

LLORIÁN GONZÁLEZ, SUSANA

UNIVERSIDAD COMPLUTENSE DE MADRID, ESPAÑA

ORCID iD: <http://orcid.org/0000-0002-0750-1620>

CLAVES DE UNA REVISIÓN DE LOS *NIVELES DE REFERENCIA PARA EL ESPAÑOL*, BASADA EN METODOLOGÍA DE CORPUS

BIODATA

Susana Llorián (slloriang@ucm.es). Doctora y máster en Lingüística Aplicada por Universidad Politécnica de Madrid y la Universidad Nebrija, respectivamente. Fue técnico del Área Académica del Instituto Cervantes hasta 2011, donde, entre otros proyectos, estuvo a cargo del SICELE; gestionó varios sistemas de certificación lingüística y diseñó el sistema de formación de examinadores del DELE. Es miembro del Consejo de Redacción del *Plan Curricular del Instituto Cervantes*. Actualmente imparte clases en la Facultad de Filología de la Universidad Complutense de Madrid y en programas de posgrado de otras instituciones; anteriormente, en universidades y otros centros públicos y privados. Es autora de materiales didácticos, artículos y libros.

RESUMEN

Las especificaciones de la lengua que interiorizan los aprendientes usuarios del español en cada uno de los niveles descritos en el *Marco común europeo de referencia (MCER)* se han hecho hasta la fecha utilizando procedimientos intuitivos, en una dirección de arriba abajo: *Un Nivel Umbral*, 1979; *Niveles de referencia para el español*, incluidos en el *Plan Curricular del Instituto Cervantes*, 2007. En la actualidad, es posible hacer este tipo de descripciones a partir de bases empíricas, utilizando metodología de corpus. De esta forma, siguiendo una dirección de abajo arriba, en un proceso de investigación permanente, podrían especificarse los elementos de la lengua que realmente comprenden y producen los aprendientes usuarios cuando alcanzan los niveles de la escala de referencia del *MCER*, que es, sin lugar a dudas, diferente de la que se supone que necesitan aprender. En este artículo se exponen las claves, a partir de las cuales podría materializarse un proyecto que lograra poner al servicio de los responsables curriculares, diseñadores de cursos, autores de materiales, entidades examinadoras e investigadores, un conjunto de recursos que permitiría vincular con garantías los productos del trabajo de estos agentes a los niveles de referencia del *MCER*.

PALABRAS CLAVE: niveles comunes de referencia, corpus de referencia, corpus de aprendientes, *Nivel Umbral*, plan curricular

KEYS FOR A CORPUS-BASED REVISION OF THE *NIVELES DE REFERENCIA PARA EL ESPAÑOL*

The specification of the language that the learners of Spanish have internalized at the levels of the *Common European Framework of Reference (CEFR)* has been done using intuitive procedures, in a top-down approach. Currently, it is possible to make this description empirically, using corpus-based methodology, in a bottom-up process of permanent investigation. The components of the language that learners really understand and produce can be specified when they reach the levels of the *CEFR*. This article

presents the keys of a project that would be able to bring to curricular managers, course designers, authors of didactic materials, examiners and researchers a set of resources that would allow them to link the products of their work to the *CEFR* reference levels.

KEY WORDS: common reference levels, reference corpora, learner corpora, T-Level Series, descriptions validation

1. INTRODUCCIÓN

El comienzo año 2017 coincide con la efeméride de la publicación de los *Niveles de referencia para el español, Plan Curricular del Instituto Cervantes (NRE-PCIC*, o simplemente *NRE*, en adelante). A una década de este acontecimiento, es momento de reflexionar acerca de algunos de los avances que ha experimentado la didáctica de lenguas en el panorama internacional y de plantear la necesidad de que se revise, en virtud de estos progresos, un documento cuyo impacto en el desarrollo de proyectos curriculares, de materiales didácticos y de sistemas de exámenes certificativos es indiscutible.

Las páginas introductorias de los *NRE-PCIC* aluden a los destinatarios de la obra: por un lado, los equipos docentes de la red de los centros de la institución y, por otro, el resto de los profesionales relacionados la enseñanza del español como lengua extranjera (ELE): editores, responsables educativos y autores de materiales didácticos (García Santa-Cecilia, 2007: 11). Por lo tanto, según el propio documento, habría que entender la utilidad de los *NRE-PCIC* en su doble dimensión, que aparece reflejada en la estructura del propio título. En primer lugar, hay que considerarlo como «el desarrollo, en términos de objetivos y contenidos de enseñanza y aprendizaje, de los niveles comunes de referencia (A1-A2, B1-B2 y C1-C2)

establecidos por el *Marco común europeo de referencia para las lenguas: aprendizaje, enseñanza, evaluación*». En segundo lugar, el Instituto Cervantes toma estos objetivos y contenidos «como propios en la actualización de su currículo» (García Santa-Cecilia, 2007: 11). El interés de los *NRE-PCIC* para los profesionales ajenos al Instituto Cervantes –y, por consiguiente, para el contenido de este artículo– radica en que constituyen la continuidad para el español de la especificación de los documentos de la serie *Nivel Umbral (Threshold Level Series)*.

En el año 2007 se abrió la página electrónica del proyecto del *English Profile*¹. Los artífices de esta novedosa iniciativa lo definen como un programa de investigación que se orienta a proporcionar descripciones detalladas de los elementos lingüísticos (gramática, vocabulario, funciones) de los niveles de la escala del *MCER* para la lengua inglesa. En este caso, las organizaciones a cargo son varias: la Universidad de Cambridge, la Universidad de Bedfordshire, Cambridge University Press, Cambridge ESOL, el British Council y English UK. No obstante, la red de las entidades implicadas se extiende a una considerable cantidad de personalidades académicas, responsables gubernamentales en la materia y profesionales de la educación en todo el mundo. Las instituciones

¹ <http://www.englishprofile.org>

nombradas cuentan con el soporte de la Comisión Europea, que incardina el proyecto en el marco del *Programa de Aprendizaje Permanente (Long Life Learning Program)*. Los trabajos de investigación propiamente dichos se inician en el año 2009 y ven los primeros resultados, en lo que a recursos lingüísticos se refiere, en el año 2012: *English Vocabulary Profile*², *English Grammar Profile*³, disponibles en línea. Este material se acompaña de un extenso cuerpo de publicaciones⁴, algunas de las cuales serán objeto de mención a lo largo de este artículo.

La diferencia más importante entre las descripciones del material lingüístico que se derivan del *English Profile* y las que se han desarrollado para el resto de las lenguas nacionales o regionales⁵, incluidas las de los *NRE-PCIC* para el español, estriba en que las primeras se han elaborado a partir de la aplicación de metodología de corpus. Esto supone la dotación, por primera vez en la historia, de bases empíricas de gran solidez a los repertorios de material lingüístico que figuran en este tipo de descripciones. El trabajo está fundamentado en el enfoque sociocognitivo desarrollado por el profesor C. Weir (2005a), a la cabeza del proyecto.

Tradicionalmente, las especificaciones de los documentos de la serie *Nivel Umbral (T-Level Series o T-Series)* se vienen desarrollando por parte de equipos de expertos, de forma intuitiva. En el caso de los *NRE-PCIC*, se alude a la participación de 234 profesores (*NRE-PCIC*, 2007: 49) en la revisión del material. No está claro si esta revisión responde a un proceso sistemático de validación o si este trabajo, que no se explica en qué consiste, se realiza de forma asistemática.

² <http://www.englishprofile.org/wordlists>

³ <http://www.englishprofile.org/english-grammar-profile>

⁴ <http://www.englishprofile.org/resources>

⁵ http://www.coe.int/t/dg4/linguistic/DNR_EN.asp

El Instituto Cervantes, al margen de esta breve alusión, no ha publicado información acerca de la metodología empleada ni de los resultados. Tampoco se han difundido trabajos de investigación en que se refleje la implementación de las especificaciones de los *NRE-PCIC* en programas de curso con los consiguientes resultados de una evaluación. Estos vendrían a aportar evidencias a partir de las cuales se podría demostrar en qué grado este material se corresponde realmente con los niveles de referencia, tal y como estos aparecen descritos en el *Marco común europeo de referencia para las lenguas: aprendizaje, enseñanza y evaluación (MCER, Marco o Marco de referencia, a partir de aquí)*

El español continúa, por lo tanto, con una asignatura pendiente: la especificación del material lingüístico que corresponde realmente a los niveles comunes de referencia, descritos en el *MCER*. Esta tarea puede hacerse siguiendo dos posibles caminos, que no son excluyentes. El primero consiste en validar y ajustar las especificaciones previamente realizadas a partir de la evaluación de los programas derivados de la implementación de los *NRE-PCIC*. El segundo, en reemprender la labor de especificación del material lingüístico, esta vez sobre bases empíricas, empleando para ello metodología de corpus.

A lo largo de este artículo, después de hacer un recorrido por los antecedentes históricos de las descripciones de los documentos correspondientes a las series *T-Level* en español y otras lenguas, plantearemos una serie de claves sobre las cuales podría realizarse la revisión de los *NRE*. Esta revisión, tal y como se viene anunciando desde el propio título, habrá de hacerse aplicando metodología de corpus informatizados, en combinación con otros procedimientos. Haremos un repaso de los corpus de aprendientes accesibles a través de Internet en la actualidad y de los problemas que plantean para

realizar esta tarea y, en consecuencia, de los que están pendientes de construcción y de desarrollo.

2. ANTECEDENTES Y EVOLUCIÓN

El rastreo de los orígenes de las descripciones de los documentos de la serie *Nivel Umbral (T-Level series)* obliga a retrotraerse a los comienzos de la década de los años setenta del pasado siglo, en que irrumpen en el panorama de las lenguas extranjeras y segundas, bajo los auspicios del Consejo de Europa, un documento cuyo impacto solamente será superado por el *MCER* tres décadas más tarde: el *Threshold Level (Nivel Umbral)*. La perspectiva actual deja pocas dudas acerca de que esta publicación hiciera posible, a este lado del Atlántico, la implantación en la práctica de los primeros desarrollos de los enfoques comunicativos. Esto sucedió gracias a que incorporaba unidades programáticas, las funciones comunicativas y las nociones generales y específicas, que permitían canalizar hacia las aulas los presupuestos aportados por las disciplinas que confluyen en este importante cambio de paradigma. Sánchez (1997: 186) encuentra los antecedentes inmediatos en el conocido como proyecto *Número 4* (1971-1973), dentro del marco del programa de *Lenguas Vivas*.

Las novedades que se introducen en el planteamiento de los procesos de diseño curricular favorecen el despegue hacia la era actual de la didáctica de las lenguas extranjeras o segundas, al alejarse parcialmente los presupuestos conductistas y adentrarse en los constructivistas. Se establece como punto de partida las necesidades de los destinatarios del currículo, es decir, los alumnos; los objetivos se formulan a partir de las unidades de análisis definidas y acuñadas anteriormente por Wilkins (1972); los objetivos

se seleccionan, por primera vez, y como consecuencia de lo anterior, con un criterio funcional (Green, 2012).

En el planteamiento de este proyecto subyace ya la idea de «unidades crédito», que busca la compatibilidad y la posibilidad de que los diferentes sistemas de enseñanza de idiomas en el continente europeo se armonicen y se homologuen. Por consiguiente, este, el objetivo fundamental del *Marco de referencia*, no emerge sin más en el año 1991, cuando se trazan las líneas maestras para su desarrollo, sino que está latente desde la década de los años setenta (Hulstijn *et al.*, 2010).

El programa de *Lenguas Vivas* toma cuerpo, tres años más tarde, en 1976, cuando confluyen dos publicaciones, que serán decisivas para la didáctica de lenguas en las cuatro décadas que suceden a esta fecha. El profesor D. Wilkins recoge los presupuestos teóricos, por un lado, en *Notional Syllabuses*; por otro, van Ek se ocupa de la especificación práctica en el *Threshold Level for Modern Language Learning in Schools*.

John Trim (2010) narra en un apasionante relato cómo se fragua el proceso que aquí desemboca, iniciado en 1971. En este mismo año, él mismo y David Wilkins pergeñan, durante la celebración de un congreso en la localidad de Rüşkilon (Suiza), las bases sobre las cuales pudiera construirse una enseñanza de lenguas modernas para adultos en la cambiante sociedad de la época, en la cual se multiplican los desplazamientos transfronterizos. Se convencen de la idoneidad, en el debate que mantienen, de crear un sistema de unidades de crédito que pudieran hacerse operativas en cursos modulares, con objetivos formulados a corto plazo, lo cual se ajusta a las circunstancias en las que aprenden normalmente los adultos.

Se proponen como solución establecer los puntos de partida para definir los objetivos de estos cursos en las situaciones y los diálogos contextualizados que en estas se dieran, pero encuentran una serie de inconvenientes derivados de la imposibilidad de concretar y generalizar las mencionadas situaciones, que acaban por revelarse como un concepto poco operativo para este fin. Contemplan asimismo la opción de establecer como base un escalado de enunciados a modo de *can do statements* o estándares de logro en varios niveles, en un sistema similar al actual *Marco de referencia*. Lo descartan igualmente debido a los problemas que plantea la segmentación totalmente arbitraria del continuo en la progresión del proceso de apropiación de una lengua y a la dificultad que entraña la redacción de estos enunciados (Trim, 2010: xxvii).

No desechan, sin embargo, la idea de 'nivel' implícita en toda esta serie de esbozos, pese a los problemas que reporta su definición y consideración. Reparar en que, durante las primeras fases del proceso de apropiación de las lenguas extranjeras, los usuarios disponen de una capacidad y unos recursos limitados para usarla en las situaciones reales de comunicación. Sin embargo, intuyen un «umbral» a partir del cual la escasa competencia comunicativa que han desarrollado los capacita para hacer frente a las necesidades básicas de la vida cotidiana. Trim recuerda que, en un proyecto que se desarrolla en el seno del Consejo de Europa, estas necesidades no pueden limitarse a satisfacer meras transacciones que impliquen procurarse bienes y servicios, sino que tienen que orientarse hacia el entendimiento mutuo entre personas de diferentes culturas y a desarrollar en el usuario de la conciencia y la confianza de su propia capacidad de aprender.

Los criterios para describir este «nivel umbral» con las nuevas unidades de análisis que D. Wilkins y J. Trim van esbozando, las

nociones y las funciones, se desprenderán de las fuentes que nutren la mítica biblioteca del profesor Trim, en un momento en el que los progresos en la lingüística teórica se encuentran en plena ebullición, gracias a los trabajos de Firth (1957), Austin (1962), Searle (1969), entre otros. También influyen decisivamente, según confiesa Trim, los libros de las «frases más útiles en una lengua», que se comienzan a comercializar en aquella época, destinados a turistas y a trabajadores migrantes.

Van Ek será el encargado de dar forma a todo este conglomerado de ideas que van fraguando a lo largo de los tres días que dura el congreso. El resultado es un trabajo, finalizado en 1975, que antecede a los documentos más arriba mencionados: *The Threshold Level in an European unit / credit system for Modern language learning by adults* (Trim, 2010: xviii). Los componentes, presentados en forma de listas, son los siguientes: situaciones de uso de la lengua, temas, actividades comunicativas, funciones comunicativas, descripción del comportamiento del usuario en relación con cada uno de los temas, nociones generales, nociones específicas, formas lingüísticas (exponentes), grado de destreza con el que los usuarios ejecutarían las funciones y nociones.

El enfoque, en coherencia con los desarrollos de las disciplinas lingüísticas del momento, parte de un análisis de las intenciones comunicativas, las funciones; los significados conceptuales, las nociones generales; los temas de comunicación, las nociones específicas. Intenta así dar respuesta al dilema que planteaba considerar una gran diversidad de factores situacionales, contextuales y extralingüísticos, a la hora de establecer una organización de los contenidos de la enseñanza en los programas derivados del planteamiento. El problema de encontrar una fórmula que permita establecer un equilibrio coherente entre las

dimensiones comunicativa y formal constituye, todavía hoy, un reto difícil de afrontar tanto en el plano teórico como en el práctico. Las funciones (*solicitar acciones del interlocutor, presentarse y ser presentado, expresar tristeza, describir algo*, etc.), las nociones generales (*existencia, frecuencia, distancia, tamaño, pensamiento*, etc.) y las nociones específicas relacionadas con temas (*tiempo atmosférico, comidas y bebidas, relaciones personales*, etc.), tal y como las concibe Wilkins (1972, 1976), son unidades cuantificables, combinables y acumulables, que se hacen corresponder con las porciones de necesidades de los alumnos en las situaciones meta de uso de la lengua, es decir, aquellas en las que el alumno se verá en la necesidad de usar la lengua objeto como vehículo de comunicación. El *Threshold Level*, en la versión de van Ek de 1976, contiene un repertorio de funciones comunicativas, una lista de nociones específicas y otra de nociones generales, además de un elenco de las estructuras que se derivan de las funciones comunicativas. Todos estos elementos se acompañan de formas lingüísticas ilustrativas, que, en español, se acuñan como 'exponentes'. Esta versión será, a lo largo de los 15 años sucesivos, adaptada a 25 lenguas.

Es necesario tener presente la observación de Sánchez (1997: 191) de que este documento no tiene naturaleza curricular, puesto que no especifica los *medios*, es decir, los procedimientos metodológicos que se precisan aplicar en su implementación en las aulas. Uno de los primeros programas de enseñanza y aprendizaje a los cuales se aplica el *Nivel Umbral* como sílabo es el conocido curso multimedia *Follow Me!* Su desarrollo en Austria permite constatar que, en este contexto, la mayor parte de los alumnos adultos no es capaz de llegar al nivel al cabo de un año de instrucción, por lo que se requiere la definición de objetivos de menor alcance. Estamos ante el germen del conocido como *Waystage Level* (Trim, 2010).

Los documentos que contienen las descripciones de los dos niveles, *Thresholdy Waystage*, experimentan una revisión en el año 1990. Al año siguiente, Cambridge University Press publica las versiones, en las que se añaden algunos elementos que no presentan las de las dos décadas precedentes. Estas contienen una definición de objetivos, redactados en dos grados de especificidad. Los objetivos de la lista más específica se clasifican en tres grandes grupos: el primero recoge enunciados en términos de las situaciones meta de uso de la lengua centradas en el intercambio de bienes y servicios de carácter básico, en interacciones sociales más o menos ritualizadas y en el procesamiento de textos; el segundo grupo, más reducido, se orienta al desarrollo de la capacidad para interactuar con miembros de otras culturas; el tercero, con extensión similar al anterior, establece objetivos relacionados con la capacidad de aprender. De aquí en adelante se hablará de documentos de la serie *T-Level*.

Es preciso tener en cuenta que, en estas primeras fases, los documentos de la serie *T-Level* no se destinan a la evaluación. Esto explica el criterio que se sigue para la redacción de los textos de los objetivos (p. e., *Learners should be able to book accommodation by letter or telephone*, van Ek y Trim, 1990: 13) y la especificación de los elementos (funciones, nociones, etc.). Se trata, en el primer caso, de enunciados muy breves y sencillos, de fácil manejo, en los cuales no se hace mención al grado en que se espera que se acometan ni a las condiciones en que se realizan las acciones a las que aluden. Según Trim (2010), es el profesor Oscarson el primero que, en la década de los años ochenta, se comienza a preocupar de todas estas cuestiones y empieza a proponer enunciados más complejos en forma de *can do statements*, que se puedan usar en la evaluación.

Los tradicionales repertorios de funciones y de nociones se completan en esta nueva versión con un grupo de inventarios

centrados en las actividades comunicativas de la lengua. El primero es un elenco de patrones de interacción oral, que viene a ser un esbozo de los turnos que se dan en las conversaciones transaccionales más estereotipadas. El segundo se centra en la expresión e interacción escritas y en él se especifican los géneros a los que se asocian los textos que producen los usuarios, las operaciones que realizan y las estrategias que emplean. El último se refiere al procesamiento de textos y, en la línea de los anteriores, se hace mención en él a los tipos de texto que se procesan, a las operaciones cognitivas implicadas en la comprensión y a las estrategias. La lista que figura después recoge los tipos de conocimiento declarativo que se precisa para desarrollar la competencia sociocultural. A este sucede una relación de estrategias de compensación y otra de estrategias metacognitivas. En el último capítulo se describen los rasgos que definen el grado de dominio en cada uno de los dos niveles. Los documentos se cierran con apéndices que recogen especificaciones de gramática y de patrones entonativos.

Bien entrada la década de los años ochenta, surge el reto de hacer frente a las necesidades de comunicación más complejas que aquellas a las que da respuesta el *Umbral*. Esto hace que se comience a retomar y a concebir el concepto de 'nivel' en términos algo diferentes, que serán los que guíen a Trim y a van Ek a formular la base sobre la que se construiría el nivel inmediatamente superior, el *Vantage Level*, cuya progresión desde el *Umbral* se plantearía sobre fundamentos cualitativos. Esta maduración de la controvertida idea de 'nivel', en la que nos extendemos más abajo, pone asimismo sobre la mesa la posibilidad de desarrollar un estadio inferior al *Waystage Level*, que, en un periodo de instrucción muy reducido, dotase a las poblaciones migrantes de los más mínimos recursos que

se pudieran considerar como un primer paso hacia el dominio de la lengua: el *Breakthrough Level* (Trim, 2010: xxxii).

Durante los últimos veinte años del siglo XX se experimenta en Europa un incremento exponencial de la movilidad de los ciudadanos, impulsada por motivos turísticos, laborales y de estudios. En este escenario, la necesidad de mostrar y demostrar ante terceros el grado de dominio en las lenguas de los países receptores dispara la consiguiente demanda de armonizar los sistemas de aprendizaje y enseñanza de lenguas en el continente. Este fenómeno pone en valor el trabajo previo del Departamento de Política Lingüística del Consejo de Europa y conduce a la celebración, 20 años más tarde, de un segundo simposio en la misma localidad de Rüşkilon (Suiza), de cuyas conclusiones se desprende el proyecto de desarrollar el *Marco común europeo de referencia para las lenguas: aprendizaje, enseñanza y evaluación*. La versión definitiva verá la luz diez años después, a pesar de que los primeros borradores circularon entre los expertos en la materia designados por la organización promotora a partir de la mitad de la década de los años noventa del pasado siglo (Llorián, 2007: 26).

Las descripciones de los niveles comunes de referencia posteriores a la aparición del *MCER* van a verse condicionadas, en las diferentes lenguas nacionales y regionales, de un lado, por la dimensión vertical del documento y, de otro, por la horizontal. La dimensión vertical del *MCER* establece las tres grandes etapas en la progresión del grado de dominio, A, B, C, que, al bifurcarse, dan lugar al conocido sistema de seis niveles comunes de referencia (A1, A2, B1...). En cuanto a la horizontal, el esquema descriptivo del *MCER* establece las categorías y los parámetros (competencias, estrategias, actividades comunicativas de la lengua, textos, contextos, ámbitos, procesos y tareas) que servirán de base para determinación de la

naturaleza de las unidades a partir de las cuales se describirá el material lingüístico o no lingüístico que el usuario habrá interiorizado para alcanzar el nivel descrito en cada caso.

El Consejo de Europa decide establecer una línea de continuidad con los trabajos anteriores a la publicación del *MCER* en cuanto a los documentos de la serie *Nivel Umbral* (*Threshold, Waystage, Vantage*). En el año 2005, el Consejo difunde una guía con la cual pretendía garantizar los requisitos mínimos de transparencia y coherencia en las descripciones de los niveles de referencia para las lenguas nacionales y regionales posteriores al *Marco de referencia: Reference Level Descriptions for National and Regional (RLD) Languages. Guide for de production of LRD*. En ella establecía las directrices siguientes: unos rasgos comunes y metodología de desarrollo para las descripciones de todas las lenguas. Dentro de los rasgos comunes, fijaba unos principios generales, unos elementos mínimos obligatorios y otros, deseables u opcionales. Entre los obligatorios, por ejemplo, se considera especificación de las funciones y de las nociones.

3. LOS NIVELES DE REFERENCIA PARA EL ESPAÑOL Y EL CURRÍCULO DE ESPAÑOL, LENGUA EXTRANJERA

3.1. LOS PRIMEROS DESARROLLOS: UN NIVEL UMBRAL

En el año 1979, el Consejo de Europa difunde la versión española del *Threshold Level, Nivel Umbral*. El artífice es el profesor Peter Slagter de la Universidad de Utrecht (Países Bajos). Para Sánchez (1997) constituye, a diferencia de lo que se hizo en otras lenguas, una mera traslación de la versión inglesa. El propio Slagter reconoce

este extremo, en el ejercicio de modestia que refleja en la introducción a la versión del documento que se recupera para la *Biblioteca Histórica* de la revista *marcoELE*. Aduce razones restrictivas que se le imponen desde las altas instancias.

A finales de la década de los años ochenta y comienzos de la siguiente del pasado siglo, se publican en España una serie manuales de enseñanza de ELE, cuyos sílabos, tal y como reconoce el propio Slagter y reivindica lícitamente en varias conferencias y clases Lourdes Miquel, van a constituirse en el auténtico *Nivel Umbral* para el español. Nos referimos a la serie *Para empezar* (Equipo Pragma, 1988) o *Esto funciona* (Equipo Pragma, 1989). El proceso de reflexión subyacente se consolida en la serie *Intercambio* y sus complementos (Miquel y Sans, 1989). No desmerece en calidad la contribución de un material que se publica unos años más tarde: *ELE 1* (Borobio, 1994). Estos manuales utilizan un modelo de secuencia didáctica conocida como PPP (Presentación, Práctica intensiva y Producción), que se presta al aprendizaje de los exponentes de las funciones y de las nociones. Al margen de estas propuestas, se puede decir que la descripción nociofuncional del español culmina con la *Gramática comunicativa del español* de Francisco Matte Bon (1992). Se trata de un trabajo titánico, de gran precisión y calidad, realizado además en solitario.

3.2. LOS NIVELES DE REFERENCIA PARA EL ESPAÑOL, PLAN CURRICULAR DEL INSTITUTO CERVANTES

El Instituto Cervantes se hace eco inmediatamente de la irrupción del *Marco de referencia* en el panorama de la didáctica de lenguas extranjeras. Asume la tarea de su traducción al español, que será una de las más valoradas por parte del Consejo de Europa.

La organización contaba, desde 1994, con un documento de primer nivel de adaptación del currículo institucional, mediante el cual se regulaba la actividad docente de los centros de su red. Este había trascendido los límites de la institución y había sido reclamado durante años por otras instancias hacia las que actuaba como referencia para la programación de cursos y para la elaboración de sílabos materiales de enseñanza. El *Plan Curricular del Instituto Cervantes* de 1994 presentaba una estructura vertical en cuatro niveles, *Inicial, Intermedio, Avanzado y Superior*. Definía los cuatro fines generales del área académica de la institución, a modo de declaración de intenciones que situaba la docencia en perfecta sintonía con la política lingüística del Consejo de Europa. Las especificaciones de los componentes curriculares, objetivos, contenidos, orientaciones metodológicas y evaluación, eran muy generales. No obstante, presentaban para la época una sorprendente modernidad, en especial, en lo tocante al desarrollo de la autonomía y al tratamiento del componente intercultural. Apuntaban claramente al desarrollo de una competencia plurilingüe, a pesar de que el concepto de plurilingüismo, tal y como se describe en el *MCER*, no se encontrara aún del todo depurado.

No es la única institución que adapta, en el ámbito del español, las especificaciones de su documentación curricular al *MCER*. Las Escuelas Oficiales de Idiomas, por ejemplo, sin modificar la estructura vertical de sus currículos, ajustan sus componentes a los parámetros y categorías del esquema descriptivo del *Marco de referencia* y orientan los fines de la enseñanza al desarrollo de una competencia plurilingüe (Fernández, S. 2011, 2003). Sin embargo, la especificación de los objetivos y de los contenidos no es tan detallada en este caso. Los componentes léxico, semántico, gramatical o pragmático discursivo no se ciñen de manera tan rígida

al paradigma nociofuncional, sino que la especificación, que queda más abierta, es más propia de un documento curricular, susceptible de ser adaptado a lo largo de sucesivos niveles de concreción. Por lo demás, en estos documentos aparecen explícitos los apartados correspondientes a la metodología y la evaluación.

Parece claro que, para definir los objetivos de aprendizaje en un documento curricular de primer o segundo nivel de concreción, es útil contar con repertorios de referencia como los que figuran en los documentos de la serie *T-Level*. Lo que no está tan claro es que estos sean objeto de integración en un documento curricular que, por lo demás, no especifica los *medios*, es decir, las orientaciones metodológicas y tampoco los procedimientos de evaluación. Esto es así porque las descripciones de la serie *Nivel Umbral* no son, como hemos dicho, de naturaleza curricular, al menos, no pueden serlo desde la perspectiva de un currículo que se incardine en la corriente constructivista, abierto y centrado en el alumno. En un documento curricular no se realizan descripciones de la lengua que utilizan los usuarios, sino la que deberían aprender lo largo de los estadios que establece el currículo en los procesos de aprendizaje.

Los *Niveles de referencia para el español*, insertos en el *Plan Curricular del Instituto Cervantes*, publicado a comienzos del año 2007, presentan una lista de objetivos generales y doce inventarios de contenidos. El conjunto apela al desarrollo de las competencias generales y comunicativas de la lengua del *MCER*.

Los objetivos generales representan un tratamiento de las competencias desde la perspectiva del alumno (*NRE-PCIC*, 2007: 33). Su desarrollo se prevé en tres perfiles o facetas del sujeto destinatario, que están en consonancia con los tres grupos de objetivos que figuraban en las series tradicionales de documentos *T-*

Level, a los que hicimos mención más arriba. Aquí se denominan perfil de *agente social*, de *hablante intercultural* y de *aprendiente autónomo*. Como hemos dicho, constituyen una nueva forma de traslación de los fines institucionales definidos en el documento curricular de 1994 hacia la actividad práctica de las aulas. La progresión del perfil del *agente social* a lo largo de la estructura vertical del currículo se prevé en tres etapas, que se corresponden con los grandes niveles del *MCER*, A, B y C. Los objetivos no se concretan en la serie sucesiva, A1, A2; B1... La de los otros perfiles, *hablante intercultural* y *aprendiente autónomo*, orientados al desarrollo de las competencias generales del *Marco de referencia*, en tres fases: *aproximación, profundización y consolidación*.

La redacción del texto de los objetivos del perfil del *agente social* difiere de la que se emplea en los documentos de la serie *T-Level*, en la medida en que estos no están focalizados en contextos de uso de la lengua (*comer en un restaurante, preguntar por una dirección en la calle, etc.*), sino que están planteados de manera más abstracta. Se asimilan, en cuanto a su formulación, al modelo de los *behavioural objectives* (Green, 2012: 40-43). Sin embargo, no se pueden hacer operativos en una dimensión programática si no se realiza una adaptación ulterior en la que se concreten, sobre todo, en relación con los seis niveles comunes de referencia, más allá de las etapas A, B y C. Son demasiado generales como para que permitan prever los medios (tareas, actividades, etc.), a partir de los cuales se pueden alcanzar. El documento debería haber aportado este nivel de especificación o, en su defecto, indicar las líneas que se deben seguir para redactarlos.

La concreción de los objetivos de los perfiles de proceso, *hablante intercultural* y *aprendiente autónomo*, requiere conocimientos y destrezas avanzadas en programación. No se pueden interpretar

como objetivos de producto, en la línea de los anteriores. Tampoco se puede incurrir en el absurdo, al intentar analizarlos buscando en los textos con los que se formulan verbos de acciones observables, en línea con las taxonomías de Bloom, un nivel de logro y unas condiciones de desempeño. Si a esto se añade la dificultad que entraña el manejo de las tres fases de desarrollo, se entiende que, sin una formación adecuada de profesores, autores de materiales, incluso de formadores, este tipo de especificaciones no lleguen ponerse en práctica en las aulas.

Los doce inventarios restantes de los *NRE* se interpretan en el *PCIC* en términos de contenidos. Se trataría de la dimensión del objeto, tal y como se indica en la Introducción (*NRE*, 2007: 31), o lo que es lo mismo, del material lingüístico o no lingüístico que debería interiorizar el sujeto en su contexto de aprendizaje y uso de la lengua para desarrollar las competencias generales y las competencias comunicativas de la lengua en el grado descrito en el *MCER* para cada uno de los niveles comunes de referencia.

Este material se presenta especificado de manera taxonómica, en una serie de listas de elementos discretos. Dentro del paradigma lingüístico, se encuentran lo siguiente: el componente gramatical, que abarca los inventarios de «Gramática», «Ortografía», «Pronunciación y prosodia» y que se orienta al desarrollo de las competencias gramatical, ortoépica, ortográfica y fonológica; el componente léxico semántico, que se materializa en los repertorios de «Nociones generales» y de «Nociones específicas» y se relaciona con las competencias gramatical, en el primer caso, léxica y semántica, en ambos; el componente pragmático discursivo, que abarca los inventarios de «Funciones comunicativas», «Tácticas y estrategias pragmáticas» y «Géneros discursivos y productos textuales» y se orientaría al desarrollo de las competencias

pragmáticas y, junto con el anterior, a la sociolingüística. El resto de las listas se relacionan con las competencias generales. Por un lado, el conocimiento sociocultural, la conciencia intercultural y las destrezas interculturales, junto con algunas habilidades prácticas, se verían cubiertos con los tres inventarios que se integran en el componente cultural. El elenco de «Procedimientos de aprendizaje», finalmente, se relaciona con la enseñanza explícita de estrategias; se orienta, por lo tanto, al desarrollo de la capacidad de aprender. De esta forma, con los *NRE-PCIC*, se ven cubiertos todos los rasgos, tanto los mínimos como los opcionales, señalados en la *Guía* del Consejo de Europa, 2005 (García Santa-Cecilia, 2007: 24-25).

Los *NRE-PCIC* no incluyen especificaciones relativas a las actividades comunicativas de la lengua y a las estrategias de comunicación. No contemplan las características de los textos o géneros que los alumnos producen, comprenden o actúan como soporte de la mediación en cada nivel. Tampoco, a las operaciones cognitivas implícitas en las actividades de comprensión, interacción, expresión o mediación. Es cierta la afirmación del propio documento de que constituyen una de las descripciones más completas de los documentos *T-Series* en la fase posterior al *MCER* (García Santa-Cecilia, Introducción General, 2007), aunque no tan alabada como la que se realiza para el alemán (Trim, 2010: xi), por ejemplo. La dimensión de la actuación lingüística comunicativa no figura tan desarrollada en los *NRE-PCIC* como sería deseable y como para que se puedan programar tareas de la «vida real» y pedagógicas en contextos significativos para los alumnos destinatarios del currículo.

4. LAS INNOVACIONES DEL *ENGLISH PROFILE PROGRAM*

4.1. DESCRIPCIÓN GENERAL DEL *ENGLISH PROFILE PROGRAM (EPP)*

El español no ha sido la única lengua que, en la fase posterior a la difusión del *MCER*, hace descripciones de los niveles comunes de referencia siguiendo la tradición de los documentos de la serie *T-Level* incorporando nuevos aspectos presentes en el *Marco de referencia*. Estas descripciones reflejan de un modo desigual, en las diferentes lenguas nacionales o regionales, los «principios comunes» y los rasgos «mínimos» u «opcionales» de la *Guía* del Consejo de Europa del año 2005. La lengua cuyas descripciones más se hacen esperar en esta fase *postmilenum* y la que más expectativas despierta es el inglés, siempre pionera e innovadora. Más arriba hicimos mención a un proyecto, el *English Profile Program (EPP)*, en lo sucesivo), que tiene como producto la especificación de los elementos léxicos, gramaticales y funcionales que los usuarios de las lenguas producen en cada uno de los niveles de referencia del *MCER*.

Se consideró que una nueva descripción de los niveles de referencia del inglés, posterior al *MCER*, en la línea tradicional seguida hasta la fecha, no superaría los problemas generados por el *Marco de referencia*. Los textos de los descriptores ilustrativos de las escalas del *MCER* se revelaban a la luz de los análisis sistemáticos (Weir, 2005b; Alderson *et al.*, 2006; Fulcher, 2010, entre otros) como inconsistentes, vagos y ambiguos, de lo cual nos hicimos ya eco de forma detallada en otro lugar (Llorián, 2012). No aportaban información suficiente y relevante relativa a la gramática, las nociones, las funciones o cualesquiera que fueran los criterios empleados para describir e inventariar el material lingüístico.

Por lo demás, las metodologías utilizadas hasta el momento para la descripción de los elementos en los documentos de la serie *T-Level* eran de tipo *top-down*, es decir, se encargaba a un grupo de expertos una interpretación intuitiva acerca de los ítems léxicos, gramaticales funcionales, etc., propios de cada nivel (Green, 2012: 2). La falta de evidencias que sustentase estas intuiciones creaba controversia y se había quedado obsoleta para un momento como el actual en que imperan las decisiones tomadas sobre bases empíricas.

Como contrapartida a todo esto, se proyectó el *EPP*. Se trata, como se indicó más arriba, de un programa de investigación colaborativo y multidisciplinar, diseñado para especificar en detalle lo que significan los niveles de referencia, descritos en el *MCER*, en términos del material lingüístico que utilizan –y no que necesitan utilizar– los usuarios de la lengua inglesa cuando alcanzan cada uno de los estadios de la escala de niveles comunes de referencia del *MCER*. La metodología que emplean es, al contrario de lo que marca la tradición, preferentemente de tipo *bottom-up*, o sea, parte del estudio directo de la lengua de los usuarios de cada nivel. Para ello es preciso disponer de muestras reales de uso, de hablantes de diferentes lenguas, en cantidad considerable y procesarlas con metodología de corpus. Esta se combina con otros procedimientos sistemáticos como, por ejemplo, la metodología de juicio experto, la encuesta, etc.

El proyecto se concibe como un proceso de investigación permanente, que va a permitir responder, entre otras, a preguntas como cuál es la influencia en el desarrollo del grado de dominio en la interiorización de las formas lingüísticas inventariadas, de las variables que resultan determinantes para el aprendizaje: la(s) lengua(s) materna(s) y de otras lenguas, los procesos de aprendizaje,

el tipo de *input* recibido por los informantes, las tareas que realizan, etc.

Lo primero que hace el *EPP* es definir nuevos descriptores ilustrativos, *can do statements*, a modo de objetivos para los niveles C, puesto que se considera que son los que peor definidos se encuentran en el *MCER*. Se procede de un modo distinto, sin perder del todo de vista la metodología utilizada por North dos décadas antes. Se compilan descriptores y objetivos de estos dos niveles en diferentes contextos, recurriendo a fuentes diversas (material didáctico, sílabos, etc.), se introducen en bases de datos y se procesan con las herramientas empleadas por los gestores de corpus informatizados, fundamentalmente *Key Word in Context* (palabras clave en contexto). Se extraen de los textos de los objetivos los rasgos siguientes, en relación con cada uno de los niveles comunes de referencia, en consonancia con los tradicionales *behavioural objectives* (Green, 2012: 40-45): la actividad que se espera que realice el alumno, el tema, el texto de entrada, el texto de salida, el nivel de logro, las condiciones y restricciones. Se ajusta la redacción de los *can do statements* para cada nivel (C1 y C2) y, a partir de aquí, se replica la metodología empleada por B. North para redactar los descriptores del *MCER*, encuestando a 700 informantes en todo el mundo.

En lo concerniente a los inventarios de contenidos o de formas lingüísticas, el *EPP* se limita a especificar los elementos constituyentes del componente gramatical, al léxico y a los exponentes de las funciones comunicativas. Dispone los materiales resultantes en un soporte electrónico, en Internet, desde el que son accesibles a través de diversos criterios de búsqueda. El principio que rige la elaboración de estos inventarios es el denominado «rasgo criterial» o «rasgo de criterio» (*criterial feature*). Los procedimientos

de extracción de los rasgos criteriosales de cada nivel varían de un inventario a otro y se basan en metodología de corpus y en procedimientos de la lingüística computacional.

4.2. LOS RASGOS CRITERIALES

Son las propiedades léxicas y gramaticales, indicadoras y definitivas de cada uno de los niveles de referencia de la escala del *MCER*. En español, a falta de suficientes datos avalados por la investigación, y siempre como ejemplo hipotético, podríamos señalar que el uso del pretérito perfecto sería uno de los rasgos criteriosales del nivel A2. Esto significa, entre otras cosas, que no aflora en las producciones de los aprendientes hasta que estos alcanzan este nivel. Lo que define un nivel en cada uno de los componentes (gramática, léxico, etc.) es un haz de rasgos criteriosales.

Tras el concepto de rasgo criterial subyace la idea de que, con independencia de que los usuarios y los candidatos a examen resuelvan una tarea, hay características léxicas y gramaticales, que son propias de cada nivel (Harrison y Barker, 2015: 6). Esto implica que las producen todos los informantes que lo alcanzan, al margen de variables como lengua materna, las tareas, etc. Las propiedades se refieren a la gramática, al léxico, a la fonética y a la correspondencia forma-función, es decir, a los exponentes funcionales. Los rasgos léxicos son las unidades léxicas y los significados o acepciones concretos, con los que estas unidades léxicas se empiezan a utilizar en cada nivel. Para la extracción de los rasgos criteriosales del componente gramatical (*English Grammar Profile*), se utilizan procedimientos tanto inductivos como deductivos. Gracias al *EPP* se evidenciaron múltiples rasgos difícilmente predecibles haciendo uso de metodologías *top-down*.

Las investigaciones de la lingüística teórica, por lo demás, dan lugar a descripciones que funcionan como base de prácticas deductivas.

Los rasgos criteriosales pueden persistir a lo largo de varios niveles. Cuando emergen y marcan la diferencia de un nivel con el inmediatamente anterior, sin llegar a consolidarse hasta que se alcanza, se llaman rasgos criteriosales «transitorios». Según el enfoque del *EPP*, los rasgos criteriosales se clasifican en dos grandes grupos: «positivos» y «negativos». Los primeros evidencian el progreso de los aprendientes y son las propiedades lingüísticas que los usuarios producen correctamente y que, por regla general, persisten en los niveles superiores. La consideración de un rasgo criterial «positivo» depende de la frecuencia de uso, que se corresponde con la distribución normal equivalente a la de un hablante nativo (Howkins y Filipovic, 2012: 33). Los rasgos criteriosales «negativos» son los errores léxicos y gramaticales que, con una determinada frecuencia, se cometen en cada uno de los niveles de la escala del *MCER*. Un error *P* con una frecuencia *F* puede considerarse criterial de un nivel *N* (Howkins y Filipovic, 2012: 25). La distribución frecuencia de los errores varía normalmente de un nivel a otro.

La consideración de los rasgos, o propiedades léxicas o gramaticales, como criteriosales es una cuestión que responde a criterios de cómputo. Cabe preguntarse cuándo se consideran realmente criteriosales los rasgos identificados y cuántas ocurrencias o frecuencias se precisan para considerarse criteriosales de un nivel o del contiguo. Para resolver este importante dilema se estableció como regla la proporción de una ratio de 10 a 1 (Howkins y Filipovic, 2012: 37). Esta se resuelve del modo siguiente: dados dos niveles adyacentes, si la cantidad de ocurrencias de un rasgo supera en uno de los niveles la proporción 10 a 1, el rasgo se considera criterial de

este nivel. En caso de una ratio menor, se atribuye al nivel inferior (Hawkins y Filipovic, 2012: 37).

Otro aspecto importante, como cabía esperar, es la influencia de las lenguas maternas de los usuarios. Un rasgo positivo o negativo puede tener una frecuencia de ocurrencia diferente en virtud de la lengua materna o de la familia a la que pertenezca la lengua, debido a la posibilidad de transferencia positiva o a la propensión a cometer errores de interferencia. Se estudia asimismo el efecto de otras variables como el efecto de la instrucción, el tipo de *input* recibido por los usuarios, etc.

4.3. CORPUS EMPLEADOS EN EL *ENGLISH PROFILE PROGRAM*

En el *English Profile* se han utilizado tres corpus, uno de ellos *ad hoc*. Dos de los corpus son de hablantes no nativos, es decir, corpus de aprendientes. El tercero, un corpus de lengua nativa, algo particular como se explica a continuación.

La fuente fundamental del *English Profile* es el *Cambridge Learner Corpus (CLC)*, en adelante). Como su propio nombre indica, es un corpus de aprendientes. Contiene 45 millones de formas, procedentes de los textos escritos, resultado de realizar las tareas de los exámenes de la factoría Cambridge ESOL, que están referenciados a la escala del *MCER*. Los exámenes son generales o con fines específicos. Cambridge facilita y pone a disposición pública a través de diferentes canales los datos empíricos, cualitativos y cuantitativos, de los procesos de vinculación de sus exámenes al *MCER*. Esto constituye una garantía de que los textos incluidos en el corpus se corresponden realmente con los niveles del *MCER* a los que se asocian.

La selección de los textos del *CLC* responde a un muestreo cuidadosamente estratificado. Los candidatos a examen, autores de los textos que configuran el corpus, presentan diferentes niveles de escala del *MCER* y son hablantes de unas 130 lenguas maternas. Las muestras de cada nivel fueron calificadas en una escala de cinco bandas, de la *A* a la *E*, en las que la banda *A* correspondería a las producciones mejor puntuadas. Se tuvieron en cuenta únicamente las producciones escritas correspondientes a las tres primeras bandas: *A*, *B* y *C* y se desecharon las demás. Las producciones de los candidatos se archivaron teniendo en cuenta las variables siguientes del informante: género, edad, lengua materna, examen, fecha y lugar de realización del examen, nota global del examen y de cada una de las pruebas. El 52 por ciento de los textos se codificó en virtud de los errores, de tipo léxico, sintáctico y morfológico. Se identificaron 76 tipos de errores. El corpus fue analizado y etiquetado automáticamente, empleando el *Robust Accurate Statistical Parser (RASP)* (Harrison, 2015: 5) y, también, lematizado.

La segunda fuente es otro corpus, el *Cambridge English Profile Corpus (CEPC)*, creado *ad hoc* para el *EPP*. La necesidad de disponer de un nuevo corpus se debe a que el *CLC* se nutre únicamente de textos que son producto de tareas de examen. Los expertos consideraron necesario contar con datos de tareas de diferente naturaleza, cuyos productos no se vieran condicionados por las restricciones impuestas por la situación de examen: trabajos de clase, deberes y, lo que es más importante, muestras de producción oral. Las tareas pueden haber sido creadas por los expertos del programa, pero se cuenta también con textos de escritura libre, en que los informantes pueden expresarse sin condicionamientos ni formatos predeterminados. En todos los casos, las tareas y los productos de su ejecución fueron alineados al *MCER*, empleando

procedimientos sistemáticos y rigurosos. En la configuración del corpus, se intenta mantener un equilibrio entre los parámetros siguientes: contextos de aprendizaje, funciones lingüísticas presentes en los textos, lenguas maternas, edades u otros datos demográficos y géneros de las muestras orales (conversaciones, presentaciones, juegos de rol, etc.). La procedencia de los informantes es variada y está cuidadosamente muestreada. Aparte de los que reclutan las Cambridge ESOL y Cambridge University Press, se cuenta con una amplia red de colaboradores voluntarios que se extiende por todo el mundo (Harrison, 2015: 5-7).

Finalmente, se trabaja con un corpus de referencia de hablantes nativos, el *Cambridge English Corpus (CEC)*, con el que se comparan las producciones de los no nativos. El *CEC* contiene más de un billón de palabras, de textos tanto orales y escritos. Se nutre de diversas fuentes y los textos pertenecen a géneros muy variados. Las fuentes de los textos son los manuales, los materiales didácticos complementarios, los textos de entrada de exámenes, etc. Lo que se intenta, en definitiva, es que sea representativo del material de entrada (*input*) que reciben los informantes, es decir, los alumnos, durante los procesos de instrucción.

5. EL CONCEPTO DE 'NIVEL' Y EL PROYECTO *LONGDALE*

A una década y media de la publicación del *MCER*, podría explicarse que, en parte, la extraordinaria acogida que ha tenido se debe a que aporta una respuesta al problema que plantea el concepto de 'nivel'. Pese a que este se encuentre interiorizado en la mente del colectivo de usuarios, es una idea relativamente nueva (Trim, 2010), problemática y no exenta de controversia.

En términos muy simplificados, un nivel, entendido como grado de dominio, equivaldría a un corte limpio y transversal en el continuo que describe el proceso de apropiación de un idioma, y la consiguiente descripción de las características de la lengua que producen o procesa el grueso de los usuarios que alcanza el punto en que se realiza el corte. La complejidad del problema hace brotar preguntas como las que siguen y que se plantean aquí, a modo de botón de muestra: ¿Se consideran en las descripciones los procesos de adquisición y aprendizaje? ¿Cómo se tienen en cuenta variables como las lenguas primeras, segundas o extranjeras y el efecto de la instrucción? ¿Es realmente posible hacer un corte transversal y limpio, válido para todos los usuarios y usos, o es preciso hacer descripciones en términos de «perfil»? ¿Cuáles son los criterios para trazar el escalado de los niveles? ¿Cómo se dibuja la línea que separa el paso de un nivel a otro? ¿Se realizan las descripciones en términos globales o parciales, en virtud del desarrollo de cada uno de los parámetros que definen el dominio de una lengua extranjera? ¿Es posible hacer descripciones al margen de los contextos de uso de las lenguas?...

La comunidad de evaluadores, en especial, la que conforman los profesionales que se ocupan de los exámenes certificativos de alto perfil, es especialmente sensible al modo en el que realiza la definición y la descripción de los niveles y de los grados de dominio de las lenguas. De hecho, su labor consiste precisamente en desarrollar instrumentos capaces de medir el constructo del dominio de la lengua que presentan los candidatos en un momento determinado y, en el caso de los exámenes referenciados un criterio, de comparar los resultados de la medición con descripciones estandarizadas de este criterio, es decir, estándares como los que se encuentran en las escalas de los descriptores ilustrativos del *MCER*. La medición de constructos, de la que se ocupa la psicometría, es

una cuestión compleja, que no admite imprecisiones o ambigüedades.

Una de las críticas principales que se esgrimen en relación con el trabajo de B. North, artífice de las escalas del *MCER*, es la falta de fundamento en los procesos de adquisición y la relación que estos establecen con los niveles comunes de referencia, ya que el escalado de los niveles en el *MCER* se establece a partir de las percepciones de los profesores (North, 2014: 23). En la otra cara de la moneda, Salamoura y Saville (2010: 7, 8) señalan que los estudios de adquisición en una dimensión longitudinal presentan históricamente debilidades que atribuyen a varios factores. Se refieren, en primer lugar, a la falta de rigor con la que los investigadores tienen en cuenta la variable grado de dominio en sus trabajos. Estos, ajenos al ámbito de la evaluación, emplean habitualmente instrumentos no calibrados para establecer las mediciones. Los autores mencionados critican que estos investigadores utilicen terminología ambigua y no normalizada para referirse a los grados de dominio (*principiantes, alumnos de niveles avanzados, etc.*). Constatan también que la variabilidad propia de la interlengua presenta problemas para generalizar los resultados. En segundo lugar, atribuyen los fracasos previos de los estudios longitudinales a que estos se hayan centrado tradicionalmente en un solo aspecto de la gramática, lo cual no resulta rentable, debido al tiempo que requieren y al despliegue de recursos que precisan. En tercer y último lugar, se refieren a que los estudios, para que realmente tengan sentido, deben hacerse en relación con varias lenguas. La metodología de corpus y el uso de instrumentos debidamente calibrados y referenciados a un sistema externo como el del *MCER* harán que vuelvan a cobrar sentido los estudios longitudinales de adquisición de lenguas extranjeras y segundas y que las descripciones de los niveles o grados de dominio estén avaladas por fundamentos anclados en los procesos de

adquisición. El proyecto *LONGDALE (LONGitudinal DATABASE of LEarner English)*⁶ se sitúa en esta línea en relación con la lengua inglesa (Llorián, S., 2016). Por estas razones, es fundamental para las comunidades científicas que se describan de manera sólida o se revisen los elementos lingüísticos de los niveles de referencia en las diferentes lenguas nacionales y regionales, es decir, se retomen desde estas nuevas ópticas los documentos de la serie *T-Level*.

6. LAS CLAVES DE UN PROYECTO DE REVISIÓN DE LOS NIVELES DE REFERENCIA PARA EL ESPAÑOL

6.1. LAS CLAVES DEL PROCESO

A pesar de su escaso impacto, el *English Profile Program*, si lo comparamos con el del *Threshold Level* o el del *Marco de referencia*, este proyecto podría servir de inspiración para guiar una revisión de las descripciones los *Niveles de Referencia para el Español*, al menos en lo referido a los procedimientos empleados. En la página electrónica oficial del *English Profile*, el Dr. Green, uno de los miembros del equipo más activos, firma un documento⁷ en el cual no solamente narra la experiencia global, sino que deja una serie de recomendaciones que se desprenden de las experiencias vividas, destinadas a quienes decidan realizar una réplica del trabajo. Recurriremos varias veces a estas recomendaciones para establecer nuestras claves de revisión.

⁶ <https://uclouvain.be/en/research-institutes/ilc/cecl/longdale.html>

⁷ Recommendations Report:
<http://www.englishprofile.org/resources/recommendations-report>

A nuestro juicio, un proceso de revisión de los *NRE* debería plantearse como una investigación colaborativa, a gran escala, de alcance panhispánico, debido a su envergadura y trascendencia. Por lo tanto, lo idóneo sería que corriera a cargo de un consorcio de universidades de varios países de habla hispana. El planteamiento debe ser pluridisciplinar. Como se ha visto, y se verá con algo más de detalle, se precisa la participación expertos en documentación curricular y de referencia, en sistemas de referencia externos, en desarrollo y adaptación del currículo, en validación de sistemas de examen, en lingüística computacional, en metodología de corpus y en las áreas de los componentes de las descripciones: gramática, léxico, etc.

El planteamiento seguido en el *EPP*, cooperativo y abierto a colaboraciones en red, facilita la implicación de los miembros de la comunidad académica y profesional, además de la disposición de informantes de diferentes contextos. Contribuye, finalmente, a que los resultados no estén sesgados por la instrucción, por el *input* de los manuales o por una documentación curricular peligrosamente unificada. Se podría formular como un proceso de investigación permanente en el que los investigadores se servirían de los recursos del proyecto y, a su vez, lo retroalimentarían. Este tipo de planteamiento requiere planificaciones detalladas a medio plazo y tiene que estar libre de intereses políticos.

6.2. LAS CLAVES DEL PRODUCTO

Se obtendría así un producto en permanente revisión y actualización, propiedad del colectivo de profesionales de la enseñanza y la evaluación del español como lengua española o segunda. En los tiempos actuales no tiene mucho sentido generar

documentos en soporte papel, sino espacios en Internet, que alberguen recursos de diferente naturaleza: los corpus utilizados, las listas de objetivos, los repertorios del material lingüístico que produce y procesa un usuario del español cuando alcanza cada uno de los niveles de la escala de referencia del *MCER*, los resultados de las investigaciones desarrolladas en torno al proyecto, etc. Las actualizaciones de los recursos que nos interesan, por constituir el foco de interés en este artículo, es decir, los inventarios de elementos léxicos, gramaticales, funcionales, pragmático-discursivos, etc., podrían preverse en plazos de unos diez años. Este tipo de reajuste repercutiría de forma muy positiva en la revisión de currículos institucionales y en los ciclos de vida de los exámenes certificativos. Más adelante, y en relación con la naturaleza de cada uno de estos repertorios, nos detendremos en el contenido y en la presentación formal que adquirirían en este formato electrónico: procedimientos de elaboración, naturaleza de la información aportada, secuenciación, etc.

El resultado, en definitiva, podría constituirse en un bien colectivo. Sería un patrimonio compartido entre las comunidades científicas, docentes y evaluadoras de todo el mundo hispano.

6.3. LAS CLAVES DE LOS CORPUS

Como se viene indicando, una de las cuestiones más importantes en relación con el proyecto en cuyas claves nos centramos son los corpus susceptibles de ser interrogados para la extracción de los elementos de los repertorios gramatical, léxico, discursivo, etc. En español, tal y como hemos reseñado recientemente (Llorián, S., 2016), existen corpus informatizados de aprendientes, en número escaso respecto a otras lenguas, varios de ellos disponibles a través

de Internet, con libre acceso. Algunos de ellos se articulan en virtud de los niveles de referencia del *MCER*, como sucede con el corpus *CAES (Corpus de Aprendices de Español* ⁸). Dedicamos un espacio independiente a este corpus en el epígrafe siguiente, por ser el mejor se aproxima a las ideas que aquí esbozamos, aunque, en su versión actual, todavía no dispone de algunas de las características a las que se debe ceñir un corpus que se utilice para los fines que se requieren en un proyecto de la naturaleza que estamos describiendo. Como se verá en virtud de las claves que detallamos a continuación, ninguno de los corpus existentes se puede utilizar para este propósito.

Los corpus utilizados deben estar correctamente parametrizados y cuidadosamente muestreados, para lo cual han de emplearse criterios claramente definidos, con el fin de garantizar la necesaria representatividad. De lo contrario, el resultado de los trabajos que del él se desprendan queda automáticamente invalidado. Se precisaría, además, contar con un corpus lematizado y codificado no solo gramaticalmente, sino en virtud de los errores de los informantes.

Es imprescindible que se aporte y que se haga pública una sólida garantía de que las muestras de lengua tomadas para alimentar el corpus se atribuyen a informantes que presentan grados de dominio equivalentes a cada los niveles de referencia del *MCER*. Para ello, tendrían que medirse estos grados de dominio con pruebas rigurosamente referenciadas a los estándares del documento, aplicando los procedimientos descritos en la documentación correspondiente. En español, son muy escasos los exámenes o pruebas aisladas que cumplan este requisito.

⁸ <http://galvan.usc.es/caes/search>

El resto de las variables consideradas en los parámetros del corpus tienen que reflejarse en un muestreo cuidadosamente estratificado, con el fin de que no se introduzcan sesgos indeseables: lenguas maternas, instrucción recibida español, edad, sexo, tipo de exposición al español, nivel de estudios, tareas utilizadas para extraer las muestras, *input* recibido, etc.

Las muestras que alimenten los corpus deben ser orales y escritas en una proporción adecuada. Deben extraerse de tareas calibradas con rigor. En este sentido, resulta de gran ayuda tomar como referencia la experiencia previa del *EPP* y no limitar la procedencia de las muestras a las tareas de los exámenes, sino que conviene –siempre que esté claro a partir de mediciones rigurosas cuál es el nivel de los alumnos– tomar los productos resultantes de tareas del aula y de deberes o de situaciones de interacción de los aprendientes con hablantes nativos.

En caso de que se opte por una réplica de la metodología del *EPP*, el proyecto tendría que contar con corpus que aglutine las fuentes de *input*, es decir, manuales, *syllabus*, etc. En el caso del español, la influencia de los *NRE-PCIC* en los últimos años podría contaminar los resultados, debido a que es la principal de las fuentes que, en España, alimenta la elaboración de manuales, programas de cursos, currículos institucionales, etc. Esto puede generar un efecto circular. De ahí la importancia de que el proyecto se extienda a otros contextos de enseñanza del español.

7. EL CORPUS CAES

El *Corpus de Aprendices de Español como lengua extranjera (CAES)* es un proyecto financiado por el Instituto Cervantes y desarrollado en la Universidad de Santiago de Compostela, desde el año 2013, por parte de los investigadores Guillermo Rojo e Ignacio Palacios. Se define a sí mismo como «un conjunto de textos escritos producidos por estudiantes de español con diferentes grados de dominio lingüístico» (<http://galvan.usc.es/caes>).

Según la información de la página electrónica⁹, la versión 1.0 de 2014, actualmente disponible, se compone de textos únicamente escritos, de aprendientes de seis lenguas maternas: árabe, chino mandarín, francés, inglés, portugués y ruso. Las previsiones de futuro, según Palacios: (<http://slideflix.net/doc/1016289/ignacio-palacios>), apuntan a incorporar muestras del nivel C2, actualmente inexistentes, textos orales grabados en vídeo y un sistema de codificación de errores. Consta en el momento actual de 575.000 de formas.

A través de esta página de Internet, se proporcionan asimismo otros datos acerca de los informantes como los países de procedencia, la edad, el sexo, etc. No se facilita información acerca del modo en el que se ha realizado en muestreo ni el criterio empleado. Tampoco se justifica la razón por la cual se han seleccionado estas lenguas y no otras. Parece ser las muestras se han recopilado en el seno del Instituto Cervantes (Parodi, 2015), a través de los centros de su red.

Los destinatarios del corpus son, según esta misma fuente, profesionales del campo de ELE (profesores, investigadores,

evaluadores, autores de materiales didácticos, responsables y equipos de centros e instituciones lingüísticas, etc.) que se propongan investigar sobre las producciones de los aprendientes.

Esta primera versión 1.0 se desarrolla, según la documentación accesible sobre el proyecto, en cuatro fases. Durante la primera, se diseña el programa informático destinado a la compilación de los datos, de manera que los colaboradores pueden introducir las muestras directamente el ordenador. En una segunda fase, se lleva a cabo la recogida de los materiales. Se prevé para ello que los informantes completen un mínimo de tareas escritas: «*These tasks were designed according to the CEFR descriptors and DELE tests as well as in accordance with the CI's General Curricular Document*» (Palacios, <http://slideflix.net/doc/1016289/ignacio-palacios>). Estas tareas responden a la siguiente tipología: correos electrónicos a amigos o parientes, reseña crítica de un libro, reserva de la habitación de un hotel, formulación una queja, redacción de una historia en clave de humor. La tercera fase consiste en la codificación y en la anotación del corpus. El etiquetado se realiza automáticamente y la desambiguación es manual. Se utiliza el sistema *FreeLing*¹⁰ en su versión 5 (Parodi, 2015). Finalmente, en la última fase, se desarrolla la interfaz de búsqueda.

Entre las utilidades, señalan sus artífices que el corpus puede usarse para el diseño de *syllabus*, el desarrollo de materiales didácticos, la programación de clases o estudios sobre adquisición, además de asegurar aplicaciones al ámbito de la traducción. Desconocemos la utilidad que pueda tener un corpus de aprendientes para el ámbito de la traducción, pero, en lo tocante al tema que nos ocupa, más relacionado con el resto de las atribuidas, cabe formularse una única

⁹ <http://galvan.usc.es/caes/search>

¹⁰ <http://nlp.lsi.upc.edu/freeling/node/1>

pregunta: ¿puede utilizarse el corpus *CAES*, al menos, en esta primera versión 1.0, como fuente para la revisión de los *Niveles de referencia para el español (NRE)*? A continuación justificamos detalladamente las razones por las cuales la respuesta es negativa, en el estado en el que se encuentran las cosas en el momento de la redacción de este artículo.

Cuestionamos, en primer lugar, el tamaño del corpus en esta versión 1.0. Con este número de elementos, el corpus no puede ser representativo (EAGLES, 1996, Nelson, 2010). Sin entrar en los requisitos mínimos de cualquier corpus para garantizar la representatividad, hay que tener en cuenta que 570.000 formas es la centésima parte del tamaño del *CLC*. El *CAES*, en su versión actual, es demasiado reducido, más cuando se pretende que, con este número de formas, queden representadas todas las lenguas maternas de los informantes.

La segunda limitación del corpus *CAES* para la utilidad en que nos centramos aquí es el número de lenguas que se ven representadas: 6, frente a las 130 que se consideran en el *CLC*. Es cierto que las 6 lenguas del *CAES* son de naturaleza muy diversa. Solo dos de ellas pertenecen a la misma familia.

Por lo demás, como hemos señalado más arriba, no se ha difundido información sobre el criterio de muestreo que se ha seguido si es que se ha empleado alguno, además del de conveniencia. Este problema, el del muestreo, afecta al resto de los parámetros que se han tenido en cuenta: el sexo, la edad, el país en el que se ha tomado la muestra y, por supuesto, el nivel, el más decisivo, al que nos referimos a continuación de forma detallada. Las muestras se extraen, según la documentación adjunta y la información que se ha difundido sobre el *CAES* (Parodi, 2015; Palacios:

<http://slideflix.net/doc/1016289/ignacio-palacios>), mediante la administración de unas tareas, seis en concreto, cuya tipología hemos enumerado más arriba. De ellas, se aportan únicamente los siguientes datos: se diseñan en consonancia con los descriptores del *MCER*, con los exámenes del sistema *DELE*¹¹ y con la documentación curricular del Instituto Cervantes (*NRE-PCIC*).

Un aspecto de la máxima importancia para los investigadores que interrogan un corpus con fines como los que aquí describimos es que se explique en la documentación adjunta cómo se garantiza cuál es verdadero nivel que presentan los informantes en el momento de la administración de las tareas: ¿se les han administrado en fechas próximas pruebas de exámenes previamente vinculados al *MCER*? En caso afirmativo, ¿qué exámenes han realizado? ¿Se pide la nota global de estos exámenes entre los datos que se tienen que aportar de los informantes? ¿Cuál es el criterio que se ha seguido para seleccionar las tareas que tienen que realizar los diferentes grupos de informantes? ¿Se pide, por ejemplo, en caso de asignación aleatoria, a los informantes que presentan supuestamente un nivel A1 que escriban una carta de queja o una reseña crítica? Lo más lógico, aunque no se justifique esta importante decisión, es que los alumnos a los que se les presume un nivel A1 realicen la tarea de la reserva o la del correo electrónico, por ejemplo. Si el número global de tareas, seis, es ya de por sí escaso, al limitar su administración a los grupos de informantes a los que se supone un nivel de la escala del *MCER*, se sesgan inevitablemente los resultados, condicionados por el efecto de las tareas. Hay que pensar en que el efecto de la tarea llega a cuestionarse en los resultados obtenidos del *EPP* (Green, 2012: 3), en el que uno solo de

¹¹ DELE (Diploma de Español como Lengua Extranjera). Información disponible en: <https://exámenes.cervantes.es/es/dele/que-es>

los corpus utilizados multiplica por 100 el tamaño del CAES, como hemos dicho. A esto podemos añadir que, en la alimentación de los corpus del *English Profile* se manejan textos obtenidos de la realización cientos de tareas diversas, tanto de examen como otras más propias de la actividad cotidiana del aula.

La siguiente área problemática de este corpus en relación con este tipo de uso es la interfaz. La identificación de las formas lingüísticas correspondientes a los componentes a los que nos referiremos más abajo precisa algunas de las herramientas que ofrecen los gestores de corpus más comúnmente utilizados: clúster, listas de palabras, palabras clave en contexto, frecuencias, líneas de concordancia, etc. (Scott, 2010). Los resultados que arroja la interfaz del corpus CAES son líneas de concordancia, con un contorno bastante reducido. Hay otros aspectos de la interfaz que, no obstante, resultan bastante útiles como los filtros que se pueden hacer en relación con los parámetros considerados: lengua materna, país de procedencia, etc.

Nos referimos, finalmente, a la codificación de los errores de los alumnos, un elemento básico en un corpus de aprendientes. Sin este, no se pueden extraer los rasgos criterios, tal y como se ha explicado más arriba.

En definitiva, el CAES, en su versión actual, podría seguramente tener cientos de valiosas utilidades, pero, a la vista de lo que acabamos de argumentar, no se puede emplear para la elaboración de documentos de referencia, en la línea que aquí nos interesa. Permanecemos a la espera que la declaración de intenciones proyectadas hacia versiones futuras se vea materializada.

8. LAS CLAVES DE REVISIÓN DE LOS COMPONENTES DE LOS NRE

8.1. CUESTIONES GENERALES

En este último epígrafe concretamos las claves de la revisión relativas a los inventarios de contenidos lingüísticos, en relación con los componentes siguientes: el léxico, el gramatical y el pragmático discursivo. Se hace especial énfasis en el componente léxico semántico, debido a lo productiva que, para este, resulta la metodología en la cual nos centramos aquí. Estas claves constituyen el fruto de una reflexión individual que se desarrolla al hilo de la propia experiencia en trabajos de investigación y en el desarrollo de documentos curriculares y de referencia en varias organizaciones. Las claves aportadas respecto a la revisión de los inventarios de los NRE considerados aquí no tratan, en ningún caso, de cuestionarlos ni mucho menos de invalidarlos, sino de aportar una visión, como se indica en el título de este artículo, sobre la forma en la que podrían actualizarse, a la luz de la metodología de corpus. Nuestro planteamiento es, por supuesto, discutible y complementario, y compatible con otros puntos de vista que se decida adoptar para llevarla a cabo.

8.2. LAS CLAVES RELATIVAS AL COMPONENTE LÉXICO SEMÁNTICO

Los NRE-PCIC despliegan este componente en dos inventarios, que conectan con la tradición de la serie de documentos *T-Level*, al utilizar como unidad de análisis las nociones generales, por un lado, y las nociones específicas, por otro. Las entradas de especificaciones de realizan, por consiguiente, a partir del índice establecido en los documentos de la serie *Nivel Umbral*, con algunas modificaciones.

Lo que varía y lo que marca la progresión a lo largo de la estructura vertical es el número de exponentes y ejemplos ilustrativos, cada vez más amplio y variado. Al igual que venía haciéndose en los documentos *T-Level*, se diferencia el vocabulario que se espera se maneje en un nivel receptivo (R), del productivo. Es importante recordar aquí que, en las especificaciones de este tipo, no ha habido nunca afán lexicográfico, por lo tanto, de exhaustividad, sino que la idea es aportar material ilustrativo de cada uno de los niveles comunes de referencia.

Entre las novedades que se introducen en los *NRE-PCIC* de 2007, destaca la trascendencia de la idea de 'vocabulario' a favor de la de 'léxico', al considerar la *unidad léxica pluriverbal*, además de la *palabra*, a la que se restringen los documentos tradicionales. Se da cabida, por lo tanto, al concepto de combinatoria léxica y se tienen en cuenta las colocaciones y otras unidades fraseológicas como las expresiones idiomáticas o las locuciones (Pérez Serrano, 2016: 225). El concepto de fraseología es, en consecuencia, abierto y flexible, en la línea de Lewis (1993, 2000), principal referente, Corpas (1996) o Higuera (2006). Las especificaciones de este tipo de unidades no resultan sencillas, debido al tratamiento asistemático e inconsistente que los componentes léxico y semántico reciben en el *MCER* (Corpas, 2008: 337; Pérez Serrano, 2016: 219).

Las claves que aquí planteamos parten de la base de conservar esta dimensión horizontal del léxico y de tomar la unidad léxica mono y pluriverbal como unidad de análisis de la especificación. No obstante, el formato electrónico y el acceso a través de sistemas de búsqueda como los que facilitan las interfaces llevan a preguntarse si las entradas pueden y deben continuar siendo las nociones, tal y como las concibiera D. Wilkins en 1976. Esta cuestión no es baladí, a pesar de que pudiera parecerlo, al menos, a efectos programáticos,

puesto que estas listas servirán de referencia para la elaboración de currículos o programas de curso. El concepto de la noción general (*distancia, posterioridad, existencia, etc.*) o el tema de la noción específica (*comidas y bebidas, relaciones personales, trabajo, etc.*) puede ser el elemento seleccionador de las unidades léxicas susceptibles de ser tratadas en las aulas. Por lo tanto, una de las claves reside en que las interfaces de acceso al componente léxico incluyan filtros fundamentados en las tradicionales nociones generales y nociones específicas.

Otra cuestión, relacionada también con el formato electrónico en el que se presente el producto, es la que plantea Pérez Serrano (2016), en relación con las colocaciones. Esta autora, en su pormenorizado análisis del tratamiento de este tipo de unidad léxica en los *NRE-PCIC*, se detiene en el criterio selección del elemento especificado, es decir, se plantea si la entrada de las unidades es la base, es el colocativo o no es ninguno de los dos. El problema se repite en el formato electrónico y se extiende a otras unidades pluriverbales: modismos y locuciones.

Por último, recordamos, en relación con las especificaciones actuales, una cuestión que es también una novedad de los *NRE-PCIC*. En lugar de repetir la unidad léxica especificada en niveles previos, se introducen variaciones determinadas por los colocativos, es decir, a medida que se progresa en la estructura vertical, aumenta el número de colocativos asociados a una base que se va repitiendo de un nivel a otro. Otros criterios de progresión son el movimiento a lo largo de un continuo, en cuyos extremos se sitúa la neutralidad (niveles inferiores) y la restricción de contextos de uso (niveles superiores).

Lo más importante aquí son las claves de la revisión de ese componente en cuanto a lo que puede suponer la aplicación la metodología de corpus para el análisis y la extracción de las unidades léxicas, junto con el uso otros procedimientos tecnológicos. Las aportaciones de la lingüística de corpus en relación con el componente léxico semántico radican en ampliación de las posibilidades de estudio de las palabras en contexto, al evidenciar las tendencias de las palabras a aparecer unas junto a las otras. Timmis (2015: 27) señala la influencia que ha tenido este enfoque en la recuperación del análisis del léxico en una dimensión horizontal y en las relaciones que establece este componente con la gramática. Destaca la utilidad de las herramientas de los programas informáticos con los que se gestionan los corpus, es decir, las líneas de concordancia, palabras clave en contexto, clústeres para la extraer unidades léxicas pluriverbales. Gracias a estas funcionalidades de los programas, además de las colocaciones, las locuciones y las expresiones idiomáticas, se pueden identificar patrones léxicos (Timmis, 2015: 28). Volveremos sobre esta noción más abajo, al hilo del componente pragmático discursivo.

Hay varios estudios realizados sobre el español como lengua extranjera o segunda que utilizan metodología de corpus. Lozano (2015) hace una minuciosa y rica recopilación comentada de los trabajos relativos a las colocaciones, a los marcadores del discurso y diversas palabras gramaticales, realizados hasta la fecha de la publicación de su artículo. Las investigaciones reseñadas se han llevado a cabo con corpus que no están referenciados a los niveles del *MCER* y que parten, como hemos comentado más arriba, de estimaciones muy vagas acerca del concepto de nivel, que los invalidan para los usos que aquí nos ocupan.

Hay un aspecto que resulta también controvertido e interesante derivado del uso de la metodología de corpus. Nos referimos a la recuperación del criterio de frecuencia, arrastrada a este campo por los programas gestores de corpus. La frecuencia como criterio de selección y secuenciación del léxico quedó erradicada del diseño curricular desde la irrupción de los enfoques comunicativos y los avances de la psicolingüística. Los detractores (Timmis, 2015; Capel, 2015, entre otros) de que la frecuencia vuelva a considerarse como criterio de secuenciación curricular aducen razones de peso como, por ejemplo, la ruptura de paradigmas (p. e. demostrativos, artículos, pronombres personales, etc.) o el hecho de las investigaciones demuestren que no existen correspondencia entre este criterio, la frecuencia de uso, y los procesos de aprendizaje y adquisición del léxico (Capel, 2015: 13).

El componente léxico semántico se recoge en el *EPP* en el *English Vocabulary Profile (EVP)*, en este artículo). En el desarrollo de este recurso, se ven involucrados lexicógrafos, especialistas en Lingüística de Corpus y en el inventariado de este tipo de unidades en documentos de esta naturaleza. La investigación que subyace a la especificación de las unidades léxicas en el producto final se desarrolla en dos fases. En primer lugar, se hace para los cuatro primeros niveles de la escala del *MCER* y, posteriormente, se trabaja en los niveles C.

Los primeros resultados revelan que, en los niveles A1 y A2, los informantes producen palabras aisladas o frases léxicas muy estereotipadas, generalmente sustantivos relacionados con temas generales, verbos de acción y palabras gramaticales muy básicas. A medida en que se progresa en la estructura vertical de la escala de referencia, en los niveles B, se van incorporando palabras polisémicas y se van ampliando progresivamente los usos con

significados distintos, además de multiplicarse el uso de unidades léxicas pluriverbales (Capel, 2015: 10). Estos hallazgos vienen a validar, en cierto modo, la decisión intuitiva tomada en los *NRE-PCIC* respecto a las colocaciones. No obstante, como primera clave de la revisión, consideramos que es necesario que se retome la repetición de unidades léxicas de nivel a nivel, teniendo en cuenta la ampliación del abanico de significados con los que se usan o las expresiones idiomáticas en las que aparecen y no solo los colocativos que se añaden de forma progresiva. La segunda clave sería, obviamente, la revisión de las unidades léxicas pluriverbales, aplicando funcionalidades que ofrecen los corpus informatizados como las líneas de concordancia o los clústeres.

Un problema que reviste la máxima importancia es la de la cobertura de los niveles, es decir, la garantía de que, con las especificaciones que se hacen, quedan representado el léxico receptivo y productivo que manejan los usuarios de cada uno de los niveles comunes de referencia, en los diferentes contextos de uso. Esto afecta especialmente a los niveles C. Las versiones de los documentos de la serie *T-Level* anteriores al *EPP*, entre las que se incluyen los *NRE*, resolvían el problema gracias a las delimitaciones establecidas por las listas de las nociones generales y los temas de las nociones específicas. Este tipo de criterio no es compatible con un enfoque construido sobre cimientos empíricos. Lo que se hace en el *EPP* para realizar las especificaciones, además de aplicar la metodología referida a los rasgos criteriosales, es utilizar repertorios ya existentes (listas, listados de frecuencias, lexicones, diccionarios, etc.) como puntos de partida o referencia y completar los hallazgos con procedimientos computacionales. Estos se podrían considerar como la tercera de las claves de revisión del componente léxico en los *NRE*.

La siguiente clave, para la que resulta muy útil tomar el *EVP* como antecedente y referente, es el análisis de las unidades léxicas que aparecen en las respuestas de los usuarios como consecuencia de los efectos de las tareas de extracción, de la instrucción que hayan recibido o de la transferencia de las lenguas maternas. Las evidencias halladas en el *EVP* (Capel, 2015: 14) son del máximo interés.

Las claves de la revisión en las que nos centramos seguidamente corresponden al formato de presentación, es decir, a las entradas léxicas, a la información que acompaña a cada una y a la interfaz de búsqueda. En el *EVP* las entradas se conciben, en primer término, en forma de palabras ordenadas alfabéticamente. Las palabras que se insertan en unidades léxicas pluriverbales se escriben con mayúscula. En relación con cada entrada, se facilita la siguiente información, bastante acorde, aunque no del todo, con los repertorios léxicos que se pueden encontrar en la actualidad en soportes electrónicos: archivo de audio, transcripción fonética, categoría gramatical, definición lexicográfica correspondiente al significado del nivel en el que aflora en las interlenguas, ejemplo de uso de hablantes nativos, frecuencia de uso en la lengua nativa, colocados más frecuentes y expresiones idiomáticas en las que aparece; como novedad, información derivativa y afijación, que remite a los niveles en los que se produce. Lo más importante, un ejemplo de producción de usuarios no nativos. Esta última información viene acompañada del examen del que se toma la muestra; obviamente, del nivel del examen y de la lengua materna del informante.

Además de las aportaciones que pueda reportar la referencia del *EVP*, tal y como las acabamos de describir, nos parece interesante tomar en consideración otras claves para la presentación de las unidades léxicas en el producto final, en relación con la información

que se puede añadir a la entrada. Nos inspiramos en los perfiles de unidad léxica con origen en las propuestas de Nation (2001), perfeccionados más tarde, sobre todo, gracias las aportaciones de los corpus como sucede con el de modelo de O'Keefe *et al.* (2007). De este modo, además de lo dicho, podrían considerarse las restricciones combinatorias de las unidades, para lo cual contamos con una inestimable fuente lexicográfica de referencia: Redes (Bosque, 2006). Resultaría bastante rentable, para finalizar, poder acompañar la entrada léxica con información acerca de las variedades del español con las cuales se puede asociar su uso. Las investigaciones asistidas por corpus que se están llevando a cabo en la actualidad para el español en una dimensión panhispánica resultarían de gran ayuda.

8.3. LAS CLAVES RELACIONADAS CON EL COMPONENTE GRAMATICAL

La información de este epígrafe se refiere únicamente al inventario que, en la versión actual de los *NRE-PCIC*, lleva el nombre de «Gramática», aunque empleemos, por coherencia con los demás epígrafes, la idea de 'componente'. Excluimos, por lo tanto, la consideración de los de «Ortografía» y «Pronunciación y prosodia», los otros dos inventarios considerados dentro de este componente gramatical en los *NRE-PCIC*.

Para tratar la gramática desde la perspectiva y el enfoque que venimos adoptando hasta aquí, es necesario que los corpus que se utilicen estén lematizados, anotados y codificados con un analizador automático potente y, posteriormente, desambiguados combinando métodos tanto manuales como automáticos. Es conveniente que se cuente, además, con información estadística

(Lozano, 2015). Lo más importante es la codificación de los errores, que ha de hacerse necesariamente de forma manual. Esta tarea se declara especialmente compleja para el tratamiento de los textos orales, que requiere un trabajo previo de transcripción.

El inventario de «Gramática» de los *NRE* presenta una organización de sus especificaciones, a partir de las unidades de análisis clásicas de esta disciplina: clases de palabras, sintagma y oración (*NRE*, 2007: 102). No obstante, estas se realizan con criterios pedagógicos (*NRE*, 2007: 101), buscando el equilibrio entre norma y uso que se requiere en este tipo de documento. El enfoque adoptado permite, por una parte, hacer mención a las áreas que son proclives a que los hablantes extranjeros cometan errores; estos se muestran en los ejemplos junto con las formas correctas. Por otro lado, se presta a explicar las numerosas restricciones semánticas que se dan en las construcciones gramaticales. Este es uno de los muchos puntos fuertes de este inventario. La adopción de una metodología basada en las claves que aquí venimos describiendo, utilizando corpus en los que se codificaran los errores, permitiría hacer una redistribución fundamentada de los fenómenos ejemplificados en los *NRE* a lo largo de la estructura vertical del currículo.

Por todo lo hasta aquí dicho, las claves de la revisión de los *NRE* respecto a las especificaciones gramaticales se podrían plantear en dos direcciones, de forma que se diera una interacción entre corpus de hablantes nativos y corpus de informantes no nativos. Gracias a los corpus de hablantes nativos, podrían aportarse ejemplos basados en el uso real de la lengua e ir incorporando resultados relevantes de los estudios sobre la gramática, que está experimentando avances espectaculares en los últimos años.

Sin perder de vista lo anterior, la clave que se constituiría en el núcleo de la revisión es, sin duda, la que se refiere a la identificación de los rasgos criteriosales, tanto positivos como negativos. Seguimos, por lo tanto, la estela proporcionada por el *EPP*, en la que nos detenemos sucintamente a continuación.

El resultado de la investigación que se lleva a cabo en el *EPP* en torno a la gramática tiene como producto lo que se llama *English Grammar Profile* (*EGP*, en adelante). Es un repertorio de rasgos gramaticales, que entraña una serie de particularidades. No se trata de un mero elenco de los aspectos de la gramática que han adquirido los usuarios cuando alcanzan cada uno de los niveles de la escala del *MCER*, sino que va mucho más lejos. En el *EGP*, se especifica lo siguiente: en primer lugar, la evolución de los rasgos gramaticales a lo largo de la escala, es decir, las formas correspondientes a cada nivel y la capacidad del usuario para emplearlas en diferentes estructuras y en distintas partes de la oración (sujeto, objeto, etc.); en segundo lugar, el desarrollo léxico gramatical, o sea, la habilidad de usar un determinado rasgo de la gramática con un rango de vocabulario; en tercer lugar, la habilidad de utilizar el rasgo especificado en una serie de funciones y de contextos; finalmente, los análisis que permite la metodología de corpus evidencian la adquisición de la lengua se produce en forma de patrones. Estos patrones serían de dos tipos: por un lado, las unidades fraseológicas, conocidas en inglés tradicionalmente como *'chunk'* (bloques), que son unidades cerradas y completas, intercambiables por otra unidad dentro de un enunciado; por otro, unidades incompletas, abiertas en su formulación, que se denominan *'frames'* (estructura, esquema), que combinan para completarse con otras secuencias, más complicadas a medida en que se progresa en la escala del *MCER* (Harrison y Barker; 2015: 38). Estas tienen una parte fija y otra variable. Un ejemplo de

progresión de un *frame*, podría ser el siguiente: *No me importa, No me importa si vienes, No me importa que vengas, No me importa que vengas o que te quedes, Me importa un bledo que vengas o que te quedes en tu casa...* De esta manera, se puede ver cómo va evolucionando cada uno de los rasgos de la gramática desde que aparecen en un determinado nivel, a lo largo de la escala, en relación con el vocabulario con el que se acompaña y con las funciones y contextos en los cuales se usa.

Lo más interesante del *EGP* no es tanto la especificación que figura en el *EPP*, sino los hallazgos en relación con los procesos de adquisición y su relación con el aprendizaje y la enseñanza. En inglés, se evidenció, por ejemplo, que hay numerosas formas que se enseñan tradicionalmente en ciertos niveles y que, sin embargo, se adquieren mucho antes. En cualquier caso, se demostró que la adquisición se produce solamente si se tiene la oportunidad de utilizar las formas en contextos reales de comunicación. También, que los rasgos gramaticales se adquieren de forma progresiva y esta progresión está estrechamente relacionada con la ampliación del vocabulario con el que se emplean (Harrison y Barker, 2015: 34). En definitiva, en un breve espacio de tiempo se han realizado espectaculares descubrimientos, que sin duda continuarán poniéndose de manifiesto, gracias a la oportunidad que se brinda a los investigadores de trabajar con este valioso material.

Hay otras cuestiones que los profesionales de la enseñanza de lenguas venían intuyendo y que se vienen corroborando a lo largo de los desarrollos del enfoque comunicativo. A medida que se progresa en el grado de dominio, los usuarios van siendo capaces de seleccionar las formas gramaticales para adecuarlas al grado de cortesía que se requiere en los contextos de uso y también a las unidades fraseológicas idiosincrásicas de las áreas académicas o

profesionales en las que se desenvuelva. Los niveles de referencia para español han hecho una separación que se ha revelado, a nuestro juicio, bastante poco operativa desde el punto de vista didáctico entre gramática y pragmática. El inventario de «Tácticas y estrategias pragmáticas», de hecho, es prácticamente en sí mismo un conjunto de referencias cruzadas a los de «Gramática», «Funciones» y «Géneros discursivos y productos textuales». Se trata de una descripción más propia de la lingüística teórica que de la didáctica del español como lengua extranjera. Desde el punto de vista programático, no tiene mucho sentido que se considere como un inventario independiente. La manera de canalizar su contenido hacia las aulas es a través de unidades de análisis usadas en la programación de contenidos como las funciones, las oraciones o los textos.

En resumidas cuentas, las claves de la revisión consistirían en retomar de forma conjunta los inventarios «Gramática» y «Tácticas y estrategias pragmáticas» y en la identificación de los rasgos criterios de cada nivel. Los importantes resultados de la investigación obtenidos con el *English Profile* invitan al estudio conjunto de la gramática y el léxico; también, a considerar la relación forma-función, tal y como hiciera años atrás Francisco Matte Bon (1992), pero esta vez con la asistencia de corpus con las características más arriba descritas.

8.4. LAS CLAVES DEL COMPONENTE PRAGMÁTICO DISCURSIVO

En el epígrafe anterior nos hemos referido al inventario de «Tácticas y estrategias pragmáticas» de la versión de los *NRE* del Instituto Cervantes (2007) y a su relación con el inventario de «Gramática». Las referencias, directas y sobre todo indirectas, que haremos aquí al

primero de estos inventarios, se realizarán en relación con de «Funciones», por las razones que han quedado expuestas más arriba. Nos centraremos, por lo tanto, en facilitar en este epígrafe las claves de la revisión respecto a los otros dos inventarios de este componente: «Géneros discursivos y productos textuales» y «Funciones». Comenzaremos por el primero, dada la naturaleza más amplia y abarcadora de sus unidades de análisis: el texto y la macrofunción.

El español es la única lengua que, en los documentos de la serie *7-Level*, se aventura más allá de las nociones y funciones o de la oración como unidades de análisis para hacer las especificaciones del material lingüístico. En el inventario de «Géneros discursivos y productos textuales» de los *NRE* se toman el texto y la macrofunción para hacer una de las descripciones, a nuestro parecer, más novedosas y rentables desde el punto de vista pedagógico, en todos los niveles de la escala del *MCER*, en especial, los niveles C1 y C2, en que, según Hawkins y Filipovic (2012: 22), la extensión de los enunciados aumenta exponencialmente respecto a los niveles anteriores.

La idea de 'macrofunción' aparece por primera vez en el *MCER*, que, a su vez, según Green (2012: 28), la toma de Cook (1989), y responde a la de 'secuencia textual', si se sigue el modelo de Adam (1992). En los *NRE* se consideran las secuencias propuestas por este autor, a excepción de la dialógica: exposición-explicación, descripción, narración, argumentación. Además de tener en cuenta la forma prototípica en la que se disponen las proposiciones en cada una de estas secuencias, se trata el modo en el que operan en ellas las denominadas funciones retóricas o estrategias discursivas: definición, analogía, enumeración, etc.

Las claves de la revisión de las macrofunciones en los *NRE-PCIC* serían relativamente sencillas, si se comparan con las del resto de los elementos de este inventario. Lo más importante sería realizar un diseño adecuado de las tareas que permitieran obtener productos textuales en que se pudiera apreciar claramente una secuencia textual envolvente. El análisis de los textos con la metodología de corpus permitiría la identificación de los patrones y regularidades que utilizan los usuarios en cada nivel de la escala del *MCER* para ejecutar cada una de las funciones retóricas descritas, además de los rasgos criteriosales y las unidades léxicas mono- o pluriverbales características de las secuencias textuales.

Aparte de las macrofunciones, el documento de los *Niveles de Referencia para el Español*, en su versión actual de 2007, presenta en este mismo inventario una lista de los géneros discursivos susceptibles de ser comprendidos "(R)" o producidos "(P)" por los usuarios en cada nivel. Se presentan muestras ilustrativas (*muestras de géneros*) de algunos de ellos, que aparecen analizados en virtud de su estructura retórica (movimientos y pasos). Se muestran, además, ejemplos de producción de los usuarios.

La primera de las claves de la revisión de estas dos partes consistiría en dotar de fundamento empírico a la selección de los textos de la lista de cada nivel, lo cual se complica a medida en que se avanza en la escala de los niveles comunes de referencia. Nos referimos a los problemas implícitos en la elaboración de la lista de géneros discursivos que procesan o producen los usuarios de los niveles C1 y C2, incluso del nivel B2. A esto se añade la dificultad que entraña la clasificación de los textos en virtud de los géneros discursivos, como señalan, por ejemplo, dentro del ámbito académico, Parodi, Julio y Vásquez (2015). Hay que tener en cuenta, por lo demás, que la noción de género discursivo figura en los descriptores ilustrativos del

MCER únicamente en la escala de *escritura creativa* (*MCER*, 2002: 65). El resto de las alusiones que se hacen a los tipos de textos que comprenden los usuarios no son consistentes (Alderson *et al.* 2006); tampoco, las de los que produce (Llorián, 2012). No obstante, en el *MCER* se menciona, entre las referencias citadas, la que se ha empleado en los *NRE* con fuente fundamental: Swales (1990). Por lo tanto, para relacionar los géneros con los niveles de la escala del *MCER*, sería necesario un trabajo de investigación, basado en una evaluación rigurosa mediante pruebas de exámenes referenciados a la escala del documento, que logre identificar los rasgos léxicos, gramaticales, funcionales, etc., característicos de los textos que comprenden y producen los alumnos en cada nivel. La estimación de la dificultad de los textos orales o escritos susceptibles de comprensión se puede calcular a través de procedimientos cualitativos y también cuantitativos, con parámetros como la sintaxis, la cohesión, la densidad léxica, el contenido, etc., y vincular esta dificultad a la escala de niveles del *MCER*, tal y como describe Green (2012: 115-146). La réplica de estos procedimientos constituiría una importante clave de la revisión de los actuales *NRE*, de gran trascendencia e importancia crucial para la elaboración, sobre todo, de pruebas de examen y de materiales didácticos. En el primer caso, ofrecería una sólida garantía de validez.

La siguiente clave consistiría en realizar el análisis de los géneros de los cuales se van a proporcionar muestras ilustrativas, en virtud de la organización retórica de los textos: movimientos, pasos y patrones característicos de cada uno de ellos. Las que se facilitan en el documento de 2007 están fundamentadas en las fuentes bibliográficas referenciadas en los *NRE*. Algunas, como la del *trabajo de tipo académico*, en trabajos de investigación (Gil, 2005). Sería necesario, por consiguiente, apoyarse en investigaciones previas para describir la organización retórica de los textos de cada género

seleccionado, tal y como se establece entre los miembros de las comunidades discursivas de los hablantes nativos.

Las muestras de producción de los hablantes no nativos facilitadas en el documento de referencia serían producto de la aplicación de una combinación de procedimientos de análisis del género, manuales y basados en corpus, que permitan identificar los patrones y las regularidades que se dan en cada una de las partes de los textos (Biber, 2010). El desarrollo de este tipo de inventario requeriría investigaciones que irían nutriendo los documentos de referencia, con técnicas como las propuestas por Biber *et al.* (2007).

Completamos este epígrafe con las claves de la revisión de las funciones en los *NRE*. La elaboración de este inventario en la versión de 2007 siguió las pautas facilitadas por el documento guía del Consejo de Europa (2005). La especificación se hizo a partir de una interpretación intuitiva de las escalas del *MCER* en que se hacía mención directa o indirecta a las funciones comunicativas. Hay que recordar que esas alusiones no son consistentes (Green, 2012; Harrison, 2015). En el desarrollo de los inventarios, se establecieron unos criterios de progresión (*NRE-PCIC*: 209, 210), que llevan no solo a añadir algunas funciones comunicativas a la lista, a medida que se progresa en los niveles de la escala, sino también –y lo más importante–, a ampliar el abanico de exponentes, su complejidad léxica y morfosintáctica, y a introducir restricciones establecidas por los contextos de uso. Todo esto implica que, sobre todo, más allá del nivel B2, se multiplique el espectro de los registros de uso de la lengua y que los exponentes vayan siendo cada vez más largos y complejos. Estos criterios son bastante similares a los que siguen Trim y van Ek en los documentos de las series del *T-Level* (Green, 2012: 29-32).

La especificación de los exponentes funcionales sobre bases empíricas requiere el uso de corpus, sobre todo, orales. De hecho, una parte considerable de las funciones responde al concepto de pare adyacente que se da en las conversaciones tanto transaccionales como casuales. En el *EPP*, la especificación de las funciones se efectuó estableciendo el punto de partida en las versiones de las series *T-Level* (1990, 2001). Sobre esta base, se hicieron los ajustes necesarios en los exponentes de los niveles A1, A2, B1 y B2, tomando como referencia las descripciones del *English Vocabulary Profile* y del *English Grammatical Profile* más arriba descritos (Green, 2012; Harrison, 2015).

Con independencia de que el *EVP* y el *EGP* ofrezcan datos extraídos de fuentes empíricas, hay problemas que persisten. El primero es la correspondencia entre el nivel B2 del *MCER* y el *Vantage Level*, que parece que nunca ha estado del todo claro (Green, 2012: 32); el segundo, la progresión de las funciones a lo largo del desarrollo del grado de dominio de los usuarios; finalmente, la especificación de las funciones correspondientes a los niveles C1 y C2, que resulta problemática, puesto que la competencia funcional, en estos niveles, requiere del usuario la capacidad para adaptarse a los contextos de uso de la lengua, que ya no se pueden delimitar, al igual que los temas sobre los que versa la comunicación.

Parece ser que ya Van Ek y Trim, citados en Green (2012: 31, 32), estiman que algunas de las especificaciones del *Vantage Level* se corresponden con aspectos más bien propios de los niveles C1 y C2. Por otra parte, en la progresión de las funciones, se hallan implícitas algunas variables que no son estrictamente lingüísticas, sino que corresponden a convenciones socioculturales, que, por lo tanto, no dependen de los procesos naturales de adquisición y tampoco son comparables entre las lenguas. La propuesta de Green (2012), en el

marco del *English Profile Program*, para los niveles C1 y C2, es trascender el concepto de función, tal y como se concibe en los documentos de la serie *T-Level*, y cubrir la dimensión funcional de la competencia lingüística con la definición de objetivos en un grado de especificación que él considera *generativo*, es decir, aquel en el que los enunciados tipo *can do statements* permitan diseñar tareas, a partir de las cuales se obtengan muestras ilustrativas de los niveles. Esto requiere que los enunciados contengan los siguientes elementos, a los que ya hicimos mención más arriba: la actividad lingüística (verbo de acción), el tema o situación, el texto de entrada, el texto de salida, el nivel de logro en términos de competencias y las condiciones o restricciones en las que se realiza la actividad.

En resumidas cuentas, las claves de la revisión de las funciones requieren ajustes en las especificaciones de los *NRE*, en virtud de los resultados del análisis de los componentes léxico y gramatical. Sobre todo, investigación acerca de la lengua real que procesan y producen los usuarios, en todos los niveles de referencia, en especial, en los niveles C. El producto de la especificación, en soporte electrónico, debería incluir archivos de audio que permitan apreciar aspectos de la lengua oral como el grado de fluidez, los patrones entonativos o la pronunciación.

9. CONCLUSIONES

Las claves de la ineludible revisión basada en metodología de corpus de los niveles de referencia para el español, expuestas a lo largo de esta argumentación, se establecen en virtud de las líneas que se exponen a seguidamente:

- la réplica de algunos de los procedimientos utilizados en proyectos similares, desarrollados con anterioridad como el *English Profile Program*;
- las características de los corpus empleados, que, en el caso del español, están pendientes de desarrollo;
- la metodología utilizada para la especificación a partir de los corpus;
- el proyecto, de dimensión panhispánica, a cargo de un consorcio internacional de universidades, entendido como un proceso de investigación permanente;
- el producto, que se concibe como una fuente de recursos, en soporte electrónico, con acceso libre a través de Internet, patrimonio de toda la comunidad docente e investigadora, en continua actualización.

Estas claves apuntan a que los currículos, los programas de curso, los exámenes de español y los materiales didácticos que declaran estar referenciados a la escala de niveles de referencia del *MCER* ofrezcan garantías reales de esto es realmente así, en especial, en el caso de los exámenes certificativos. También, a que el español se sume a las iniciativas tomadas ya en otras lenguas, sobre todo el inglés, para crear materiales que medien entre el *Marco de referencia* y los usuarios, con el fin de salvar sus carencias, ambigüedades y dificultades de uso. Finalmente, a contribuir al logro los objetivos de transparencia y coherencia en el aprendizaje, la enseñanza y la evaluación de lenguas en el continente europeo y en todo el territorio mundial.

10. REFERENCIAS

Adam, J.M. (1992), *Les textes: types et prototypes. Récit, description, argumentation, explication et dialogue*. París: Nathan.

Alderson, J.C., Figueras, N., Kuijper, H., Nold, G., Takala, S., y Tardieu, C. (2006), "Analysing tests of reading and listening in relation to the *Common European Framework of Reference*. The experience of the *Dutch CEFR construct Project*". *Language Assessment Quarterly*, 3(1): 3-30. Disponible en: http://kiesplang.fi/blog/wp-content/uploads/2011/10/CEF-Construct_article_2006.pdf

Austin, J.L. (1962), *How to Do Things With Words*. Oxford: Oxford University Press.

Biber, D. (2010), "What can a corpus tell us about registers and genres?" en *The Routledge Handbook of Corpus Linguistics*. New York: Routledge.

Biber, D., Connor, U. & Upton, T. (2007), *Discourse on the move*. Amsterdam: John Benjamins.

Borobio, V. (1994), *ELE 1*. Madrid: S.M.

Bosque, I. (dir.). (2006), *Diccionario combinatorio práctico del español contemporáneo*. Madrid: S.M.

Capel, A. (2015), "The English Vocabulary Profile" en Harrison, J. y Barker, F. (2015), *English Profile in Practice*. Cambridge: Cambridge University Press.

Consejo de Europa (2005), *Reference Level Descriptions for National and Regional (RLD) Languages. Guide for de production of LRD*. En línea Disponible en la web:

http://www.coe.int/t/dq4/linguistic/DNR_EN.asp#P26_1791

[consulta: 30 de junio de 2017].

Consejo de Europa (2002), *Marco común europeo de referencia para las lenguas: aprendizaje, enseñanza y evaluación*. Madrid: Anaya, Instituto Cervantes. Disponible en la web:

http://cvc.cervantes.es/ensenanza/biblioteca_ele/marco/cvc_mer.pdf

[consulta: 30 de junio de 2017]

Corpas, G. (2008). "Detección, descripción y contraste de las unidades fraseológicas mediante tecnologías lingüísticas". En Olza I., Manero, E. (2008) (eds). *Fraseopragmática. Colección Romanistik*.

Berlín: Frank y Timme, 2013. Disponible en la web:

<http://www.Franktimme.de/verlag/verlagsprogramm/buch/verlagsprogramm/bd-15-ines-olzaelviramanero-richard-eds-fraseopragmatica/backPID/romanistik-1.html>

[consulta: 30 de junio de 2017]

Corpas, G. (1996), *Manual de fraseología española*. Madrid: Gredos.

Cook, G. (1989), *Discourse*. Oxford: Oxford University Press.

EAGLES (Expert Advisory Group on Language Engineering Standards) (1996), *Preliminary recommendations on Corpus Typology*. Disponible en la web:

<http://www.ilc.cnr.it/EAGLES/corpusstyp/corpusstyp.html>

[consulta: 30 de junio de 2017]

Equipo Pragma (1989), *Esto funciona*. Madrid: Edelsa.

- Equipo Pragma (1988), *Para empezar*. Madrid: Edelsa.
- Fernández, S. (2011), "Nuevos desarrollos y propuestas curriculares. Programar a partir del *MCER*". Suplementos *marcoELE*, 12, enero-junio.
<http://marcoele.com/suplementos/programar-a-partir-del-mcer>
- Fernández, S. (2003), *Propuesta curricular y Marco común europeo de referencia. Desarrollo por tareas*. Madrid: MEC y Edinumen.
- Firth, J.R. (1957), *Papers in Linguistics 1934-1951*. London: Oxford University Press.
- Fulcher, G. (2010), *Practical Language Testing*. London: Hodder Education.
- García Santa-Cecilia, Á. (2007), "Introducción general" en *Plan Curricular del Instituto Cervantes. Niveles de referencia para el español*. Madrid: Biblioteca Nueva. Disponible en la web:
http://cvc.cervantes.es/ENSENANZA/biblioteca_ele/plan_curricular/introduccion.htm
 [consulta: 30 de junio de 2017]
- Gil, M. (2005). *La escritura académica en el aula de español, lengua extranjera*. Tesis doctoral no publicada.
- Green, A. (2012), *Language Functions Revisited. Theoretical and Empirical Bases for Language Construct Definition Across the Ability Range*. Cambridge: Cambridge University Press.
- Harrison, J. (2015), "What is English Profile?" en Harrison, J. y Barker, F. (2015), *English Profile in Practice*. Cambridge: Cambridge University Press.
- Harrison, J. & Barker, F. (2015), *English Profile in Practice*. Cambridge: Cambridge University Press.
- Higueras, M. (2006), *Las colocaciones y su enseñanza en la clase de ELE*. Madrid: Arco Libros.
- Howkins, J. & Filipovic, L. (2012), *Criterial Features in L2 English*. Cambridge: Cambridge University Press.
- Hulstijn, J.H., Alderson, J.C., & Schoonen, R. (2010), "Developmental stages in secondlanguage acquisition and levels of second-language proficiency: Are there links between them?" en Bartning, I., Martin, M., & Vedder, I. (eds.), (2010), *Communicative proficiency and linguistic development: intersections between SLA and language testing research (pp. 1120)*. *Eurosla Monographs Series*, 1. Disponible en la web:
<http://eurosla.org/monographs/EM01/EM01home.html>
 [consulta: 30 de junio de 2017]
- Instituto Cervantes (2007), *Plan Curricular del Instituto cervantes. Niveles de referencia para el español*. Madrid: Biblioteca Nueva. Disponible en la web:
http://cvc.cervantes.es/ENSENANZA/biblioteca_ele/plan_curricular/default.htm
 [consulta: 30 de junio de 2017]
- Instituto Cervantes (1994), *Plan Curricular del Instituto Cervantes*. Publicaciones internas.
- Lewis, M. (2000), *Teaching Collocation. Further Developments in the Lexical Approach*. Londres: The Language Teaching Publications.

Lewis, M. (1993), *The Lexical Approach*. Londres: Language Teaching Publications.

Llorián, S. (2016), "La descripción empírica de los niveles de referencia del español a partir de corpus de aprendientes" en *Revista Nebrija de lingüística aplicada a la enseñanza de lenguas*, 20. Disponible en la web:

<http://www.nebrija.com/revista-linguistica/la-descripcion-empirica-de-los-niveles-de-referencia-del-esp%C3%B1ol-a-partir-de-corpus-de-aprendientes.html>

[consulta: 30 de junio de 2017].

Llorián, S. (2012), "Los retos de la certificación del español con fines específicos" en *Revista Nebrija de lingüística aplicada a la enseñanza de lenguas*, 12. [Disponible en la web:

www.nebrija.com/revista-linguistica/files/articulosPDF/articulo_527b6a3c4a626.pdf

[consulta: 30 de junio de 2017].

Llorián, S. (2007), *Entender y utilizar el Marco común europeo de referencia desde el punto de vista del profesor de idiomas*. Madrid: Santillana.

Lozano, C. (2015), "Learner corpora as a research tool for the investigation of lexical competence in L2 Spanish", en *Journal of Spanish Language Teaching*, 2: 180-193.

Matte Bon, F. (1992), *Gramática comunicativa del español*. Madrid: Edelsa.

Miquel, L. y Sans, N. (1989), *Intercambio 1, 2*. Barcelona: Difusión.

Nation, I.S.P. (2001), *Learning Vocabulary in an other Language*. Cambridge: Cambridge University Press.

Nelson, M. (2010), "Building a written corpus: what are the basics?" en O'Keefe, A. & McCarthy, M. *The Routledge Handbook of Corpus Linguistics*. New York: Routledge.

North, B. (2014), *The CFR in Practice*. Cambridge: Cambridge University Press.

O'Keefe, A. & McCarthy, M. (2010), *The Routledge Handbook of Corpus Linguistics*. New York: Routledge.

O'Keefe, A., McCarthy M. & Carter, R. (2007), *From corpus to classroom*. Cambridge: Cambridge University Press.

Parodi, G. (2015), "Corpus de aprendices de español (CAES)" en *Journal of Spanish Language Teaching*, 1.

Parodi, G., Julio, C. & Vásquez, L. (2015), "Los géneros del Corpus PUCV-UCSC-2013 del discurso académico de la economía. El caso del Informe de Política Monetaria." en *Revista Latinoamericana de estudios del discurso*, 15(2): 179-200. Disponible en la web:

<https://dialnet.unirioja.es/servlet/articulo?codigo=5958949>

[consulta: 30 de junio de 2017].

Pérez Serrano, M. (2016), "Tratamiento de la combinatoria léxica en documentos de referencia y curriculares: el caso del MCER y del PCIC" en *Revista de Investigación Lingüística*, 18: 213-232.

Salamoura, A. & 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: 101–132.

Sánchez, A. (1997), *Los métodos en la enseñanza de idiomas. Evolución histórica y análisis didáctico*. Madrid: SGEL.

Scott, M. (2010), "What can corpus software do?". En O'Keeffe, A. y McCarthy, M. *The Routledge Handbook of Corpus Linguistics*. New York: Routledge.

Searle, J. (1969), *Speech Acts: An essay in the Philosophy of language*. Cambridge: Cambridge University Press.

Slagter, P.J. (1979), *Un nivel umbral*. Edición de Tolosa. J. y Yagüe, A. (eds.) (2007). *Monográficos marcoELE*, 5. Disponible en la web: <http://marcoele.com/monograficos/un-nivel-umbral> [consulta: 30 de junio de 2017]

Swales, J. (1990), *Genre Analysis*. Cambridge: Cambridge University Press.

Timmis, I. (2015), *Corpus Linguistics for ELT: Research and Practice*. New York: Routledge.

Trim, J. (2010), "Preface". En Green, A. (2012). *Language Functions Revisited. Theoretical and Empirical Bases for Language Construct Definition Across the Ability Range*. Cambridge: Cambridge University Press.

Van Ek, J. (1976), *Threshold Level for Modern Language Learning in Schools*. Wolters-Noordhoff: Longman.

Van Ek, J. & Trim, J.L.M. (1990), *Threshold 1990*. Cambridge: Cambridge University Press.

Van Ek, J. & Trim, J.L.M. (1990), *Waystage 1990*. Cambridge: Cambridge University Press.

Van Ek, J. & Trim, J.L.M. (2001), *Vantage*. Cambridge: Cambridge University Press.

Weir, C.J. (2005a), *Language testing and validation: An evidence-based approach*. Basingstoke: Palgrave MacMillan.

Weir, C.J. (2005b), "Limitations of the Common European Language" en *Language Testing* 22 (3); 281–300.

Wilkins, D. (1976), *Notional Syllabuses*. Oxford: Oxford University Press.

Wilkins, D. (1972), *Linguistics in Language Teaching*. Londres: Edward Arnold.

FECHA DE ENVÍO: 10 DE JULIO DE 2011

FECHA DE ACEPTACIÓN: DÍA 18 DE AGOSTO DE 2017