Procesos de programación curricular del español con fines específicos asistidos por corpus

Inmaculada Martínez Martínez
Centro Internacional de Estudios Superiores del Español (CIESE-Comillas)

Susana Llorián González
Universidad Complutense de Madrid

1 Introducción

La aplicación de la lingüística de corpus a la didáctica de lenguas extranjeras y segundas, en general, y de fines específicos, en particular, ha sido una constante a lo largo de los últimos 15 años. No en vano Rundell y Stock (1992) acuñaron el término *revolución de corpus*, cuyo impacto ha sido fundamental y duradero.

Los corpus son herramientas potentísimas, no solo en la investigación, sino también en su aplicación didáctica en el aprendizaje de español con fines específicos (EFE). Debemos comenzar por preguntarnos, entonces, por qué, a diferencia del inglés con fines específicos, en donde los corpus son herramientas cotidianas, tanto para la perspectiva investigadora como para la pedagógica, nos resulta difícil hablar de la misma situación en el caso de la didáctica del discurso especializado en español (Parodi 2005). A este respecto, debe señalarse que, en el caso del inglés, el uso de corpus se ha mostrado especialmente fructífero en relación con el componente metodológico del currículo y también, aunque en menor medida, con el de la evaluación.

Todo ello provoca el que sean inexistentes los trabajos dedicados al uso de los corpus para guiar el desarrollo de sílabos en los currículos de enseñanza de EFE. En este artículo se describe el proceso de elaboración de un corpus de español de los negocios, así como los procesos de programación curricular centrados en el desarrollo de sílabos de español con fines específicos, en los que se combina el uso de corpus de hablantes nativos con el de aprendientes en constante interacción. Estos procesos abarcan las fases que van desde la elaboración de estos corpus hasta su interrogatorio para seleccionar las unidades que configurarán el sílabo de estos cursos de español. Con ello, se consigue un segundo objetivo importante: ofrecer una metodología protocolizada de compilación de corpus de especialidad (Seghiri 2011) que pueda servir de referencia a otros corpus especializados e, igualmente, al diseño curricular de los cursos de español.

2 Fundamentos y antecedentes

2.1 Concepto y caracterización de un corpus especializado

Partimos de la definición inserta en la que se considera la obra fundacional de la lingüística de corpus, Sinclair (1991), completada con nuevas aportaciones en 2005. Tendremos en cuenta esta definición para conseguir la caracterización de un corpus textual, en primer lugar, y de un corpus textual con fines específicos, en segundo. Así pues, señala Sinclair que un corpus es "un conjunto de (fragmentos de) textos en formato electrónico, seleccionados conforme a criterios externos para representar, en la medida de lo posible, una lengua o variedad de lengua como fuente de datos para la

investigación lingüística" (1991, 8; traducción nuestra)[1]. Los rasgos que se derivan de esta definición y que, a nuestro juicio, deben caracterizar un corpus de discurso especializado son los siguientes:

1. Ha de estar codificado en formato electrónico. Los textos que forman un corpus deben estar codificados de modo que sea posible lograr la recuperación selectiva de la información para su estudio posterior. Como señala Rojo, no sería viable de otra forma "extraer con rapidez y comodidad los casos de, por ejemplo, una determinada expresión en un corpus de trescientos millones de formas" (2021, 2). Y, como recuerda también Sinclair, dichos textos han de estar "documentados conforme a su origen y procedencia"[2] (1996, 4; traducción nuestra), para lo cual es preciso incorporar los datos que caracterizan cada uno de los textos incluidos en el corpus, los llamados *metadatos*[3].
2. Ha de abarcar géneros orales y escritos en situaciones naturales. Es absolutamente necesario observar el lenguaje de estudio tal y como se utiliza en sus contextos específicos (Vargas 2006). Observar el lenguaje ―y, por tanto, las palabras, para conocer su comportamiento― significa observarlo en su entorno natural; no olvidemos, en este sentido, que el entorno natural de las palabras son los textos. Antes de su inclusión en el corpus, los textos (o *fragmentos de textos*, como recuerda Sinclair 1991) fueron creados como géneros discursivos escritos u orales. Es esencial, por tanto, integrar los textos generados "con la intención real de comunicar algo" (Rojo 2021, 1).
3. Ha de ser representativo y equilibrado y para conseguirlo resultan decisivos los criterios externos que se delimiten (Biber 1993), tales como el tamaño, el periodo de tiempo que abarca o el peso porcentual de las áreas geográficas implicadas, entre otros. El análisis del conjunto de textos integrados en un corpus debe dar una visión adecuada de aquello que pretende representar, es decir, ha de ser representativo. Asimismo, debe estar equilibrado, esto es, debe contener un número suficiente de textos de los diferentes tipos que integra (textos orales y escritos) y de los diferentes eventos. Es importante recordar que tanto la representatividad como el equilibrio son valores "de imposible consecución, pero a los que hay que tender" (Rojo 2021, 69).
4. Ha de facilitar el estudio de la lengua de referencia. Un corpus debe poseer como objetivo contribuir a "facilitar el estudio" (Rojo 2021, 63), lo que da lugar al llamado análisis de corpus, tanto con una finalidad científica (por ejemplo, para un mejor conocimiento de la terminología utilizada en el español de los negocios) como con una finalidad aplicada, orientada hacia la metodología, la evaluación o el diseño curricular.
5. Ha de representar una variedad de lengua. Por último, la variedad de lengua que aspira a representar nuestro corpus se enclava dentro de los llamados discursos especializados, entre los que Ciapuscio y López Ferrero consideran al "discurso científico, el académico y el técnico-profesional" (2022, 268). Como Cabré nos recuerda también, un corpus de especialidad integra "las producciones lingüísticas, orales y escritas, que se producen en escenarios de comunicación profesional y sirven exclusivamente a una finalidad profesional" (2007, 90).

[1] Cita original: "A corpus is a collection of pieces of language text in electronic form, selected according to external criteria to represent, as far as possible, a language or language variety as a source of data for linguistic research" (Sinclair 1991, 8).

[2] Cita orginal: "Its constituent pieces of language are documented as to their origins and provenance" (Sinclair 1996, 4).

[3] Los metadatos se refieren a aquellos *datos que hablan de los datos*, es decir, que describen el contenido de los archivos. Así, por ejemplo, si el archivo codificado pertenece a una noticia de prensa, los metadatos contendrán la información acerca del periódico en que fue publicado: país, fecha o nombre del autor, entre otros. Su inclusión hace posible, en palabras de Guillermo Rojo, "su estudio científico y, más concretamente, la recuperación selectiva de la información que contiene" (2021, 2).

2.2. Elaboración de un corpus de discurso especializado

La posibilidad de aplicar los corpus textuales a cuestiones relacionadas con la enseñanza y el aprendizaje de segundas lenguas se hizo presente muy pronto en la lingüística de corpus, especialmente en Europa. Fundamental es, por ejemplo, el proyecto COBUILD[4] para el inglés, dirigido por Sinclair en los años ochenta del siglo pasado.

Otra aplicación tiene que ver con los llamados corpus de aprendices o corpus de no nativos. El propósito de estos corpus es proporcionar un recurso en el que sea posible analizar, con las técnicas utilizadas habitualmente en lingüística de corpus, las características de la interlengua de los estudiantes: los errores que contienen, el modo en que evoluciona su conocimiento, los componentes en los que hay interferencia con su lengua materna, entre otros aspectos. Bien conocido es, en este sentido, el proyecto CAES (Corpus de Aprendices de Español como Lengua Extranjera), compilación de textos escritos producidos por estudiantes de español con diferentes niveles de dominio y diversas lenguas maternas.

El proceso de elaboración de un corpus abarca desde su diseño hasta su análisis en lo que constituye ya la explotación o interrogación al propio corpus. Destacaremos tan solo algunos aspectos esenciales de estas etapas en la Figura 1.

En la fase preliminar del diseño se fijan los parámetros que caracterizarán los textos, los rasgos que luego serán utilizados en el análisis y con ellos se crea la DTD (*Document Type Definition*), que integra la cabecera en la que se incluyen estos rasgos y el propio texto codificado. En nuestro caso —y como rasgo identificativo de nuestro corpus, pues no lo hemos encontrado en proyectos similares— destacamos los géneros discursivos como uno de sus parámetros esenciales. Pretendemos que los programas didácticos nacidos a partir de nuestro corpus especializado conviertan a los estudiantes de EFE —en nuestro caso, de español de los negocios— en miembros competentes de la comunidad discursiva correspondiente que "construyen, interpretan y usan estos géneros para lograr sus metas comunicativas" (Bhatia 2002, 6).

La segunda fase está integrada por la adquisición y compilación textual. La búsqueda y selección de los textos que integran el corpus se lleva a cabo conforme a los criterios de representatividad y equilibrio ya comentados. Se consideran también los criterios externos señalados por Sinclair y mencionados líneas más arriba.

El soporte físico será el que dará cobertura a la codificación. Se ha elegido el lenguaje de codificación más extendido en los últimos años, XML[5], a través del programa Oxygen. La cabecera, como zona especial diferenciada del cuerpo del texto, está destinada a contener los metadatos, lo cual garantiza que las aplicaciones de consulta sepan dónde deben ir a buscar cada elemento de información y cómo deben interpretarlo.

---

[4] COBUILD es un acrónimo que se corresponde con *Collins Birmingham University International Language Database*. Se trata de un laboratorio experimental británico desarrollado en la Universidad de Birmingham en 1980 y fundado por la editorial Collins. Este laboratorio fue dirigido por el profesor Sinclair y destaca entre sus logros la creación y análisis de un corpus electrónico contemporáneo, el corpus Collins, que más tarde condujo al desarrollo del *Bank of English* y de numerosos diccionarios y manuales de gramática inglesa.

[5] XML se corresponde con la sigla en inglés de *Extensible Markup Language* ("lenguaje de marcas extensible"). Es un metalenguaje extensible de etiquetas desarrollado por el World Wide Web Consortium (W3C), que constituye una simplificación y adaptación del SGML, acrónimo de *Standard Generalized Markup Language* o "lenguaje de señalización general normalizado". Por lo tanto, XML no es un lenguaje particular, sino una manera de definir lenguajes para diferentes necesidades.

Por último, la fase de explotación e interrogación constituye la verdadera aplicación del corpus a todo tipo de fines científicos.

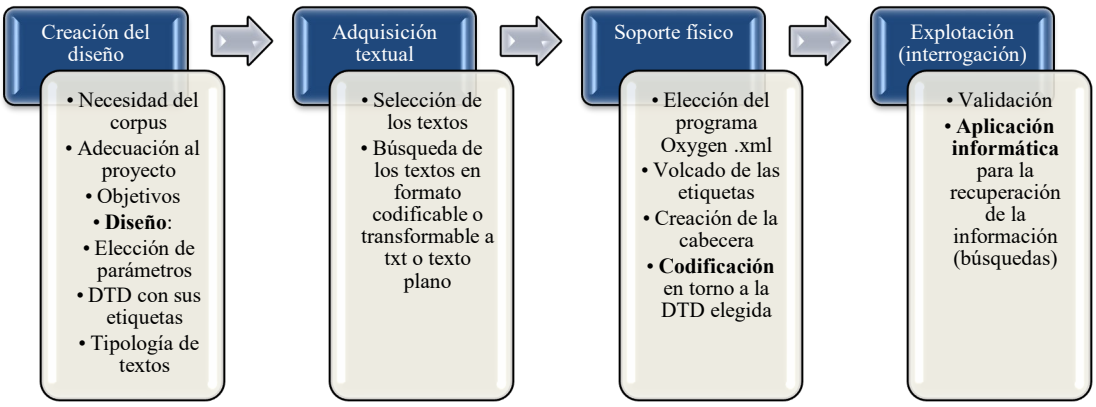

Figura 1. Etapas en la elaboración de un corpus (adaptado de Vargas 2006 y Rojo 2021)

3 El Corpus del Español de los Negocios (CORPEN), corpus de discurso especializado

3.1 Características generales

Todo lo anterior nos lleva a caracterizar CORPEN como un corpus:

1. Mixto, pues integra muestras escritas y orales conforme a eventos de uso. Son, por lo tanto, textos en su dimensión comunicativa real. Entre los escritos figuran el informe, el manual de procedimiento o el orden del día, por citar tan solo algunos géneros; entre los orales, la entrevista profesional, la defensa de un proyecto, una presentación o una reunión, por ejemplo. En este sentido, cabe señalar que no podría llevarse a cabo un plan curricular del español de los negocios sin considerar las interacciones comunicativas orales que se producen diariamente en esta lengua de especialidad. Los porcentajes atribuibles a la división textos escritos y orales deben ser necesariamente distintos a los de los corpus de referencia ya existentes, dada la transcendencia que la lengua oral posee en este contexto específico. Así, la distribución es la que sigue: un 75 % para los textos escritos y un 25 % para los textos orales, ligeramente superior al 10 % que se reserva para estos últimos en ciertos corpus como el British National Corpus (BNC) o en el propio Corpus del Español del Siglo XXI (CORPES). El objetivo último es que refleje la realidad de la lengua de especialidad y que aspire a la neutralidad de la muestra mediante dichos porcentajes.
2. Electrónico. Ya se ha señalado anteriormente que solo un corpus digital permite la recuperación posterior y el análisis ordenado de los datos para todo aquel investigador que se acerque a él con la finalidad de conocer el comportamiento de esta lengua de especialidad. La recuperación y análisis ordenado de datos es posible gracias al sistema de codificación antes aludido, Oxygen XML, programa que nos asegura "la correcta introducción de todas las etiquetas y, en procesos posteriores, la adecuada manipulación de las marcas para conseguir la recuperación selectiva de la información introducida" (RAE 2020, 10). Es esta una fase de gran importancia en el proceso de creación de todo corpus y ha de integrar dos niveles: una cabecera, con los datos bibliográficos y los parámetros que caracterizan al texto electrónico, y

el texto propiamente dicho, con la marcación tipográfica habitual en torno a párrafos y marcas de formato[6]. Para CORPEN se ha creado una cabecera específica, cuyo principal objetivo es la codificación de los textos en un formato estandarizado que permite, como ya se ha mencionado, su tratamiento posterior mediante procesos informáticos.

3. Representativo y equilibrado. El CORPEN se caracteriza, por un lado, por ser representativo de la realidad que quiere reflejar. Para ello, debemos aspirar a la neutralidad de la muestra, mucho más que a la cantidad de textos que ha de componerlo (Vargas 2006, 6). En nuestro caso, trabajamos con muestras proporcionales de todos los aspectos que se consideran esenciales en la realidad lingüística del español de los negocios. Este aspecto concreto de la distribución coherente y controlada en los porcentajes de representación de los textos en CORPEN se debe cuidar de manera muy especial dada su importancia. En su trabajo sobre los criterios básicos para desarrollar un corpus lingüístico, Sinclair (2004) insiste en que todo aquel que acometa la tarea de construir un corpus deberá afanarse por conseguir que este sea lo más representativo posible de la lengua a la que representa. Más adelante añade un dato esencial: los criterios estructurales en los que se fundamente un corpus lingüístico deberán elegirse con sumo cuidado porque las preocupaciones básicas de equilibrio y representatividad dependen de aquellos criterios de carácter estructural (Sinclair 2004, 2-3).
4. De discurso especializado o específico. También llamado "técnico" por otros autores (Rojo 2021), aspira a reflejar el español de los negocios a través de los distintos géneros discursivos que lo integran. La elección de los géneros discursivos que se integran en él constituye el verdadero eje vertebrador de este corpus.

3.2 Proceso de elaboración

El proceso de elaboración de CORPEN toma las referencias anteriores como punto de partida, sobre todo, en la medida en que establece el género discursivo como unidad de análisis. Lo primero que hemos hecho es desarrollar una matriz de textos asociados a los géneros discursivos que tendrán que procesar los destinatarios del currículo durante los eventos de comunicación. Una vez desarrollada, la matriz entra en proceso de validación a través de la metodología de jueces expertos.

Un panel constituido por profesionales del sector, profesores de español para los negocios y una muestra de potenciales alumnos verifican o refutan las hipótesis que han guiado la elaboración de la matriz. El resultado se somete a una consulta a gran escala, que permite seleccionar la tipología de textos que alimentará el corpus a lo largo de las fases de su desarrollo para ordenarlo después en torno a criterios de prioridad y dificultad.

Como se verá inmediatamente, los parámetros de la matriz se vuelcan en la cabecera del corpus y, a partir de aquí, se empezará a recopilar los textos y a su alimentación, con la consiguiente codificación y etiquetado. Finalmente, se interroga el corpus para desarrollar el currículo, tal y como se describirá más abajo.

La matriz de los textos se ha configurado con arreglo a doce categorías cuidadosamente seleccionadas[7]. Los criterios empleados para la selección responden a los propósitos hacia los cuales se orientan las aplicaciones de CORPEN. Nos limitamos

[6] Las marcas tipográficas, como la cursiva, se introducen mediante una etiqueta. Las etiquetas delimitan, por tanto, "fragmentos de texto que tienen una determinada característica" (Rojo 2021, 97).

[7] Evento, actividad; géneros discursivos (macrogéneros); canal; soporte habitual; actividades comunicativas de la lengua; fuentes; sector empresarial; modos de organización del discurso; participantes; códigos de relevancia (densidad del discurso especializado); registro; modalidad.

aquí a la especificación de los componentes del currículo de ENE a lo largo de los niveles de adaptación curricular. Estas categorías definen la arquitectura del corpus y, por lo tanto, permiten que este sea interrogado en relación con los mencionados propósitos. En este caso, se pretende que los resultados de la interrogación al corpus den lugar a una especificación detallada del material lingüístico y no lingüístico objeto de aprendizaje, enseñanza y evaluación en los cursos y/o exámenes certificativos de ENE. Esto se constituirá en la base para la elaboración de materiales didácticos y para el diseño de programas de ENE.

La primera categoría representa el punto de partida para el desarrollo de la matriz y es el evento de comunicación, es decir, el contexto situacional en el que interactúan en español los miembros de las comunidades profesionales de las que pasarán a formar parte los destinatarios de los cursos y exámenes de ENE. En esta misma dimensión contextual se sitúan otras tres categorías: los participantes en el evento, las fuentes de las que se han extraído los textos y el sector empresarial en el que se inscriben. Dentro de los confines de estos eventos resulta factible la identificación de los géneros discursivos, principal categoría, y también el canal (oral o escrito), además del soporte a través del cual se transmiten los textos.

En una dimensión programática y didáctica, conviene que se puedan identificar las actividades comunicativas de la lengua que se ejecutan durante el procesamiento de los textos, a saber, comprensión, expresión, interacción y mediación.

El resto de las categorías aportan información de utilidad para el análisis de los géneros: los modos de organización del discurso que dominan (narración, exposición, argumentación y descripción) o los códigos de relevancia (Knoch y McQueen 2020), que resultan determinantes para estimar la densidad del contenido especializado que transportan los textos. Estos códigos se definen a partir del grado de conocimiento especializado que compartan los interlocutores. Por ejemplo, la densidad de contenido especializado difiere en virtud de si la comunicación tiene lugar entre un vendedor y un cliente o si se da entre técnicos en publicidad y márquetin que hablan sobre un mismo producto. Finalmente, se consideran el registro, es decir, el grado de formalidad, y el tipo de comunicación, interna o externa, en el contexto de empresa.

Cada una de estas categorías se compone de una lista de parámetros como los modos de organización del discurso o las clases de actividades comunicativas de la lengua antes mencionados. Estas listas de elementos son los que se vuelcan en la cabecera de CORPEN y también en la del corpus de hablantes no nativos con el fin de que los dos corpus se construyan con la misma arquitectura.

4 El diseño del currículo de ENE asistido por corpus

La aplicación que se prevé como prioritaria y más importante de CORPEN consiste, como se ha venido indicando, en prestar asistencia al desarrollo de un currículo del español para los negocios, con el fin de que el material lingüístico susceptible de ser enseñado, aprendido y evaluado por parte de los destinatarios de los cursos de ENE se seleccione a partir de fuentes auténticas y, por lo tanto, el currículo pueda construirse sobre bases empíricas. Para lograr este objetivo, se ha diseñado un proceso en el que el corpus CORPEN entra en interacción con una segunda herramienta: otro corpus, en este caso, de hablantes no nativos. Este segundo corpus se alimenta de textos producidos por aprendientes de ENE, que son el resultado de la realización de las tareas basadas en los géneros discursivos que son significativos para los miembros de las comunidades

discursivas destinatarias del currículo. Las tareas se elaboran sobre la base de la misma matriz utilizada para la elaboración de CORPEN.

El proceso prevé que los dos corpus se utilicen para la especificación de los cuatro componentes curriculares. No obstante, nos centraremos aquí en los contenidos, en las orientaciones metodológicas y en los procedimientos de evaluación, que, de forma indirecta, repercuten en la formulación de los objetivos. Se prevé que ambos corpus se utilicen para la especificación a lo largo de todo el proceso de adaptación del currículo.

CORPEN, como corpus de hablantes nativos, permitirá investigar la lengua real que se utiliza en los contextos profesionales del ámbito de los negocios y de la empresa. El análisis y la descripción toman como principio organizativo el género discursivo. El uso de esta herramienta permitirá verificar las intuiciones que guían las especificaciones del contenido previas al uso del corpus, además de descubrir y evidenciar nuevos fenómenos y características, que completarán las hipótesis iniciales.

Gracias al corpus de los hablantes no nativos, se pueden identificar los llamados "rasgos criteriales" (Hawkins y Filipović 2012, 37), es decir, el conjunto de propiedades de la lengua que son características de un determinado grado de dominio y que distinguen unos niveles de los otros; marcan, en definitiva, el paso de un estadio a otro en el proceso de aprendizaje. Estos rasgos se constituyen en la base empírica para la fijación de los niveles de grado de dominio que articularán la estructura vertical del currículo y que, por lo tanto, servirán para secuenciar los contenidos.

Para describir la aplicación de los dos tipos de corpus a la especificación de los contenidos, la metodología y la evaluación a lo largo del desarrollo del currículo, nos serviremos del esquema de niveles en los que Beacco *et al.* (2016) clasifican las actividades implicadas en el proceso de diseño e implementación curricular. Los niveles se encuentran en permanente interacción y se alimentan unos a otros. El proceso se extiende desde el nivel nano al nivel supra o viceversa a lo largo de cinco estadios[8] de adaptación.

El nivel *supra* quedaría, por el momento, fuera del ámbito de las aplicaciones de CORPEN y del corpus de hablantes no nativos al desarrollo del currículo, puesto que la concreción del currículo en este nivel no tiene carácter programático. De hecho, el nivel supra se sitúa en el plano de los estándares internacionales. El volumen complementario del *Marco común europeo de referencia* (Consejo de Europa 2020) es un ejemplo de documento de este nivel. Los documentos y proyectos que se desarrollan en este nivel contribuyen a la armonización entre sistemas de aprendizaje, enseñanza y evaluación de diferentes países o regiones. En fases posteriores del desarrollo del proyecto del currículo del español de los negocios, no se puede descartar el desarrollo de este tipo de documentos o proyectos en que se describan, por ejemplo, estándares de dominio de las lenguas extranjeras del ámbito de los negocios, entre otras, el español.

En la actualidad el principal objetivo del proyecto que aquí se describe consiste en el desarrollo del Plan Curricular del Español de los Negocios de la Fundación Comillas. Pese a tratarse de un documento cuya elaboración parte de un proyecto institucional, pretende constituirse en una referencia a escala internacional para la elaboración de planes de aprendizaje y enseñanza, además de sistemas de evaluación y certificación en ENE. A diferencia de los proyectos y documentos que se localizan en el nivel supra, este documento tiene carácter programático, por lo que se puede equiparar a Beacco *et al.* (2016), ya que estos se sitúan en el nivel *macro*, aunque sus aplicaciones trasciendan los límites nacionales o regionales. Como documento curricular de referencia, integraría declaraciones acerca del enfoque adoptado, indicaciones acerca de las pautas de

[8] Nivel *supra*, nivel *macro*, nivel *meso*, nivel *micro*, nivel *nano*.

desarrollo del currículo, líneas de investigación y, por supuesto, en la dimensión programática, los cuatro componentes curriculares: objetivos, contenidos, orientaciones metodológicas y criterios y procedimientos de evaluación.

Las aplicaciones de este documento a los currículos institucionales que lleven a cabo las organizaciones que ofrecen servicios de enseñanza de ENE constituirán las múltiples concreciones que se darían en un nivel *meso*. En este mismo marco, las adaptaciones a cada grupo de aprendientes que se hagan de los programas, de los manuales de aprendizaje y enseñanza y de otros recursos se ubican en un nivel micro. No obstante, el currículo no se llega a concretar del todo hasta el momento en el que cada uno de sus destinatarios, a título individual, se involucra, en el nivel *nano*, en las experiencias de aprendizaje.

La especificación de los componentes curriculares ―objetivos, contenidos, orientaciones metodológicas y procedimientos de evaluación― está determinada por el género discursivo como principio articulador en todos los niveles de concreción curricular que se acaban de describir, lo cual está en perfecta consonancia con la arquitectura de los dos corpus creados *ad hoc* para prestar asistencia a todo este proceso. El género discursivo es en sí mismo un contenido del currículo, al cual se subordina el resto de los elementos que se inventarían en el documento con el fin de describir el material susceptible de ser aprendido y, por lo tanto, enseñado y evaluado (unidades fraseológicas, marcadores del discurso y del metadiscurso, categorías morfosintácticas, etc.). El planteamiento que rige la especificación de los contenidos va, por lo tanto, de unidades mayores a unidades menores.

Del análisis de los textos correspondientes a cada género (es decir, correos electrónicos, informes, negociaciones de ventas, etc.), se extraería una macroestructura prototípica, o sea, la parte del texto y la forma en que estas partes se integran en los modos de organización del discurso (descripción, argumentación, etc.). En el plano de la microestructura se identifican los recursos que se utilizan en cada género para apuntalar las propiedades textuales de coherencia y cohesión como, por ejemplo, los marcadores del discurso.

En la especificación del componente metodológico el género, de nuevo, asume un papel protagonista. Este se erige en el elemento al que se orientan los objetivos de las tareas, que se basarán en el procesamiento de los textos, asociados a los géneros, que sean significativos de los destinatarios del currículo de ENE en los diferentes contextos de aprendizaje, enseñanza y evaluación. Por lo tanto, el género será también la base para especificar los elementos del sílabo de los procedimientos de evaluación y de las tareas que se empleen con este propósito.

5 La especificación de los componentes del currículo

En los siguientes apartados se describe el modo en el que los dos corpus, CORPEN y el corpus de hablantes no nativos, se utilizan para especificar los componentes del currículo a lo largo de los niveles de adaptación arriba descritos. De todos ellos, nos centraremos de manera particular en el nivel de concreción macro, puesto que, como hemos señalado, es aquel en el que se ubica el proyecto Plan Curricular del Español de los Negocios de la Fundación Comillas.

5.1 Las especificaciones curriculares en el nivel *macro*

La especificación de los contenidos en un documento localizado en este nivel de concreción, por todo lo dicho hasta aquí, se fundamenta en el género discursivo como unidad de análisis. A este se subordinarán todas las unidades de análisis que se seleccionen para describir el material lingüístico, y no lingüístico, presentado en el documento en forma de inventarios de contenido (gramática, unidades fraseológicas, etc.).

Para identificar los géneros objeto de descripción, será necesario partir de los eventos comunicativos en los que se verán involucrados los miembros actuales o potenciales de las comunidades discursivas a los que se destina el currículo de ENE. En este punto, entra en juego la matriz de textos descrita más arriba, gracias a cuyas categorías se pueden señalar los elementos del contexto situacional en el que se da la comunicación, es decir, participantes, canal, propósitos comunicativos, etc. Sirva como ejemplo el género de la convocatoria de reunión. Si se describe a partir de la matriz de textos, se concibe como una comunicación interna, del ámbito de la empresa, que se transmite por escrito normalmente, a través del correo electrónico, de convocante a convocados, en un registro formal, con baja densidad de contenido especializado porque se utiliza un repertorio propio del lugar de trabajo. En el proceso de redacción de este tipo de texto, se combinan modos de organización del discurso de tipo expositivo instruccional.

A continuación, se empleará CORPEN para prestar asistencia al análisis lingüístico de los textos que se corresponden con los géneros. Nuestra propuesta de análisis se fundamenta en los dos enfoques propuestos por Biber *et al.* (2007). Uno de ellos opera de arriba abajo (*top-down*) y el otro, en la dirección opuesta (*bottom-up*). El primero requiere una estructuración de los textos *a priori*. Por ejemplo, la identificación de los movimientos, los pasos y las funciones retóricas que articulan las convocatorias de reunión (asunto, lugar, fecha y hora, orden del día, documentación, etc.). Hecho esto, habría que etiquetar el corpus a partir de este tipo de criterio y, finalmente, realizar las búsquedas para, de forma automática, identificar los patrones y los rasgos característicos de cada una de las partes de una convocatoria que se desarrolle en el ámbito de ENE. En el enfoque *bottom-up* son los datos los que emergen del corpus. Se necesita un software que haga posible el análisis automático de los textos, a partir de lo que Biber *et al.* (2007) llaman "unidad del discurso basada en el vocabulario" (*vocabulary-based discourse units*). Este tipo de análisis se realiza mediante cálculos computacionales y comparaciones de las unidades léxicas del texto. Las operaciones descritas hacen posible diferenciar de forma automática las partes del texto y su segmentación. El paso siguiente consiste en la extracción de la estructura, en su descripción y en la identificación de tendencias de aparición de patrones y regularidades, sobre todo, léxicas y gramaticales.

Según este tipo de planteamiento, se establecen varias fases, la primera de las cuales es comparativa. Consiste en el cotejo de los textos del ámbito de los negocios con otros característicos de otros registros. Siguiendo el hilo del ejemplo, podrían contrastarse las convocatorias de reunión del contexto de los negocios con las de otros ámbitos diferentes. La segunda fase es de tipo cuantitativo. Se trataría de identificar en cada uno de los registros analizados el porcentaje de combinaciones sintagmáticas de determinadas características, el rango de aparición de pronombres personales, etc. La tercera fase del análisis lingüístico se orienta a la obtención de una muestra representativa del género de la convocatoria de reunión en el ámbito de las comunicaciones internas de empresas en las que se desenvuelven los destinatarios de los

currículos. Finalmente, se buscarían las causas que pudieran explicar que una unidad fraseológica, como *orden del día*, respondiera en este contexto a determinadas características funcionales como la de establecer la secuenciación del contenido del evento de la reunión, marcado por la prioridad frente a otro ámbito en el que pudiera regirse hipotéticamente por la jerarquía que ocupa el responsable del asunto en cuestión en el organigrama de la organización.

El análisis del género asistido por corpus permite, además, poner en evidencia otros aspectos, como el modo en que se reflejan en los textos las marcas de relaciones sociales, poder, jerarquía, etc. De igual modo, la aplicación de estos procedimientos puede identificar regularidades o confluencias de rasgos léxico-gramaticales, así como los elementos del metadiscurso (Hyland 2005; Hyland 2017), que se constituyen en patrones indicadores de las identidades disciplinares o corporativas de los miembros de las comunidades discursivas a las que se asocian los géneros.

Una vez que los géneros quedan descritos, se procede a la especificación de los contenidos propiamente dicha. Para ello, es preciso adoptar un enfoque programático debido al uso que se le va a dar, es decir, la elaboración de programas en el nivel *meso* de concreción curricular o el desarrollo de materiales didácticos. Los contenidos tienen que ser inventariados, secuenciados y distribuidos en niveles o grados de dominio. Para esto último, se requiere que CORPEN, como se ha venido señalando, interactúe con el corpus de hablantes no nativos, alimentado con textos producto de la realización de tareas estandarizadas basadas en el procesamiento de los géneros discursivos que se han incorporado a CORPEN. Por ejemplo, se introducirían textos que generen los aprendientes usuarios de ENE a partir de la ejecución una tarea en la que se les pida redactar una convocatoria de reunión, sirviéndose de modelos de texto, y focalizadas en las formas gramaticales, léxicas, discursivas, etc. que haya revelado el análisis de este género. Gracias a este segundo corpus, se podrán identificar los rasgos criteriales, es decir, los patrones léxicos, gramaticales o discursivos característicos de cada estadio de desarrollo de la competencia lingüística comunicativa en español para los negocios.

La especificación de los contenidos que se desprenda del análisis de los géneros asistido por corpus precisa una toma de decisiones acerca de los tipos de inventario (léxicos, gramaticales, discursivos, etc.) en los que se van a articular. El inventario del currículo cuyo desarrollo puede resultar más fructífero desde la perspectiva de la aplicación de los corpus a estos procesos es el léxico-semántico. No en vano las utilidades de un corpus de ENE como CORPEN pueden consistir en guiar la elaboración de listas, glosarios, etc., que logren identificar el vocabulario técnico, semitécnico y general (Alcaraz 2000), de frecuente aparición en los textos orales y escritos propios de un ámbito de especialidad como ENE. Estas listas, al estar destinadas a fines didácticos, y no terminológicos o lexicográficos como la mayor parte de las existentes, se compondrían de unidades léxicas, mono- y pluriverbales, es decir, palabras simples y unidades fraseológicas (colocaciones, locuciones idiomáticas, etc.).

Anteriormente apuntamos el modo de extraer los patrones gramaticales de los textos con la ayuda del corpus y el modo en que estos patrones contribuyen a la descripción de los géneros discursivos. En este sentido, CORPEN informará cómo las palabras se combinan en los textos del ámbito de ENE para configurar patrones gramaticales que escapan a la intuición (O´Keeffe *et al*. 2007). En este caso, resulta especialmente relevante la relación que se establece entre la gramática y el vocabulario (Timmis 2015). Por lo demás, y gracias a un corpus como CORPEN, construido a partir de géneros, se podría verificar, por ejemplo, si son, en efecto, frecuentes las construcciones condicionales con *como* + *subjuntivo* en la correspondencia electrónica de correos electrónicos que vehiculen cartas para expresar la función retórica de “amenaza ante

incumplimientos" (por ejemplo: *Como no entreguen la mercancía dentro del plazo*…). Este sería el tipo de especificación que podría figurar en un documento curricular de nivel *meso* de ENE en el inventario correspondiente al análisis gramatical.

Tratamos, por último, el inventario resultante de un análisis pragmático discursivo. La unidad de análisis por excelencia a efectos programáticos ha sido, tradicionalmente, la función comunicativa, acuñada por el Consejo de Europa a mitad de los años setenta. Para Richards (2013), las funciones suponen la aplicación de la teoría de los actos de habla, las teorías de Wilkins y las de Hymes respecto a la competencia comunicativa. Gracias a las listas de funciones comunicativas que presentan los documentos primeros de la serie Nivel Umbral, T-Level Series (Van Ek 1976), se hizo posible canalizar en los programas de lenguas en Europa los principios comunicativos en sus primeros desarrollos. Pese al indiscutible impacto que ha tenido y sigue teniendo este concepto, el enfoque aquí adoptado precisa que se trascienda en favor de otros más rentables y coherentes con el constructo del género discursivo como la función retórica o paso, según la terminología de Swales (1990; 2004). Algunos de los exponentes lingüísticos que se derivan de la aplicación de este tipo de "molde" quedarían asimilados a diferentes tipos de fraseologismos. Por lo tanto, quedarían recogidos en los inventarios léxico-semánticos.

Uno de los componentes curriculares que, en este nivel macro, se ven favorecidos por la aplicación de los dos corpus es el de la evaluación. De la interacción de CORPEN con el corpus de hablantes no nativos se obtendría lo siguiente. En primer lugar, se haría posible la descripción del constructo teórico o, en otras palabras, la habilidad que es objeto de evaluación. En este caso, la competencia comunicativa específica del ámbito del español de los negocios, que se define fundamentalmente en relación con la capacidad de los usuarios para producir y comprender textos orales y escritos asociados a los géneros discursivos característicos de las situaciones de comunicación en que participan los miembros de estas comunidades de práctica, en ejercicio profesional o en formación. Además del constructo teórico, las descripciones de los géneros asistidas por corpus harían posible la especificación del constructo operacionalizado, es decir, la descripción de las pruebas y resto de procedimientos que tendrá que realizar el alumno. Por ejemplo, se podrían describir con precisión las características léxicas y gramaticales de los textos de entrada de las tareas de comprensión de lectura y de comprensión auditiva que alumnos o candidatos realizarían en las pruebas de evaluación, así como los textos de salida de las tareas de expresión e interacción orales y escritas. CORPEN guiaría, en resumidas cuentas, el desarrollo de las especificaciones de las pruebas de evaluación, que se corresponderían necesariamente con las que se realizan en las aulas.

5.2 Los niveles *meso* y *micro*

La concreción curricular en el nivel *meso* se corresponde con la programación de los cursos que se imparten en una determinada organización, en virtud de las características y del perfil del estudiante y, sobre todo, de sus necesidades. Estas serían determinantes para la selección de las situaciones de comunicación y de los géneros discursivos que resultan significativos para el destinatario de los cursos y, por lo tanto, de los contenidos discursivos, gramaticales y léxicos.

Nos hemos referido ya, de forma sucinta, en relación con el componente metodológico, al diseño de tareas basadas en el procesamiento de textos a partir de géneros. Al margen de este criterio para diseñar tareas, resultan especialmente interesantes las posibilidades que ofrece la explotación de un corpus como CORPEN

para crear secuencias didácticas basadas en el aprendizaje guiado por datos, en las que los alumnos pongan en juego procesos cognitivos que les permitan observar y advertir los fenómenos lingüísticos en su contexto, focalizar su atención, concienciarse sobre su funcionamiento y automatizar su uso. Por ejemplo, este tipo de secuencia puede resultar muy rentable para el aprendizaje del léxico, en concreto, de las colocaciones. Las líneas de concordancia permitirían estudiar en qué contexto léxico aparecería en determinados géneros una palabra como, por ejemplo, *cotización*: *cotización al alza*, *cotización a la baja*, etc.

5.3 Los niveles *micro* y *nano*: evaluación

El aprendizaje guiado por datos permite poner a disposición de los aprendientes muestras de lengua auténtica para el aprendizaje de la gramática de manera inductiva, sin reglas y a través de la reflexión (Larsen-Freeman y Long 1994, 153). Este acercamiento inductivo estimula la habilidad para descubrir patrones en la lengua objeto y para formar generalizaciones sobre su forma y uso. Se trata de un proceso que enfrenta al aprendiente, tan directamente como sea posible, con los datos reales de la lengua y hace de él un investigador lingüístico. Por todo ello, debemos señalar que la inclusión —y explotación— de los corpus en la clase de ENE no solo realza el papel de la gramática en la adquisición de una segunda lengua, sino que también nos permite acercarnos a su comprensión y uso a través del importante mecanismo cognitivo de la inducción por parte del estudiante.

En definitiva, "los profesores deben facilitar el acceso directo a los corpus en función de las propias situaciones y objetivos lingüísticos de los alumnos" (O´Keeffe *et al*. 2007, 18) y ser muy conscientes, además, de que en un corpus especializado es probable que el léxico y las estructuras específicas ocurran con unos patrones y una distribución más regular que en la lengua general (O´Keeffe *et al*. 2007, 198). De este modo, el corpus adquiere una importante función correctiva: los aprendientes, al comparar su propia producción con datos de corpus de hablantes nativos expertos, pueden encontrar la ayuda que necesitan para corregir sus propios rasgos de interlengua.

6 Conclusión

En definitiva, gracias al proyecto cuyas líneas maestras hemos presentado, y en cuyo desarrollo está previsto que participe una nutrida y tupida red de profesionales de la enseñanza del español para los negocios, se van a poder describir, especificar y secuenciar sobre bases empíricas los patrones discursivos, gramaticales y léxicos de los géneros discursivos que resultan significativos para los destinatarios de los currículos, así como la forma en la que se van a poner en juego en el contexto del aula.

Esto se hará posible porque la especificación de los contenidos del currículo, que utiliza el género discursivo como principio organizador, estará asistida por dos tipos de corpus: CORPEN y un corpus de aprendientes de ENE. Los dos corpus cumplen funciones diferentes en los distintos estadios del desarrollo del currículo. CORPEN es un corpus de hablantes nativos, desarrollado *ad hoc,* a partir del cual se realiza una descripción de los géneros discursivos declarados como significativos para los destinatarios del currículo. El resultado de esta descripción se emplea para especificar los contenidos del currículo en los materiales curriculares de referencia, localizados en un nivel *macro*, y en los programas de curso y sílabos de los materiales didácticos (nivel

meso). Por su parte, el corpus de aprendientes se utilizará *a posteriori* para secuenciar los contenidos del currículo.

Este planteamiento resulta extraordinariamente novedoso y es una garantía de respuesta a las necesidades reales de los destinatarios del currículo: en primer lugar, porque el proceso de adaptación curricular llega hasta cada uno de los aprendientes; en segundo lugar, porque la selección de los textos que alimentan CORPEN se hace a partir de un análisis de necesidades; en tercer lugar, la secuenciación de los contenidos se realiza en virtud de los rasgos lingüísticos reales que marcan el paso de un nivel a otro.

REFERENCIAS BIBLIOGRÁFICAS

Alcaraz, E. (2000). *El inglés profesional y académico*. Madrid: Alianza Editorial.

Beacco, J.C., Byram, M., Cavalli, M., Coste, D., Egli Cuenat, M., Goullier, F. y Panthier, J. (2016). *Guide for the Development and Implementation of Curricula for Plurilingual and Intercultural Education*. Estrasburgo: Consejo de Europa. Recuperado de: https://rm.coe.int/CoERMPublicCommonSearchServices/DisplayDCTMContent?documentId=09000016806ae621.

Bhatia, V.K. (2002). Applied Genre Analysis: Analytical Advances and Pedagogical Procedures. En A. M. Johns (Ed.), *Genre in the Classroom: Multiple Perspectives* (pp. 279-283). Mahwah: Erlbaum.

Biber, D. (1993). Representativeness in Corpus Design. *Literary and Linguistic Computing*, 8(4), 243-257.

Biber, D., Connor, U. y Upton, T. (2007). *Discourse on the Move: Using Corpus Analysis to Describe Discourse Structure.* Ámsterdam: John Benjamins Publishing.

Cabré, M.ª T. (2007). Construir un corpus de textos de especialidad: condiciones y posibilidades. En M. Ballard y C. Pineira-Tresmontant (Eds.), *Les corpus en linguistique et en traductologie* (pp. 89-106). Arras: Artois Presses Université.

Ciapuscio, G. y López Ferrero, C. (2022). El estudio del discurso especializado basado en corpus: el caso de las cartas. En G. Parodi, P. Cantos-Gómez y C. Howe (Eds.), *The Routledge Handbook of Spanish Corpus Linguistics* (pp. 268-282). Londres: Routledge.

Consejo de Europa (2020). *Marco común europeo de referencia para las lenguas: aprendizaje, enseñanza y evaluación. Volumen complementario con nuevos descriptores*. Edición Madrid: Ministerio de Educación y Formación profesional/ Instituto Cervantes. Recuperado de: https://cvc.cervantes.es/ensenanza/biblioteca_ele/marco_complementario/mcer_volumen-complementario.pdf.

Hawkins, J.A. y Filipović, L. (2012). *Criterial Features in L2 English: Specifying the Reference Levels of the Common European Framework*. Cambridge: Cambridge University Press.

Hyland, K. (2005). *Metadiscourse*. Londres: Continuum.

Hyland, K. (2017). Metadiscourse: what is it and where is it going? *Journal of Pragmatics*, 113, 16-29.

Knoch, U. y Macqueen, S. (2020). *Assessing English for Professional Purposes*. Londres: Routledge.

Larsen-Freeman, D. y Long, M.H. (1994). *Introducción al estudio de la adquisición de segundas lenguas*. Madrid: Gredos.

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

Parodi G. (2005). Discurso Especializado y Lingüística de Corpus: Hacia el desarrollo de una competencia psicolingüística. *Boletín de Lingüística*, 17(23), 61-88. Recuperado de: https://www.redalyc.org/pdf/347/34702304.pdf.

Real Academia Española (2020). *Corpus del español del siglo XXI (CORPES). Descripción del sistema de codificación. Libros y prensa*. Madrid: Real Academia Española. Recuperado de: www.rae.es/sites/default/files/2020_DisYCod_Escritos.pdf.

Richards, J.C. (2013). Curriculum approaches in language teaching: forward, central, and backward design. *RELC Journal*, 44(1), 5–33. Recuperado de: http://www.professorjackrichards.com/wp-content/uploads/Curriculum-Approaches-in-Language-Teaching.pdf.

Rojo, G. (2021). *Introducción a la lingüística de corpus en español*. Londres: Routledge.

Rundell, M. y Stock, P. (1992). The corpus revolution. *English Today*, 8(3), 21-32

Sheghiri, M. (2011). Metodología protocolizada de compilación de un corpus de seguros de viajes: aspectos de diseño y representatividad. *Revista de Lingüística Teórica y Aplicada*, 49 (2), 13-20.

Sinclair, J. (1991). *Corpus, concordance, collocation*. Oxford: Oxford University Press.

Sinclair, J. (1996). *Preliminary Recommendations on Corpus Typology*. Recuperado de: http://www.ilc.cnr.it/EAGLES96/corpustyp/corpustyp.html.

Sinclair, J. (Ed.) (2004). *Trust the Text. Language, corpora and discourse*. Londres: Routledge.

Sinclair, J. (2005). Corpus and text – Basic principles. En M. Wynne (Ed.), *Developing linguistic corpora: A guide to good practice* (pp. 1–16). Oxford: Oxford Text Archive. Recuperado de: http://hdl.handle.net/20.500.12024/2951.

Swales, J. (1990). *Genre Analysis English in Academic and Research Settings*. Cambridge: Cambridge University Press.

Swales, J. (2004). *Research Genres: Explorations and Applications.* Cambridge: Cambridge University Press.

Timmis, I. (2015). *Corpus linguistics for ELT: research and practice*. Nueva York: Routledge.

Van Ek, J.A. *(*1976*). The threshold level for modern language learning in schools*. Estrasburgo: Consejo de Europa.

Vargas, C. (2006). Diseño de un corpus especializado con fines terminográficos: el corpus de la piedra natural. *Debate terminológico*, 2, 1-20. Recuperado de: http://rua.ua.es/dspace/bitstream/10045/9426/1/chelo_vargas_sierra.pdf.