excluye de este análisis debido a la escasez de textos que lo representan.

		TAREA 2		
		Formas verbales	Nº de apariciones	Nº de textos con presencia
Frecuencia baja	Condicional compuesto		1	1
	Pretérito pluscuamperfecto de subjuntivo		1	1
	Pretérito perfecto compuesto de subjuntivo		2	2
	Pretérito imperfecto de subjuntivo		8	8
	Pretérito perfecto compuesto de indicativo		20	13
	Futuro simple		20	13
	Pretérito perfecto simple		23	13
	Gerundio		20	15
Frecuencia media	Condicional simple		60	17
	Pretérito imperfecto de indicativo		33	18
	Presente de subjuntivo		89	38
Frecuencia alta	Infinitivo		630	62
	Presente de indicativo		1.257	62

Tabla VI: Promedio y desviación típica (DT) del número de paradigmas verbales diferentes en los textos clasificados por niveles.

Como documenta la tabla VI, el comportamiento de las mediciones de variedad verbal es diferente en las dos escalas que manejamos. Según la clasificación de la escala holística los promedios de variedad verbal no progresarían de forma lineal a medida que se avanza de nivel, mientras que en los textos clasificados por la escala de alcance y control gramatical sí lo harían.

Para saber si las diferencias entre las medias obtenidas tienen significación estadística hemos utilizado la prueba de Kruskal-Wallis, un test no paramétrico que se aplica en lugar de la prueba F de ANOVA cuando se violan los supuestos de normalidad y homogeneidad de varianzas. También es una prueba idónea cuando las muestras son pequeñas y los tamaños muestrales son desiguales, como es nuestro caso (Herrera, Martínez y Amengual, 2011). Los resultados de la prueba muestran que las diferencias tienen significación estadística, a excepción de las que arroja la escala holística en los textos de la tarea 1. Se pueden ver los resultados del test en la tabla VII.

	Tarea 1		Tarea 2	
	H	p	H	p
Escala holística	3,203	0,202	11,225	0,004
Escala gramatical	6,268	0,044	6,139	0,046

Tabla VII: Contraste de promedios de las mediciones de variedad verbal.

La prueba de Games Howell, un test post hoc adecuado para utilizarse con muestras que presentan tamaños desiguales o que no respetan los supuestos de los análisis de varianza paramétricos como ocurre con nuestras muestras (Zaiontz, 2015), permite averiguar si las mediciones que han resultado significativas son capaces de distinguir entre niveles adyacentes. Según este test la variable de variedad verbal distingue de forma significativa entre los niveles A y B2 ($p < 0,05$) en los textos de ambas tareas clasificados con la escala de alcance y control gramatical. Encontramos, además, un

contraste significativo entre niveles consecutivos: en los textos de la tarea 2 habría diferencias significativas entre los promedios de los textos de nivel B1 y B2 calificados mediante la escala holística ($p < 0,01$). No se encuentran en ningún caso diferencias significativas entre los niveles A y B1.

Acercando la lupa al comportamiento de los diferentes paradigmas verbales, mostramos en la tabla VIII los promedios y desviaciones típicas de la frecuencia de aparición de los tiempos de frecuencia media y alta y el tiempo verbal más frecuente de los de baja frecuencia, que en ambas tareas es el gerundio.

Como se observa en la tabla VIII, en los textos de la tarea 1 hay tiempos verbales que presentan progresión lineal ascendente: infinitivo (solo en la clasificación de la escala holística), pretérito perfecto compuesto de indicativo, condicional simple (solo en la clasificación de la escala holística), presente de subjuntivo y gerundio (estos dos últimos solo en la clasificación de la escala gramatical). La progresión lineal es descendente en el caso de los siguientes tiempos verbales: el presente de indicativo (solo en la clasificación de la escala holística), el pretérito imperfecto de indicativo y el pretérito perfecto simple (este último solo en la clasificación de la escala gramatical). En el caso de los textos de la tarea 2, se documenta progresión lineal en la aparición de cuatro de las seis formas analizadas: presente de indicativo, en sentido descendente, y presente de subjuntivo (solo en los niveles de la escala gramatical), infinitivo y condicional simple en sentido ascendente.

Cuando se aplica el contraste Kruskal-Wallis a los promedios obtenidos en los textos de la tarea 1, los datos muestran que, si consideramos la aplicación de las dos escalas, solo habría diferencias estadísticamente significativas en los valores promedio del pretérito perfecto compuesto de indicativo ($H=8,404$; $p=0,015$ en la escala holística y $H=9,451$; $p=0,009$ en la escala gramatical). También las habría, según la clasificación basada en la escala de alcance y control gramatical, en el uso del condicional simple ($H=8,542$; $p=0,014$) y el pretérito perfecto simple ($H=7,571$; $p=0,023$).

Tarea 1						
		Presente de indicativo	Infinitivo	Pretérito perfecto compuesto de indicativo	Pretérito perfecto simple	
		Promedio (DT)	Promedio (DT)	Promedio (DT)	Promedio (DT)	
Escala holística	Nivel MCER					
	A (n=10)	0,475 (0,14)	0,175 (0,09)	0,066 (0,08)	0,106 (0,14)	
	B1 (n=25)	0,437 (0,11)	0,191 (0,09)	0,070 (0,08)	0,109 (0,08)	
Escala gramatical	B2 (n=22)	0,420 (0,11)	0,212 (0,07)	0,118 (0,07)	0,061 (0,06)	
	A (n=20)	0,466 (0,11)	0,186 (0,08)	0,059 (0,07)	0,126 (0,10)	
	B1 (n=30)	0,417 (0,11)	0,207 (0,07)	0,097 (0,08)	0,084 (0,08)	
Escala holística	B2 (n=11)	0,425 (0,13)	0,170 (0,09)	0,133 (0,08)	0,041 (0,04)	
			Condicional simple	Presente de subjuntivo	Pretérito imperfecto de indicativo	Gerundio
			Promedio (DT)	Promedio (DT)	Promedio (DT)	Promedio (DT)
Escala holística	Nivel MCER					
	A (n=10)	0,016 (0,02)	0,048 (0,05)	0,071 (0,09)	0,004 (0,01)	
	B1 (n=25)	0,031 (0,03)	0,030 (0,04)	0,060 (0,08)	0,021 (0,04)	
Escala gramatical	B2 (n=22)	0,041 (0,03)	0,056 (0,05)	0,046 (0,05)	0,014 (0,02)	
	A (n=20)	0,016 (0,02)	0,032 (0,04)	0,068 (0,08)	0,010 (0,02)	
	B1 (n=30)	0,045 (0,03)	0,045 (0,05)	0,053 (0,06)	0,012 (0,03)	
	B2 (n=11)	0,044 (0,05)	0,058 (0,04)	0,041 (0,06)	0,041 (0,05)	
	B1 (n=30)	0,596 (0,12)	0,287 (0,09)	0,033 (0,04)	0,008 (0,02)	
	B2 (n=16)	0,512 (0,12)	0,296 (0,09)	0,072 (0,06)	0,026 (0,04)	
	B2 (n=16)	0,035 (0,07)	0,010 (0,02)			

Tarea 2					
		Presente de indicativo	Infinitivo	Presente de subjuntivo	Pretérito imperfecto de indicativo
Nivel MCER		Promedio (DT)	Promedio (DT)	Promedio (DT)	Promedio (DT)
Escala holística	A (n=12)	0,640 (0,16)	0,258 (0,12)	0,036 (0,05)	0,024 (0,05)
	B1 (n=24)	0,617 (0,11)	0,285 (0,08)	0,033 (0,05)	0,009 (0,02)
	B2 (n=25)	0,530 (0,13)	0,302 (0,10)	0,046 (0,05)	0,019 (0,03)
Escala gramatical	A (n=15)	0,644 (0,14)	0,272 (0,11)	0,023 (0,03)	0,020 (0,05)
	B1 (n=30)	0,596 (0,12)	0,287 (0,09)	0,033 (0,04)	0,008 (0,02)
	B2 (n=16)	0,512 (0,12)	0,296 (0,09)	0,072 (0,06)	0,026 (0,04)
		Condicional simple	Gerundio		
Nivel MCER		Promedio (DT)	Promedio (DT)		
Escala holística	A (n=12)	0,006 (0,01)	0,012 (0,02)		
	B1 (n=24)	0,019 (0,05)	0,003 (0,01)		
	B2 (n=25)	0,042 (0,07)	0,013 (0,02)		
Escala gramatical	A (n=15)	0,005 (0,01)	0,004 (0,01)		
	B1 (n=30)	0,032 (0,06)	0,011 (0,02)		
	B2 (n=16)	0,035 (0,07)	0,010 (0,02)		

Tabla VIII: Promedio y desviación típica (DT) de la frecuencia de aparición de paradigmas verbales en los textos clasificados por niveles.

La prueba a posteriori Games Howell no es consistente con algunos resultados de la prueba anterior. En concreto, y en el caso de los textos clasificados según la escala holística, la aparición del pretérito perfecto compuesto de indicativo, no podría discriminar entre ninguno de los niveles establecidos por esa escala. La situación es diferente si observamos la clasificación que arroja la escala gramatical: el condicional simple puede distinguir entre los niveles adyacentes A y B1 ($p < 0,01$) y tanto el pretérito perfecto simple como compuesto de indicativo lo hace entre los niveles A y B2 (en el nivel $p < 0,01$ en el caso del primero y en el nivel $p < 0,05$ en el caso del segundo).

En cuanto a los textos de la tarea 2, el contraste de Kruskal-Wallis muestra que las diferencias entre los promedios solo resultan significativas para dos paradigmas verbales en los niveles establecidos por la escala de alcance y control gramatical: el presente de indicativo ($H=7,571$; $p=0,023$) y el de subjuntivo ($H=7,717$; $p=0,021$). El test Games Howell señala que esas formas no distinguen entre niveles consecutivos pero sí lo hacen entre A y B2 ($p < 0,05$).

En la tabla IX resumimos los contrastes estadísticamente significativos que las variables analizadas han sido capaces de establecer en los textos de nuestro corpus.

Tarea 1						
	Escala holística			Escala gramatical		
	A/B1	B1/B2	A/B2	A/B1	B1/B2	A/B2
Nº de paradigmas verbales diferentes						✓
Pretérito perfecto compuesto de indicativo						✓
Pretérito perfecto simple						✓
Condicional simple				✓		
Tarea 2						
	Escala holística			Escala gramatical		
	A/B1	B1/B2	A/B2	A/B1	B1/B2	A/B2
Nº de paradigmas verbales diferentes		✓				✓
Presente de indicativo						✓
Presente de subjuntivo						✓

Tabla IX: Resumen de las distinciones estadísticamente significativas halladas entre niveles

Se observa claramente la práctica incapacidad de las variables de discriminar entre niveles consecutivos. También es evidente que este tipo de mediciones son mucho más efectivas en la distinción de los niveles que establece la escala de alcance y control gramatical que en los de la escala holística, que representa un constructo más amplio.

3. DISCUSIÓN

3.1. LA RELACIÓN ENTRE LAS MEDICIONES CUANTITATIVAS DE COMPLEJIDAD GRAMATICAL Y LOS JUICIOS SUBJETIVOS DE CALIFICACIÓN MEDIANTE ESCALAS

Nuestra primera pregunta de investigación estaba dedicada a investigar el grado de correspondencia entre la medición objetiva de la variedad verbal y las calificaciones de los textos de examen derivadas de las escalas de calificación. Como hemos visto en la exposición de resultados, encontramos correlaciones débiles aunque estadísticamente significativas. Esas correlaciones indicarían que hay relación entre los niveles funcionales del MCER que certifica el examen de la Universidad Nebrija y las mediciones de complejidad lingüística. Las correlaciones halladas pueden constituir también evidencia de la validez de criterio del examen que investigamos ya que se encuentra correspondencia entre las calificaciones y otros instrumentos de medida como son nuestras variables de complejidad gramatical.

Una cuestión interesante es que las correlaciones entre la variedad verbal y las calificaciones son, prácticamente en todos los casos, más fuertes con la escala holística que con la gramatical. Como ya hemos apuntado, no es sorprendente la debilidad de las correlaciones ya que ambas escalas cubren constructos más amplios que la variedad verbal, a cuya medición se limitan nuestras variables. Lo que quizá sí lo sea es que la correlación sea mayor con la escala holística (que contiene descriptores sobre logro y eficacia comunicativa, adaptación al género, mecanismos de cohesión y coherencia y alcance y control léxico-gramatical) que con la escala gramatical (que se limita a observar el nivel de complejidad y corrección de las estructuras gramaticales que aparecen en los textos). La explicación puede estar en el hecho de que al juzgar de acuerdo con la escala gramatical, que focaliza en alcance y control, el análisis analítico haya priorizado los factores relacionados con la corrección, que es un aspecto que nosotros no hemos medido. Prácticamente toda la bibliografía consultada resalta la importancia de la interacción entre corrección y complejidad a la hora de delimitar niveles de dominio y desarrollo y de juzgar la calidad de la escritura⁸. También se nota en los resultados que la escala de alcance y control gramatical se muestra en ocasiones concretas más sensible que la escala holística a ciertos rasgos como las frecuencias de aparición de algunas formas verbales. Por ejemplo, mientras que la escala holística no aprecia correlaciones con el uso del pretérito perfecto simple (T1) y del presente de subjuntivo (T2) sí lo hace la escala gramatical⁹. Un análisis cualitativo de los textos podría mostrarnos si en esos casos se trata de una disminución de errores debido, por ejemplo, al uso correcto del subjuntivo en lugar de otras formas¹⁰ o si se trata de indicadores que contribuyen puramente a la complejización. Ese tipo de análisis está más allá de nuestros objetivos pero las incógnitas planteadas invitan a proseguir la investigación y constatan la limitación de las interpretaciones de los resultados a partir del análisis de la complejidad aislada de la corrección.

3.2. LA COMPLEJIDAD GRAMATICAL Y LA DISCRIMINACIÓN ENTRE NIVELES DE DOMINIO

La pregunta de investigación 2 se dedicó a ver si realmente, más allá de las correlaciones que acabamos de comentar, se producían diferencias estadísticamente significativas entre niveles en cuanto a la complejidad gramatical que medían nuestras variables. En general, nuestros datos muestran que la delimitación entre unos niveles y otros es borrosa, sobre todo si se trata de niveles consecutivos, pero que algunos rasgos lingüísticos que hemos analizado tienen capacidad de discriminar entre ellos. Nuestra interpretación de los resultados se complica al considerar en nuestros datos dos escalas de calificación diferentes que dan lugar a dos clasificaciones de los textos en niveles que no son coincidentes en bastantes casos. También, como ya hemos ido notando, consideramos dos tipos de textos (los de la T1 y los de la T2) cuyo análisis arroja valores y comportamientos de las variables muy diferentes. Además de todo ello, las limitaciones del propio análisis, que no aporta datos sobre la corrección de uso de las estructuras ni sobre otros fenómenos de complejidad local, oscurece en muchas ocasiones el significado de los resultados conseguidos. A pesar de todo, nuestra conclusión no es diferente de las que se pueden leer en otros estudios sin las limitaciones del nuestro. Por ejemplo, Kang (2013) concluye su completísimo análisis sobre una prueba oral de los niveles B1-C2 de los exámenes Cambridge con las siguientes palabras:

“(D)istinctive patterns in linguistic features across CEFR speaking levels (...) were found, even though there was some fuzziness of distinctions at adjacent levels. The complexity of the configuration of components in any overall judgement of proficiency, and the fuzziness of distinctions between levels seem to be unavoidable” (Kang, 2013:47)

Teniendo todo lo anterior en cuenta, intentaremos utilizar nuestros resultados para hacer una descripción de los posibles perfiles de uso de las formas verbales que caracterizan los niveles de competencia certificados por el examen. Por último, comentaremos la relación de los resultados con algunos descriptores de la escala de calificación de alcance y control gramatical.

Según los resultados de nuestras variables, los textos de nivel A11 muestran una variedad verbal muy similar a la de los textos del nivel B1 ya que se observan pocas diferencias estadísticamente significativas entre ambos niveles. Sin embargo, la escala de alcance y control gramatical muestra que los aprendices de nivel A podrían hacer un uso significativamente menor del condicional simple que los de nivel B1 (al menos en los textos de la T1 donde la variedad verbal es considerablemente mayor que en los de la T2). La conclusión más evidente es que las variables que hemos utilizado difícilmente describen un perfil lingüístico exclusivo de la producción escrita de los aprendices de nivel A y las diferencias que se encuentran son visibles únicamente en textos de comunicación interpersonal (T1).

Como consecuencia de lo anterior, el nivel B1 se caracteriza por un mayor uso del condicional simple que el que se encuentra en el nivel A. También, aunque solo lo detecta la escala holística en los textos de la T1, el nivel B1 se caracterizaría por una variedad verbal significativamente menor que la que encontramos en el nivel B2.

Por último, el nivel B2 muestra mayor diversidad verbal que el nivel B1 y, como era de esperar, claras diferencias con las características del nivel A: mayor diversidad verbal, crecimiento del uso del pretérito perfecto simple y compuesto de indicativo en las respuestas a la T1 y en las de la T2 disminución del uso del presente de indicativo y aumento del de presente de subjuntivo.

En general, pero sobre todo a propósito del análisis de la frecuencia verbal, surge una duda sobre la posible circularidad del valor distintivo de las mediciones cuando usamos los niveles que se derivan de la aplicación de la escala de alcance y control gramatical. Esta escala hacía mención explícita de algunas de las variables que hemos usado en nuestro trabajo y que recogemos en la tabla X.

<B1	... Predominio de estructuras gramaticales simples. Comete errores en el uso de los tiempos verbales del pasado.
B1	Utiliza una pequeña variedad de estructuras gramaticales complejas, por ejemplo, uso de las oraciones subordinadas con predominio del indicativo ... Es capaz de usar con corrección los pasados de indicativo. Produce oraciones subordinadas con presente de subjuntivo y, ocasionalmente, con imperfecto de subjuntivo...
B2	Usa una variedad de estructuras gramaticales complejas, lo cual se aprecia en el empleo de la subordinación donde se destaca el control del presente, imperfecto y perfecto de subjuntivo, además de la utilización del condicional y las formas del imperativo...

Tabla X: Referencias a la complejidad morfosintáctica en la escala de alcance y control gramatical del examen multinivel de español académico de la Universidad Nebrija.

La escala considera la presencia del presente de subjuntivo en B1 junto con la aparición ocasional del imperfecto de subjuntivo. En B2, en cambio, el aprendiz controlaría ya el uso de presente, imperfecto y perfecto de subjuntivo. Nosotros no hemos analizado la corrección, por lo que no podemos decir si realmente se produce ese control, pero observamos que las frecuencias de aparición de esos tiempos verbales no parece corresponderse con lo que prevén las escalas. Por ejemplo, como puede verse en nuestra exposición de resultados, el presente de subjuntivo muestra un crecimiento lineal y sostenido a lo largo del espectro de niveles pero su presencia no es puramente anecdótica en el nivel A como podría inferirse de la escala, que prevé que en ese nivel hay un “predominio de estructuras simples”. Pero los datos más dispares respecto a la escala nos los ofrecen el imperfecto y el perfecto de subjuntivo, que tienen muy baja frecuencia en los textos de respuesta. Siempre según el análisis de la clasificación de textos derivada de la escala gramatical, el imperfecto de subjuntivo tiene una frecuencia media de aparición de 0,007 en el nivel A, 0,007 también en el nivel B1 y 0,005 en el nivel B2 de la T1; en la T2 la frecuencia media es de 0,002 en el nivel A, 0,003 en el B1 y 0,006 en el nivel B2. Por su parte el perfecto de subjuntivo en la tarea 1 tiene un promedio de frecuencia de 0,002 en A, 0,006 en B1 y 0,004 en B2; en la tarea 2 es prácticamente inexistente ya que no aparece en los niveles A y B2 y registra una frecuencia media de 0,002 en el nivel B1. Estos datos parecen indicar que esos paradigmas no son precisamente característicos del nivel B2.

El resto de los tiempos mencionados en la escala son el imperativo, el condicional y los tiempos de pasado. Los dos primeros se mencionan en el nivel B2 de la escala y su comportamiento en los textos no parece corresponderse con ello. El imperativo es prácticamente inexistente en los textos. No aparece en la T2 y en la T1 tiene una frecuencia promedio de 0,008 en el nivel A, 0,005 en el B1 y 0,007 en el B2. El condicional simple, por su lado, hemos visto que irrumpe en el nivel B1 discriminándolo del nivel A y no produciendo diferencia estadística entre B1 y B2. En cuanto a los usos de los tiempos de pasado, en la escala se especula sobre todo con su grado de corrección y nosotros no podemos decir mucho al respecto. Sí parece razonable pensar que el poder de discriminación que tienen el pretérito perfecto simple y compuesto entre el nivel A y el B2 de la T1 pueda significar que esos tiempos de pasado toman su justo papel en los movimientos narrativos de los textos. Pero la frecuencia de uso parece indicar que la estabilidad del sistema aparece en B2 y no en B1 como sugieren los descriptores de la escala. En definitiva, y pese a lo limitado de un análisis como el nuestro, que no incluye variables de corrección, nos parece que podemos descartar la circularidad entre el comportamiento de las variables y los descriptores de la escala gramatical que se refieren específicamente a ellas.

Por otro lado, una de las posibles aplicaciones de estudios como el nuestro es contribuir con datos empíricos a la mejora de los descriptores de calificación. Tras el análisis que acabamos de hacer se puede poner en duda, por ejemplo, la utilidad que pueda tener para los calificadores mencionar en las escalas tiempos verbales de muy baja frecuencia y comportamiento no lineal (imperativo, perfecto e imperfecto de subjuntivo). También es dudoso el nivel en el que se sugiere que emergen ciertos tiempos (el condicional) o el momento en el que los aprendices toman control sobre ellos (tiempos de pasado).

En definitiva, las correlaciones entre los juicios subjetivos derivados de las escalas de calificación y las mediciones objetivas de variedad y frecuencia verbal muestran que algunos rasgos lingüísticos se desarrollan linealmente a medida que los aprendices avanzan en el espectro de niveles del MCER que certifica el examen multinivel de la Universidad Nebrija. Es más, algunos de esos rasgos son capaces de discriminar entre textos clasificados por los evaluadores en niveles diferentes, incluso entre textos pertenecientes a niveles consecutivos. Este resultado no está en consonancia con algunos de los trabajos que hemos consultado (Prodeau, Lopez y Véronique, 2012 y Kuiken, Vedder y Gilabert, 2010) pero si lo está con otros muchos que encuentran correlatos morfosintácticos que distinguen diferentes niveles de dominio vinculados al MCER (Forsberg y Bartning, 2010; Martin et al. 2010; Toropainen y Lahtinen, 2014; Salamoura y Saville, 2010; Gyllstad et al., 2014 y Kang, 2013). Vemos, por tanto, que a pesar de las diferencias metodológicas entre los diferentes estudios y de las limitaciones de nuestro trabajo que luego señalaremos, los resultados que hemos obtenido coinciden en lo esencial con los de mucha parte de la bibliografía.

Hay que decir además que pese a las correlaciones encontradas y la significación estadística hallada entre algunas diferencias de promedios entre grupos, nuestros datos también evidencian la variabilidad individual subyacente al análisis grupal, la falta de crecimiento lineal de algunas variables y la borrosidad que existe en muchas ocasiones entre niveles, sobre todo entre niveles adyacentes. De estos extremos también se hacen eco algunos de los trabajos que hemos consultado y nos parece que es una característica importante del estado de la cuestión. Ya hemos citado a Kang (2013) cuando se refiere a la alta variabilidad individual y a la indistinción de muchos indicadores entre niveles adyacentes. En este sentido, estamos de acuerdo con Prodeau, Lopez y Véronique (2012) que explican en su trabajo que las dimensiones pragmáticas y sociolingüísticas necesarias para conseguir éxito comunicativo imposibilitan muchas veces la tarea de aislar recursos puramente formales que singularicen a un solo nivel de competencia comunicativa del MCER.

3.3 LIMITACIONES Y LÍNEAS FUTURAS DE INVESTIGACIÓN

Algunas limitaciones de este estudio ya han salido a la luz en la descripción del método y en la discusión de los resultados. Se trata, por ejemplo, del hecho evidente de haber llevado a cabo el análisis sin conocer muchas variables de presagio que pueden haber intervenido en los resultados. Algunas de ellas, la L1 de los participantes sobre todo, pueden haber influido considerablemente en los valores de las variables de complejidad que hemos analizado. Otra cuestión relevante es la naturaleza de las escalas usadas en este estudio para clasificar los textos por niveles. Ninguna de ellas estaba totalmente exenta de referencias a cuestiones gramaticales. Aunque creemos que nuestros resultados no se ven excesivamente afectados por una posible circularidad entre variables cuantitativas y escalas de calificación, pensamos que serían más robustos si la clasificación de los textos en niveles del MCER se hubiera derivado exclusivamente de criterios funcionales y de adecuación comunicativa.

Pero la que, en nuestra opinión, es la limitación más importante de este trabajo es la que también abre interesantes líneas de investigación para el futuro. Como ya hemos tenido ocasión de notar, nuestro análisis se ha circunscrito a variables de variedad verbal cuya interpretación resulta difícil y parcial sin tener datos sobre su interacción con el grado de corrección formal de los textos. Nos parece que ese ámbito de investigación pueden aportar información transcendental para obtener una visión más completa de los rasgos lingüísticos que singularizan los diferentes niveles del MCER en español. En la misma línea, el rastreo de la evolución en los diferentes niveles de construcciones idiosincrásicas del español aportaría datos empíricos relevantes, no solo para llegar a una descripción más ajustada a la realidad de los niveles de referencia, sino también para mejorar la enseñanza y la evaluación del español como lengua extranjera.

1 Ese es el tipo de trabajo que se llevó a cabo para elaborar el *Plan curricular del Instituto Cervantes. Niveles de referencia para el español*.

“Las decisiones adoptadas en el proyecto, por tanto, en todo lo relacionado con los esquemas conceptuales de cada inventario, la selección y el tratamiento del material y su distribución a lo largo de los seis niveles de progresión curricular son el resultado de un largo y minucioso proceso basado en dos factores clave: por una parte, la experiencia docente en la enseñanza del español de cerca de 250 profesores —entre los de plantilla del Instituto y los colaboradores externos— y los conocimientos y el juicio experto de profesores de reconocido prestigio especializados en los distintos temas; por otra, un largo proceso de revisión y depuración en el que los participantes en el proyecto, a partir de sus propios conocimientos y experiencia, y desde distintas situaciones profesionales, han ido conformando el material de cada uno de los inventarios.”(Instituto Cervantes, 2006).

Véase, al respecto de esto, el trabajo empírico que se está realizando para la descripción de los niveles de referencia en inglés (Salamoura y Saville, 2010).

2 En el área de la evaluación, y más concretamente en la terminología típica de las escalas de calificación, el constructo de la complejidad aparece con frecuencia bajo el nombre de alcance léxico y gramatical.

3 Muchos pertenecen al grupo de investigación SLATE, *Second Language Acquisition and Testing in Europe*, dedicado a investigar cuáles son los rasgos lingüísticos que caracterizan la interlengua de aprendices de lenguas extranjeras situados en los niveles del MCER. El grupo trabaja con siete lenguas europeas (holandés, inglés, francés, finés, alemán, italiano y sueco).

4 En principio, la muestra consistía en 126 textos y 63 participantes. Durante el análisis nos percatamos de que uno de ellos no seguía las instrucciones de las tareas por lo que sus textos quedaron excluidos del estudio.

5 En todo el proceso de evaluación participaron 11 calificadores, todos ellos expertos en la enseñanza y evaluación de lenguas en el ámbito del MCER.

6 Se puede acceder a más información y a una versión de demostración gratuita del programa en <http://connexor.fi>

7 En el sitio web que también mantiene el autor, www.real-statistics.com, se pueden encontrar manuales y explicaciones sobre los procedimientos de análisis y las herramientas que ofrece el programa.

8 Otra cuestión interesante que surge es cómo los calificadores evalúan los riesgos que toman los aprendices y su capacidad de llevarlos a buen puerto. Las tensiones entre corrección y riesgo constituyen un tema recurrente de discusión entre los profesionales de la enseñanza y la evaluación de L2 porque en demasiadas ocasiones se penaliza la toma de riesgos o, también, la inconsistencia e inestabilidad inherente a las interlenguas en francos períodos de maduración y/o reorganización.

9 Sin embargo, el incremento de uso del condicional simple y del pretérito perfecto compuesto de indicativo a medida que se avanza de nivel (T2) es un aspecto al que solo es sensible la escala holística. En ambos casos la aparición de los tiempos puede estar relacionada con aspectos pragmáticos y discursivos. Por ejemplo, por la lectura de los textos parece que la aparición del pretérito perfecto en la T2 está relacionada con movimientos narrativos del texto que vienen a apoyar los argumentos del autor mediante el relato de experiencias personales. Esto bien puede haber influido en la percepción de la calidad de los textos por parte de los examinadores cuando usaban la escala holística. Sería necesario, de todos modos, un análisis cualitativo de mayor profundidad que la mera lectura de los textos para confirmar hipótesis de este tipo.

10 Lo mismo puede ocurrir con el comportamiento del pretérito perfecto e imperfecto. La lectura superficial de los textos durante nuestro análisis parecía invitar a pensar que el incremento de aparición del pretérito perfecto en los niveles más altos venía a corregir un error frecuente de los niveles más bajos en la T1 donde los examinados utilizaban erróneamente el pretérito imperfecto de indicativo para relatar su experiencia con la asignatura y justificar su solicitud de revisión de la nota. Finalmente, no se confirmó una correlación estadísticamente significativa del imperfecto con las calificaciones, aunque sí se constata una correlación negativa (coeficiente *tau* de Kendall -0,130) y un decremento lineal de su frecuencia según avanzan los niveles en ambas escalas de calificación.

11 Nosotros hemos etiquetado así los textos que los evaluadores, siguiendo las instrucciones de las escalas, han considerado que estaban por debajo del nivel B1. Este procedimiento no parece que haya convertido al nivel A en un cajón de sastre ya que no muestra mayor variabilidad que los niveles B1 y B2 que certifica el examen. Parece, por tanto, que se trata realmente del nivel anterior al B1, es decir, un nivel A2, aunque no lo podamos denominar de esta forma ya que ese nivel está ausente de las especificaciones del examen. ?

REFERENCIAS BIBLIOGRÁFICAS

- Bulté, B. y Housen A. 2012. Defining and operationalising L2 complexity. En Housen, A., Kuiken, F., y Vedder, I., eds. *Dimensions of L2 performance and proficiency: complexity, accuracy and fluency in SLA* (Vol. 32). John Benjamins Publishing, pp. 34-59
- Consejo de Europa, 2002. *Marco común europeo de referencia para las lenguas: aprendizaje, enseñanza, evaluación*, Madrid: Secretaría General Técnica del MEC-Subdirección General de Información y Publicaciones y Grupo ANAYA.
- Ellis, R. y Yuan, F., 2004. The effects of planning on fluency, complexity, and accuracy in second language narrative writing. *Studies in Second Language Acquisition*, 26, pp.59–84.
- Figueras, N., 2008. El MCER, más allá del éxito. *Monográficos MarcoELE*, 7, pp.26–35.
- Forsberg, F. y Bartning, I., 2010. Can linguistic features discriminate between the communicative CEFR-levels? A pilot study of written L2 French. I. Bartning, M. Martin, y A. Vedder, eds. *Communicative proficiency and linguistic development: intersections between SLA and language testing research*. Eurosla Monographs Series, 1, pp. 133–158.
- Gyllstad, H. et al., 2014. Linguistic correlates to communicative proficiency levels of the CEFR: The case of syntactic complexity in written L2 English, L3 French and L4 Italian. *EUROSLA Yearbook*, 14, pp.1–30.
- Herrera, H., Martínez, R. y Amengual, M., 2011. *Estadística aplicada a la investigación lingüística*. Madrid: EOS Universitaria.
- Housen, A., Kuiken, F., y Vedder, I., 2012. Complexity, accuracy and fluency: definitions, measurement and research. En Housen, A., Kuiken, F., y Vedder, I., eds. *Dimensions of L2 performance and proficiency: complexity, accuracy and fluency in SLA* (Vol. 32). John Benjamins Publishing, pp. 1-20
- Instituto Cervantes (2006). *Plan curricular del Instituto Cervantes. Niveles de referencia para el español*. Madrid: Instituto Cervantes-Biblioteca nueva.
- Kang, O., 2013. Linguistic analysis of speaking features distinguishing general English exams at CEFR levels. *Cambridge English Research Notes*, 52, pp.40–48.
- Kuiken, F., Vedder, I. y Gilabert, R., 2010. Communicative Adequacy and Linguistic Complexity in L2 writing. En I. Bartning, M. Martin, y A. Vedder, eds. *Communicative proficiency and linguistic development: intersections between SLA and language testing research*. Eurosla Monographs Series, 1, pp. 81–100.

- Larsen-Freeman, D., 2009. Adjusting Expectations: The Study of Complexity, Accuracy, and Fluency in Second Language Acquisition. *Applied Linguistics*, 30(4), pp.579–589.
- Macwhinney, B., 2000. *The CHILDES Project: Tools for Analyzing Talk* 3a ed., Mahwah, NJ: Lawrence Erlbaum Associates.
- Martin, M. et al., 2010. On Becoming an Independent User. En I. Bartning, M. Martin, y I. Vedder, eds. *Communicative proficiency and linguistic development: intersections between SLA and language testing research*. Eurosla Monographs Series, 1, pp. 57–80.
- Norris, J.M. y Ortega, L., 2009. Towards an Organic Approach to Investigating CAF in Instructed SLA: The Case of Complexity. *Applied Linguistics*, 30(4), pp.555–578.
- Ortega, L., 2003. Syntactic complexity measures and their relationship to L2 proficiency: A research synthesis of college? level L2 writing. *Applied linguistics*, 24(4), pp.492–518.
- Ortega, L., 2012. Interlanguage complexity. En *Linguistic Complexity: Second Language Acquisition, Indigenization, Contact*. Walter de Gruyter, pp. 127–155.
- Ortega, L., 2015. Syntactic complexity in L2 writing: Progress and expansion. *Journal of Second Language Writing*, 29, pp.82–94.
- Pallotti, G., 2009. CAF: Defining, Refining and Differentiating Constructs. *Applied Linguistics*, 30(4), pp. 590–601.
- Prodeau, M., Lopez, S. y Véronique, D., 2012. Acquisition of French as a Second Language : Do developmental stages correlate with CEFR levels? *Apples – Journal of Applied Language Studies*, 6(1), pp.47–68.
- Salamoura, A. y Saville, N., 2010. Exemplifying the CEFR: criterial features of written learner English from the English Profile Programme. En I. Bartning, M. Martin, y I. Vedder, eds. *Communicative proficiency and linguistic development: intersections between SLA and language testing research*. Eurosla Monographs Series, 1, pp. 101–132.
- Toropainen, O. y Lahtinen, S., 2014. Interrogative Clauses across CEFR Levels in Finnish and Swedish as an L2. *Apples – Journal of Applied Language Studies*, 8(3), pp.71–84.
- Yuan, F. y Ellis, R., 2003. The Effects of Pre-Task Planning and On-Line Planning on Fluency, Complexity and Accuracy in L2 Monologic Oral Production. *Applied Linguistics*, 24(1), pp.1–27.
- Zaiontz C. (2015) Real Statistics Using Excel. En línea: www.real-statistics.com [Fecha de acceso 17-10-2015]