

El análisis de la complejidad lingüística en los niveles de competencia del MCER: retos y propuestas

The analysis of linguistic complexity across CEFR proficiency levels: challenges and proposals

Joseba Ezeiza Ramos

Universidad del País Vasco .

joseba.ezeiza@ehu.eus

Ezeiza Ramos, J. (2016). El análisis de la complejidad lingüística en los niveles de competencia del MCER: retos y propuestas. *Revista Nebrija de Lingüística Aplicada* (2016) 20.

RESUMEN

En este artículo se reflexiona acerca de algunos de los retos que rodean la validación de criterios para la evaluación de la complejidad lingüística en exámenes certificativos multinivel. Igualmente, se realizan algunas propuestas para abordar dichos retos a partir de los principios en los que se sustenta el marco teórico que ofrecen la Teoría de la Complejidad (CT) y la Teoría de los Sistemas Dinámicos (DST).

Palabras clave: evaluación certificativa, complejidad lingüística, MCER, Teoría de la Complejidad, Teoría de los Sistemas Dinámicos

ABSTRACT

This report is intended to point out some of the problems that affect the validation of criteria to assess the linguistic complexity of the productions of L2 students in multilevel proficiency assessment tests. Furthermore, I will make a proposal to deal with those challenges on the basis of the principles of the theoretical framework or the Complexity Theory and the Dynamic Systems Theory.

Keywords: proficiency assessment, linguistic complexity, CEFR, Complexity Theory, Dynamic Systems Theory

Fecha de recepción: 15/2/2016

Fecha de aceptación: 30/3/2016

1. APROXIMACIÓN GENERAL A LA VALIDACIÓN DE ESCALAS Y DESCRIPTORES DE DOMINIO DE LA LENGUA

Tal y como evidencia el estudio elaborado por De la Torre García (2016), la validación de pruebas certificativas constituye un reto que plantea bastantes dificultades y que está abierto a múltiples incógnitas. El reto resulta más exigente, aún si cabe, cuando la prueba certificativa en cuestión pretende clasificar a los candidatos en un espectro multinivel que cubre un espacio tan extenso como el que, siguiendo la nomenclatura del Marco común europeo de referencia para las lenguas (MCER), convenimos denominar A-C1.

Una de los elementos críticos en este tipo de pruebas lo constituyen indudablemente los descriptores que forman parte de las escalas de nivel que se aplicarán para clasificar los candidatos que realicen dichas pruebas. De hecho, se trata de un elemento crucial del constructo de la prueba. Por una parte los descriptores y escalas constituyen una referencia fundamental tanto para los evaluadores como para los candidatos que se presenten a la misma. Por otra parte, tendrán un correlato directo en la validez y fiabilidad de los juicios que se emitan.

Los criterios generales a considerar en la elaboración de descriptores y escalas de evaluación se resumen en la guía que proporciona el MCER (2002: 197-208). En dicha guía se establece que los criterios deberán cumplir los requisitos de “precisión”, “claridad”, “brevedad” e “independencia” (MCER, 2002: 198-199). También se proponen diversos métodos para la elaboración de descriptores o escalas. Entre ellos se diferenciarían tres grupos: los denominados “métodos intuitivos”, los “métodos cualitativos” y los “métodos cuantitativos” (MCER, 2002: 200-204). En dicha guía también se sugiere una doble vertiente de análisis “a partir de los descriptores” y “a partir de las muestras de actuación” (MCER, 2002: 196).

Pero, en definitiva, en coherencia con el planteamiento general del MCER, se deja en manos de los administradores de pruebas la responsabilidad última de determinar “en qué medida se da a las calificaciones otorgadas en el sistema un sentido compartido mediante definiciones comunes” y qué métodos “se utilizan para desarrollar dichas definiciones” (MCER, 2002: 204). Por lo tanto, cualquier sistema de evaluación que pretenda referirse al MCER deberá incorporar mecanismos de control de calidad que garanticen la validez y fiabilidad de los instrumentos de medida que aplica.

Algunos de los aspectos a considerar en la validación de descriptores han sido previamente tratados, entre otros, en dos trabajos publicados anteriormente en esta misma revista por Del Moral Manzanares (2013 y 2014) y en una memoria de máster publicada en RedEle (Gutiérrez, 2012) En dichos trabajos quedan ilustradas algunas de las metodologías que pueden resultar útiles para realizar validaciones “a partir de los descriptores”.

El trabajo de De la Torre (2016), por su parte, ilustra cómo puede abordarse esta tarea “a partir de las muestras de actuación”, aplicando una metodología básicamente cuantitativa. No obstante, dado que en dicho trabajo se consideran conjuntamente juicios de evaluadores basados en dos modelos de escala (una de carácter holístico y otra de alcance y control gramatical) con mediciones objetivas de variedad verbal en un corpus de producciones de estudiantes de ELE de diferentes niveles, podríamos considerar que se trataría de una metodología “híbrida”. Así pues, se trataría de una propuesta ambiciosa que, en principio, podría presumiblemente realizar interesantes aportaciones.

Sin embargo, y dado precisamente lo prometedor del exhaustivo y riguroso trabajo de De la Torre (2016), las conclusiones prácticas del estudio pueden resultar finalmente desalentadoras. De hecho, si bien en otros órdenes este trabajo realiza aportaciones interesantes, los datos arrojados por el estudio no permiten establecer con cierto grado de certidumbre qué matices convendría incorporar a los descriptores para que puedan refinarse adecuadamente.

De hecho, los resultados del trabajo presentado parecen sugerir: a) la posible influencia de la naturaleza de la tarea comunicativa en el perfil competencial evidenciado por los candidatos; b) el bajo poder discriminatorio del rasgo analizado (al menos en niveles adyacentes); y c) diferencias en las correlaciones entre las medidas de complejidad y los juicios de los evaluadores en función de si estos han sido realizados en base a la escala holística o en base a la escala de alcance y control gramatical.

Así pues, en este comentario se tratará de realizar una breve aportación al respecto, ampliando la discusión que la autora abre en su estudio. En primer lugar, se señalarán algunas de las dificultades que plantea el objetivo abordado por De la Torre (2016) en su trabajo. A continuación, se sugerirán posibles alternativas para, por una parte, plantear la posibilidad de reinterpretar los resultados obtenidos en un marco más amplio, y, por otra, reflexionar acerca de las vías futuras de desarrollo del mismo.

2. DIFICULTADES PARA LA VALIDACIÓN DE ESCALAS DE DOMINIO SINGULARIZADAS A PARTIR DE DESCRIPTORES BASADOS EN RASGOS LINGÜÍSTICOS

A grandes rasgos, la validez, fiabilidad y, en definitiva, el potencial discriminatorio de un dispositivo de evaluación dependerá de tres factores fundamentales: a) la naturaleza de las tareas que se solicitarán a los candidatos; b) el conjunto de indicadores que se apliquen para inferir su perfil competencial; y c) las expectativas relativas a la emergencia, frecuencia y uso de determinados recursos lingüísticos que puedan ser considerados representativos de un determinado nivel de dominio.

Por este motivo, las especificaciones técnicas de las pruebas de certificación adscritas al MCER, incluyen a) una descripción general de las características, condiciones y restricciones de cada una de las tareas a realizar; b) una escala, parrilla o rúbrica para cada uno de los indicadores a considerar; y c) una serie de indicaciones, escalas o criterios relativos a los rasgos lingüísticos que se considerarán indicativos del nivel de dominio en cuestión, y, en su caso, orientaciones relativas al grado de admisibilidad de los errores.

Aquellas pruebas que deseen referirse al MCER encontrarán en su documento básico (MCER, 2002) las categorías y escalas básicas necesarias para determinar los requisitos de las tareas a realizar (qué deberán saber hacer los candidatos de un determinado nivel) y los criterios generales de evaluación a aplicar (qué grado de experticia se espera que evidencien). Por lo tanto, en estas dos dimensiones, la tarea principal de los diseñadores de pruebas consistirá en seleccionar y adaptar aquellos criterios, escalas y descriptores que se ofrece el MCER, atendiendo a los objetivos y el contexto de aplicación de la prueba. En general, el desarrollo y validación de estas escalas se realiza, en primera instancia, aplicando una aproximación “a partir de los descriptores”.

Con respecto al tercero de los aspectos, el MCER no ofrece indicación alguna que sirva para establecer los correlatos lingüísticos de los niveles de dominio definidos para las pruebas. El MCER ofrece un marco operativo que permite armar la estructura general de dos de las dimensiones del constructo de las pruebas: a) la dimensión horizontal, que se refiere a los aspectos a evaluar y b) la dimensión vertical, en la que se establecen las escalas de dominio correspondiente. Sin embargo, el MCER deja en manos de los diseñadores de la prueba la tarea de substanciar la tercera dimensión, esto es, la que se refiere a los elementos, categorías, clases, estructuras, procesos y relaciones gramaticales, lexicales, funcionales y discursivas que puedan ser consideradas representativas de cada nivel de dominio.

Tal y como apunta De la Torre (2016) en su trabajo, el abordaje de esta dimensión puede realizarse de manera apriorística (de forma “intuitiva” y “a partir de los descriptores”) o puede realizarse de forma empírica (con “metodologías cualitativas y cuantitativas” adecuadas, y “a partir de las producciones de los estudiantes”). La autora del estudio se decanta –probablemente de forma acertada– por esta segunda aproximación y propone una metodología de validación que conjuga juicios de expertos basados en dos tipos de escalas, y datos empíricos extraídos de un corpus de producciones de candidatos-tipo.

Para indagar la viabilidad de esta metodología, y siguiendo la estela de diversos expertos e investigaciones que parecen sugerir que uno de los aspectos que pueden tener cierto valor predictivo del nivel de dominio de los candidatos en el intervalo que corresponde a los niveles B1, B2 y C1 podría ser el de la “complejidad lingüística”, De la Torre (2016) delibera centrar su atención en este aspecto, y delimita el estudio a un determinado rasgo lingüístico -la variedad verbal- que, si bien, en sí mismo, no sería representativo de la complejidad lingüística en su globalidad, sí podría virtualmente acercarnos en cierta medida al análisis de este parámetro.

El estudio se realiza de forma exhaustiva y rigurosa, pero, en concordancia con lo sugerido por el conjunto de trabajos realizados previamente sobre esta cuestión, los resultados no son concluyentes. Más bien, abren las puertas a muy diversos interrogantes, no referidos estrictamente al poder discriminante y predictivo del rasgo lingüístico en cuestión (la variedad verbal), sino, más bien, acerca de los postulados de partida del trabajo, acerca del propio dispositivo metodológico y acerca de la propia interpretación de los resultados.

Con respecto a los postulados de partida, cabría realizar una primera reflexión acerca de la naturaleza del MCER y sus objetivos. En ocasiones parece que se tiende a olvidar que el MCER no deja de ser un constructo programático, un instrumento de gestión y, en definitiva, un artificio meramente convencional, concebido para desplegar de forma ordenada y sostenible las políticas lingüísticas que integran el proyecto de creación y desarrollo de la Unión Europea.

Por lo tanto, si bien es indudable que el MCER cuenta con el aval de las principales instituciones académicas, y aun siendo también cierto que se trata de un documento con bases sólidas, no parece razonable suponer que los procesos individuales de adquisición y aprendizaje de las lenguas transiten de forma natural y ordenada por carriles tan estrechos y estructurados como los que proyecta el itinerario de niveles del MCER (A1, A2, B1, B2, C1 y C2). En todo caso, el planteamiento sería el inverso, y serían los estudiantes los que, convencidos de la utilidad administrativa y social del planteamiento del MCER, dirigirían sus esfuerzos a ajustar de forma deliberada su perfil competencial hacia lo establecido por el MECR.

Pero si el propio sistema de niveles no deja de ser un “artificio” en sí mismo, más artificiales serían aún los límites que puedan establecer dentro de cada nivel. Es lógico, comprensible y útil tratar de segmentar los procesos de enseñanza-aprendizaje en periodos temporales relativamente breves (y, que por lo tanto, se definan subniveles como el A1.1, el A1.2, etc.). De hecho, gracias a la versatilidad del MCER, resulta relativamente sencillo establecer qué tareas podrían ser representativas de dichos subniveles y qué descriptores podrían ser indicativos de una realización eficiente de las tareas en cuestión por parte de hablantes que se encuentran en el nivel de dominio correspondiente.

Sin embargo, no parece tan plausible que la generalidad de hablantes que se encuentren en cada uno de esos estadios, cuyos límites han sido establecidos arbitrariamente, resuelvan las tareas utilizando un repertorio lingüístico ajustado a un

perfil tan delimitado que presente una emergencia y frecuencia de uso de determinadas estructuras lingüísticas que pueda ser considerada específicamente representativa de dicho nivel. Parece también improbable que dos grupos de hablantes de niveles adyacentes próximos evidencien en la realización de una misma tarea perfiles lingüísticos significativamente distintivos en términos de emergencia y frecuencia de uso.

Por ende, la probabilidad de detectar estas tendencias y diferencias se reduce notablemente si las producciones solicitadas son extremadamente breves (como es el caso de las producciones que habitualmente se solicitan a los candidatos en pruebas de examen). Finalmente, la probabilidad de que podamos detectar ciertas regularidades, correlaciones y diferencias significativas se reduce aún más si se presta atención arbitrariamente a una única categoría lingüística.

De hecho, aunque en el caso de una producción concreta pueda considerarse relativamente probable la emergencia de determinada categoría gramatical, si el texto es breve, el número de ocurrencias será posiblemente muy bajo y no variará mucho ni en función del hablante ni del nivel de dominio de dicho hablante, precisamente, por tratarse de un texto breve. Por otra parte, cabe la posibilidad cierta de que cualquier hablante (nativo o no nativo) pueda resolver la tarea de forma eficaz sin necesidad de utilizar la categoría gramatical observada. En cualquier caso, dado que, en general, un candidato que se presenta a una prueba de examen trata de evidenciar el dominio de un abanico más o menos amplio de recursos lingüísticos, probablemente tratará deliberadamente de introducir distintas formas de expresión en sus producciones, lo cual también puede influir en los datos de emergencia y frecuencia de uso de una determinada categoría.

Todo ello invita a considerar con mucha cautela la expectativa de detectar categorías gramaticales u otro tipo de rasgos lingüísticos que puedan correlacionar estadísticamente con un determinado nivel. Incluso aunque esta correlación aparentara ser consistente, la ausencia o baja frecuencia de una determinada categoría en una producción breve podría explicarse precisamente por la brevedad de dicha producción o, a la inversa, el uso más frecuente de una determinada categoría podría deberse a que los candidatos presupongan que se podría considerar un indicio fiable de dominio de determinado nivel.

Obviamente, estas limitaciones podrían superarse si las expectativas se plantean no en relación con una serie de categorías lingüísticas individualizadas, sino en relación con cualidades más generales como podría ser, en el caso de las producciones escritas, el rasgo que denominamos "complejidad lingüística". Sin embargo, el trabajo de De la Torre (2016) señala también algunas dificultades para operativizar la complejidad lingüística.

Por una parte, abordada la cuestión desde un plano comunicativo, el propio trabajo de De la Torre (2016) asume que las dimensiones pragmáticas y sociolingüísticas necesarias para conseguir el éxito comunicativo imposibilitan muchas veces la tarea de aislar rasgos gramaticales formales que singularicen un solo nivel de competencia del MCER. De hecho, los resultados del trabajo realizado evidencian que unos mismos candidatos presentan perfiles distintos en tareas con demandas comunicativas diferentes.

Por otra parte, desde un punto de vista psicolingüístico, se acepta que las diferencias individuales, el dinamismo de los procesos de cambio y desarrollo de la competencia comunicativa, y la propia naturaleza de los sistemas lingüísticos podrían dificultar, si no contradecir, la búsqueda de indicadores que progresen linealmente y que sean capaces de discriminar estadísticamente entre niveles. También se apunta la posibilidad de que pudieran existir relaciones dinámicas entre los niveles de complejidad y los de fluidez y corrección. Esto es, cabe pensar que los niveles de complejidad pueden fluctuar en función de la cantidad de texto producido, las estrategias de producción del mismo y del esfuerzo realizado para controlar el uso preciso, adecuado y correcto de las categorías lingüísticas introducidas en el mismo.

Finalmente, en el plano lingüístico, se asume que el constructo de la complejidad lingüística presentaría una configuración multidimensional constituida por diversos subsistemas que podrían interrelacionar entre sí de forma dinámica, lo cual entraría en contradicción con la expectativa de poder discriminar a los candidatos considerando de forma aislada los datos correspondientes a categorías que forman parte de un determinado subsistema del constructo. De hecho, en referencia a la variedad verbal, un texto con un número limitado de palabras elaborado con cláusulas extensas presentaría, probablemente, mayor complejidad sintáctica, pero menor cantidad de formas verbales que otro elaborado mediante cláusulas más breves. En el primer caso, la probabilidad de emergencia de una mayor variedad verbal disminuye, al menos, en términos de frecuencia de uso y, en función de las estructuras sintácticas utilizadas, también en términos de emergencia de determinadas formas verbales.

Por todo ello, parece arriesgado referir la validez de la prueba a un solo parámetro (la complejidad lingüística) y, por añadidura, esperar encontrar categorías gramaticales asociadas a dicho parámetro que, considerados de forma aislada, puedan aportarnos criterios válidos para la asignación fiable de las producciones de los candidatos a determinados niveles y subniveles. Sin embargo, sí cabe la posibilidad de que, considerando un conjunto de categorías lingüísticas suficientemente

amplio y representativo de la complejidad lingüística podamos obtener una imagen general más certera del perfil-tipo de las producciones prototípicas de hablantes con diferentes niveles de dominio lingüístico.

No obstante, a la hora de establecer dichos prototipos sería sensato no obviar la inherente variabilidad intra- e interhablante de las producciones lingüísticas. Por una parte, en el propio trabajo de De la Torre (2016) ya se apunta la posibilidad de que rasgos como el de la complejidad lingüística puedan verse afectados por variables individuales de los hablantes, como podrían ser la L1, la edad, el perfil sociocultural, etc. Otro aspecto que puede condicionar las posibles expectativas de emergencia y frecuencia de uso de determinadas categorías gramaticales sería la familiaridad de los candidatos con la prueba de examen, en particular, sería interesante conocer si han recibido instrucción específica para afrontar dicha prueba. Finalmente, conviene considerar seriamente las exigencias sociodiscursivas propias del género textual en el que se substanciará la tarea comunicativa propuesta (rasgos lingüísticos prototípicos, coocurrencias y prominencia de categorías lingüísticas, ejes de variación, etc.).

De hecho, si se pretende ajustar ciertas garantías las escalas de alcance y control gramatical de una prueba multinivel, parece sensato tratar de controlar aquellas variables que puedan incidir en los rasgos lingüísticos de las producciones requeridas a los candidatos, para, de este modo, poder interpretar los resultados obtenidos de los análisis de dichas producciones desde diversas perspectivas. De esta interpretación cruzada se podrá obtener una imagen más certera de cuáles son los rasgos más representativos de las producciones de cada uno de los niveles y los márgenes de variación que pueden presentar dichos rasgos dentro de cada nivel.

En cierto modo, desde esta perspectiva, la hipótesis de partida sería distinta a la planteada por De la Torre (2016) en su estudio, ya que la nueva hipótesis postularía la posibilidad de que, por una parte, la complejidad lingüística prototípica de las producciones de cada nivel quede representada por un conjunto de rasgos distinto en cada caso, y que, incluso, entre ciertos casos las diferencias entre niveles no se reflejen directamente en producciones de niveles de complejidad significativamente distintos, sino en diferencias en la fluidez o la precisión gramatical que presenten los candidatos.

3. NUEVAS VÍAS DE ESTUDIO DE LAS EXPECTATIVAS DE COMPLEJIDAD LINGÜÍSTICA EN LAS PRODUCCIONES DE PRUEBAS CERTIFICATIVAS MULTINIVEL

A la luz de los argumentos planteados en el apartado anterior, cabe sugerir algunas propuestas para líneas futuras de investigación, además de las ya planteadas por De la Torre (2016) en la parte final de su trabajo.

En primer lugar, al igual que cualquier otra tarea del ámbito de la lingüística aplicada, parece conveniente circunscribir el diseño y validación de pruebas certificativas a un marco que permita establecer las hipótesis de trabajo sobre unas bases teóricas que se sitúen más allá de lo meramente utilitario. Los interrogantes y retos evidenciados en trabajo De la Torre (2016) sugieren que el marco que ofrecen la Teoría de la Complejidad (Complexity Theory) y la Teoría de Sistemas Dinámicos (Dynamic Systems Theory) podrían ofrecer algunos criterios para situar adecuadamente la discusión acerca de la predictibilidad de determinados comportamientos de los hablantes LE en respuesta a demandas como las tareas referidas en el estudio. Al respecto, pueden resultar especialmente útiles las aportaciones de trabajos como los de Herdina y Jessner (2002), Larsen-Freeman y Cameron (2008), Ellis y Larsen-Freeman (2009), Sampson, Gil y Trudgill (2009), Verspoor, De Bot y Lowie (2011), Larsen-Freeman (2012), Zamorano (2012) o Zheng (2012).

Precisamente, una de las cuestiones que emergen de este marco teórico se refiere al papel de la evaluación y de la forma de abordarla desde esta perspectiva. Hay diversos trabajos que señalan las cuestiones más candentes que se derivan de este punto de vista, entre ellos se encontrarían los de Brown, Hudson, Norris y Bonk (2002), Lantolf y Phoener (2004), Castello (2008), Phoener (2008), Mislevy (2009) o Rodríguez (2012). Entre otras cuestiones, estos trabajos abordan la pregunta que da título al artículo de Mislevy (2009): *If Language Is a Complex Adaptive System, What Is Language Assessment?* Y también aportan algunas sugerencias metodológicas para diseñar y validar pruebas de examen, teniendo en cuenta las asunciones que dan soporte a esta aproximación teórica.

En cuanto al constructo de complejidad lingüística, podría ser interesante considerar esquemas analíticos como el de Castello (2008), más orientados al análisis de la complejidad de los textos que sirven de soporte para las tareas de comprensión, pero que, quizás, podrían ser de utilidad para ampliar con nuevas dimensiones los constructos a los que De la Torre (2016) ya se refiere en su trabajo.

Otro aspecto a considerar sería el de la influencia de la tarea en el desempeño de los hablantes LE. En las siguientes referencias pueden consultarse los hallazgos de algunos estudios interesantes: Robinson (2011), De Jong, Steinel, Florijn, Schooneny Hulstijn (2012), Kuiken y Verder (2012), Levinka y Gilabert (2012). Igualmente convendría considerar los trabajos que analizan las posibles interrelaciones dinámicas entre parámetros de desempeño como la fluidez, la corrección y la complejidad: Wolfe-Quintero, Inagaki y Hae-Young (1998), Housen, Kuilen y Vedder (2012), Mavrou (2013), etc. Por su especial interés para el tema que nos ocupa, podría destacarse el trabajo de Mavrou (2013) en el que se plantea una cuestión que sobrevuela las reflexiones desarrolladas en este trabajo: ¿Hasta qué punto compiten entre sí la precisión gramatical y la complejidad sintáctica por los recursos limitados del aprendiz?.

Por otra parte, esta nueva perspectiva requeriría aplicar metodologías de análisis de datos un poco más complejas que el análisis de correlaciones o las pruebas de comparación utilizadas en estudios como el realizado por De la Torre (2016). En primera instancia, podrían ser de utilidad explorar el potencial que ofrece para describir y clasificar por niveles las producciones de hablantes no nativos un índice que actualmente se utiliza en otros ámbitos de la Lingüística Aplicada y que Cicres y Turell (2010) denominan "índice de distancia/similitud idiolectal" (ID/SI).

Igualmente, cabría considerar el uso de pruebas de análisis factorial (Mavrou, 2015) para estudiar los patrones de coocurrencias en producciones correspondientes a distintos perfiles de candidatos, distintos tipos de tareas y distintos niveles; de este modo, podrían posiblemente detectarse perfiles prototipo y los rasgos lingüísticos más idiosincrásicos de cada perfil, independientemente de que presenten o no una distribución lineal a lo largo del continuum de niveles de dominio. Trabajos como los de Biber (1988, 1995 y 2006), Biber y Finegan (1994), Biber y Conrad (2011) o Ezeiza y Alberdi (2015) evidencian el potencial de este tipo de análisis para establecer perfiles-tipo que consideren un conjunto amplio de categorías lingüísticas y un abanico de variables que pudieran incidir en su emergencia y frecuencia de uso.

Pero si se pretende un trabajo más ambicioso, podría sugerirse también el uso de técnicas estadísticas como la del Modelo de Ecuaciones Estructurales (Swaki, 2012; Schoonen, 2013; Winke, 2014).

En definitiva, si bien trabajos como el realizado por De la Torre (2016) suponen indudablemente un avance hacia el objetivo pretendido de dotar de mayor base empírica a los sistemas certificativos, también desvelan la gran dificultad que ello conlleva y la propia complejidad del proceso. Ahora bien, una vez que se ha dado ya este primer paso, parece obligado seguir avanzando. Si algo dejan patente trabajos como el aquí referido, es la evidencia de que los sistemas de certificación podrían ser más fiables y robustos si se apoyan en la investigación.

En este trabajo se han señalado algunas líneas innovadoras de investigación que actualmente se están ya explorando. Evidentemente, sería precipitado avanzar resultados de éxito, pero los planteamientos de las teorías y diseños metodológicos a los que se hecho referencia son sólidos y, al menos, muestran alentadoras vías futuras de indagación.

Así pues, no cabe sino animar a investigadores como De la Torre a que continúen trabajando en esta senda, abriendo, al menos un espacio para la reflexión y la discusión sobre cuestiones tan cruciales como la validez y fiabilidad de los dispositivos de certificación lingüística.

REFERENCIAS BIBLIOGRÁFICAS

- Biber, D. & Conrad, S. (2011). *Register, Genre and Style*. Cambridge: Cambridge University Press.
- Biber, D. & Finegan, E. (Eds.) (1994). *Sociolinguistic perspectives on register*. Oxford: Oxford University Press.
- Biber, D. (1988). *Variation across speech and writing*. Cambridge: Cambridge University Press.
- Biber, D. (1995). *Dimensions of Register Variation: A Cross-Linguistic Perspective*. Cambridge: Cambridge University Press.
- Biber, D. (2006). *University Languages. A corpus-based study of spoken and written register*. Ámsterdam: John Benjamins.
- Brown, J. D., Hudson, T., Norris, J. y Bonk, W. J. (2002). *An investigation of second language task-based performance assessments*. Manoa: University of Hawai'i,
- Castelo, E. (2008). *Text complexity and reading comprehension tests*. Nueva York: Peter Lang.
- Cicres, Jordi; Turell, Maria Teresa (2010). Índice de distancia/similitud idiolectal (ID/SI) aplicado a la producción no nativa (Interlengua) en español con finalidades descriptivas o formativas. *MarcoELE: revista de didáctica español como lengua extranjera*

11: 1- 12.

- De Jong, N. H., Steinel, M. P., Florijn, A., Schoonen, R. y Hulstijn, J. H. (2012). The effect of task complexity on functional adequacy, fluency and lexical diversity in speaking performances of native and non-native speakers. En Housen, A., Kuilen F. y Vedder I. (eds.). *Dimensions of L2 performance and proficiency: complexity, accuracy and fluency in SLA*. Amsterdam; Philadelphia: John Benjamins Pub. Co.: 121-142.
- De la Torre, N. (2016). La complejidad lingüística en los niveles de competencia del MCER: el caso de la variedad verbal en la expresión escrita de ELE. *Revista Nebrija de Lingüística Aplicada*, 20.
- Del Moral Manzanares, F. (2013). Escalas de descriptores y fiabilidad de la evaluación de la expresión e interacción orales del usuario competente. *Revista Nebrija de Lingüística Aplicada*, 13.
- Del Moral Manzanares, F. (2014). Dificultades en el uso de los descriptores en el proceso de evaluación de la expresión oral de ELE. *Revista Nebrija de Lingüística Aplicada*, 16.
- Ellis, N. C. y Larsen-Freeman, D. (Eds.) (2009). *Language as a complex adaptive system*. Oxford: Wiley-Blackwell.
- Ezeiza, J. y Alberdi, Xabier. (2015). Los marcadores discursivos como rasgo distintivo de los textos del ámbito de las ciencias jurídicas en lengua vasca: Aproximación estadística con proyección jurilingüística. *Revista Signos*, 48(87): 3-29.
- Gutiérrez, R. (2012). La objetivación de los descriptores de evaluación de la expresión oral en la enseñanza de E/LE a alumnos adultos en inmersión lingüística. *RedEle*, 13.
- Herdina, P. y Jessner, U. (2002). *A dynamic model of multilingualism: perspectives of change in psycholinguistics*. Clevedon: Multilingual Matters.
- Housen, A., Kuilen F. y Vedder I. (eds.) (2012). *Dimensions of L2 performance and proficiency: complexity, accuracy and fluency in SLA*. Amsterdam; Philadelphia: John Benjamins Pub.
- Kuiken, F. y Vedder, I. (2012). Syntactic complexity, lexical variation and accuracy as a function of task complexity and proficiency level in L2 writing and speaking. En Housen, A., Kuilen F. y Vedder I. (eds.) *Dimensions of L2 performance and proficiency: complexity, accuracy and fluency in SLA*. Amsterdam; Philadelphia: John Benjamins Pub. Co.: 143-170.
- Lantof, J. y Phoener, M. (2004). Dynamic assessment of L2 development: bringing the past into the future. *Journal of Applied Linguistics* Vol. 1.1.: 49-72.
- Larsen-Freeman, D. (2012). Complex, dynamic systems: A new transdisciplinary theme for Applied Linguistics? *Language Teaching* 45.2: 202-214.
- Larsen-Freeman, D. y Cameron, L. (2008). *Complex systems and Applied Linguistics*. Oxford: Oxford University Press.
- Levinka, M. y Gilabert, R (2012). The effects of cognitive task complexity on L2 oral production. Housen, A., Kuilen F. y Vedder I. (eds.) *Dimensions of L2 performance and proficiency: complexity, accuracy and fluency in SLA*. Amsterdam; Philadelphia: John Benjamins Pub. Co.: 143-170.
- Mavrou, I. (2013). Precisión lingüística y complejidad sintáctica: ¿hasta qué punto compiten entre sí por los recursos limitados del aprendiz? *Revista Nebrija de Lingüística Aplicada*, 14.
- Mavrou, I. (2015). Análisis factorial exploratorio: cuestiones conceptuales y metodológicas. *Revista Nebrija de Lingüística Aplicada*, 19.
- Mislevy, R. J. (2009). If Language Is a Complex Adaptive System, What Is Language Assessment? *Language Learning* 59:Suppl.: 249-267.
- Poehner, M. E. (2008). *Dynamic assessment: a Vygotskian approach to understanding and promoting L2 development*. Pensilvania: Springer.
- Robinson P. (ed.) (2011). *Second language task complexity : researching the cognition hypothesis of language learning and performance*

. Amsterdam; Philadelphia: John Benjamins Pub. Co.

- Rodríguez Sánchez, I. (2012). La Teoría de los Sistemas Dinámicos como marco para el estudio de la adquisición de vocabulario L2. Tres estudios empíricos. En Hoyos, Amalia y Ana Lidia Munguía (coord.) *Interdisciplinas Lingüísticas y Lingüística Aplicada*. Sonora: Universidad de Sonora: 143-160.
- Sampson, G., Gil, D. y Trudgill, P. (eds.) (2009). *Language complexity as an evolving variable*. Oxford: Oxford University Press.
- Schoonen, B. (2013). Estructural Equation Modeling. En Chapelle, C. (ed.), *The Encyclopedia of Applied Linguistics*. Iowa: Blackwell. <http://onlinelibrary.wiley.com/doi/10.1002/9781405198431.wbeal1118/pdf>.
- Swaki, Y. (2012). Structural Equation Modeling in Language Assessment. En Chapelle, C. (ed.), *The Encyclopedia of Applied Linguistics*. Iowa: Blackwell. <http://onlinelibrary.wiley.com/doi/10.1002/9781405198431.wbeal1119/pdf>.
- Verspoor, M. H., De Bot, K. y Lowie, W. (eds.) (2011). *A dynamic approach to second language development: methods and techniques*. Amsterdam; Philadelphia: John Benjamins Pub. Co.
- Winke, P. (2014). Testing Hypotheses about Language Learning Using Structural Equation Modeling. *Annual Review of Applied Linguistics* 34: 102–122.
- Wolfe-Quintero, K., Inagaki, S. y Hae-Young, K. (1998). *Second language development in writing: measures of fluency, accuracy, & complexity*. Manoa: University of Hawai'i.
- Zamorano Aguilar, A. (2012). Teorías del Caos y Lingüística: aproximación Caológica a la comunicación verbal humana. *Revista Signa* 21: 679-705.
- Zheng, Y. (2012). *Dynamic vocabulary development in a foreign language*. Berna: Peter Lang.