Encuentro
Revista de investigación e innovación en la clase de idiomas

EL SOFTWARE DE ANÁLISIS DE CORPUS APLICADO A LA DOCENCIA DE LFE: PROPUESTA METODOLÓGICA

Leticia Moreno Pérez
Universidad de Valladolid - ACTRES

Resumen

Las necesidades del mundo actual han modificado la demanda de aprendizaje de lenguas extranjeras hacia la profesionalización, provocando un crecimiento exponencial de la enseñanza de lenguas para fines específicos (LFE). Esto ha supuesto un importante reto para los docentes que, en muchas ocasiones sin ser expertos, tienen que implementar metodologías efectivas que permitan formar a alumnos con perfiles muy diferentes en un periodo habitualmente limitado. El presente artículo expone una propuesta metodológica adaptativa basada en la utilización de software de análisis de corpus en la docencia de LFE. El empleo de corpus en la enseñanza de lenguas, investigado desde hace décadas por numerosos autores (Aston, 2000, 2001; Johns 1991, 1994, 1997, 2002; Santamaría-García, 1995, 1998, 2011; Sinclair, 1991), permite al alumno construir su conocimiento de forma autosuficiente, siguiendo la premisa «learn by doing» (Schank, 1995), y acercándose así al lenguaje real utilizado por los expertos. La metodología permite, a su vez, adoptar distintos enfoques en función del nivel del alumnado y solventar las carencias puntuales en el conocimiento experto del docente, el cual podrá nutrirse de los posibles conocimientos expertos del discente.

Palabras clave: Lenguas para fines específicos, análisis de corpus, lingüística del texto, didáctica de la lengua.

Abstract

The needs of our current globalized world have transformed the demand of foreign language learning towards professionalization, which results in an increasing demand of languages for specific purposes (LSP) training. Teachers, who many times are not LSP experts, are facing new challenges like finding effective methodologies to train students of very different profiles usually in limited periods of time. This paper presents an adaptive methodology proposal based on the use of corpus analysis software for LSP teaching. The use of text corpora in language teaching – researched by different scholars in the last decades (Aston, 2000, 2001; Johns, 1991, 1994, 1997, 2002; Santamaría-García, 1995, 1998, 2011; Sinclair, 1991) – allows students to build their own knowledge through «learn by doing» theory (Schank, 1995), while approaching to real language used by experts. In addition, different approaches can be adopted through this methodology according to the level of the students, and lack of expert knowledge of the teachers can also be resolved, as students will share their expert knowledge with them.

Keywords: Languages for specific purposes, corpus analysis, text linguistics, language didactics.

1. Introducción

En los últimos tiempos la enseñanza superior ha evolucionado hacia la especialización, tanto del alumnado como del profesor. En el caso de los idiomas, esto se ve reflejado en la demanda de formación en lenguas para fines específicos (LFE), que comienza tan pronto como termina la enseñanza obligatoria. Esta realidad ha dado origen a nuevos retos en la enseñanza de lenguas a los cuales los docentes han tenido que adaptarse para cubrir dicha demanda. Autores como Roberts (1991), Rodríguez Piñero y García Antuña (2009) y García Laborda y Litzler (2015), entre otros, han analizado los problemas a los que deben enfrentarse los profesores de LFE, de entre los que destacan especialmente tres. En primer lugar, la diversidad de perfiles de los discentes, que tienen niveles de partida de lengua extranjera dispares,

conviviendo en el aula alumnos con una base previa afianzada de la lengua extranjera con compañeros que tienen un nivel relativamente bajo. Esto también sucede, según nuestra experiencia, en el nivel de conocimientos de la especialidad en la lengua materna, pues algunos alumnos ya son expertos, mientras que otros no. En segundo lugar señalan que, en la mayoría de los casos, los docentes de LFE no son expertos en el dominio de especialidad que enseñan, sino profesores de lengua extranjera general que deben reorientar sus conocimientos hacia una especialización por materias dentro de dicha lengua. Si unimos este condicionante a las nuevas modalidades de aprendizaje (la presencialidad ha dejado de ser imprescindible y la docencia se basa, cada vez más, en las nuevas tecnologías como apoyo, o en sustitución, de la clase magistral), observamos como el docente de LFE se ha convertido en un profesional «todoterreno» que debe formarse en múltiples aspectos para poder enfrentarse a los nuevos retos de la enseñanza. En tercer lugar cabe destacar las restricciones temporales que suelen caracterizar los cursos de LFE frente al volumen de contenidos exigido, lo cual añade otro condicionante que complica aún más el panorama del docente. Todos estos escollos tienen como consecuencia la necesidad vital de encontrar metodologías efectivas que permitan al docente ser altamente productivo y adaptativo para que el alumno alcance el mayor rendimiento equilibrando las posibles discordancias y carencias de alumnado y profesorado en el periodo limitado de que se dispone.

2. Las LFE y su aprendizaje

Las LFE se contraponen a la lengua general por su carácter especializado, pues se trata de lenguajes con «[…] una temática específica, utilizada en situaciones pragmáticas determinadas» (Aguado, 2002: 18). Esa temática específica corresponde al saber de los distintos dominios de especialidad, por ejemplo la medicina o el derecho. Las situaciones pragmáticas, por su parte, son aquellas que se producen cuando los expertos de dichos dominios, por ejemplo médicos o juristas, comunican esa temática específica. Esta especialización temática y situacional tiene como consecuencia la existencia de características que hacen que las LFE difieran de la lengua general. De entre estas características podemos destacar las siguientes:

- Terminología: se trata del léxico empleado dentro de un dominio de especialidad. Está compuesta por los denominados términos, «[…] una unidad descrita por un conjunto de características lingüísticas sistemáticas y dotada de la propiedad de referirse a un elemento de la realidad […] utilizada en un dominio de especialidad» (Cabré, 2005: 25). Surgen de la necesidad de sus usuarios, los expertos, de denominar un concepto para poder referirse a él dentro de su disciplina profesional (íbid), y son los elementos más característicos de las LFE (Cabré y Gómez de Enterría, 2006).
- Elementos sintáctico-gramaticales: existen ciertos patrones y estructuras distintivas de las LFE, bien sean patrones considerados generales a todas ellas, como la tendencia a la nominalización (Eckkrammer, 2009), o patrones propios de una LFE concreta, como la sobreutilización de la estructura pasiva en el discurso médico o la mayor longitud oracional propia del discurso jurídico.
- Géneros: tanto terminología como sintaxis y gramática se ven reflejados en la comunicación especializada, que se organiza en distintos géneros textuales dependiendo de la finalidad de dicha comunicación. Swales (1990: 58) define el concepto de género como «…a class of communicative events, the members of which share some set of communicative purposes. These purposes are recognized by the expert members of the parent discourse community, and thereby constitute the rationale for the genre. This rationale shapes the schematic structure of the discourse and influences and constrains choice of content and style». La estructura, contenidos y estilo característicos que menciona

este autor en su definición implican que estos textos especializados tengan reglas específicas de elaboración que es necesario aprender, en contraposición a la espontaneidad del aprendizaje de la lengua general (Cabré y Gómez de Enterría, 2006).

Dado que estas características son las que distinguen a las LFE, si nos trasladamos al ámbito de la enseñanza serán los pilares en los que su docencia deberá basarse para permitir cubrir las necesidades de aprendizaje esperadas. La terminología es sin duda el escollo más inmediato con el que se encontrará cualquier aprendiz de una LFE, pues los términos delimitan conceptos muy específicos que generalmente no pueden ser parafraseados o explicados de otro modo, como puede hacerse con los conceptos de la lengua general. Esto hace que la necesidad más inmediata de los aprendices sea conocer dicha terminología en la lengua extranjera. Por tanto, sea experto o no en su lengua materna, su introducción a la LFE deberá asegurar que adquiere la terminología específica del dominio de especialidad, ligada a los conceptos que denomina. En cuanto a los elementos sintáctico-gramaticales, el aprendizaje pasará por reconocer las estructuras y patrones característicos para, por un lado, ser capaz de comprender el discurso (ciertas estructuras fuera de lo habitual hacen que el lector no familiarizado con ellas tenga problemas a la hora de entender el sentido de una oración) y, por otro, ser capaz de generarlo de acuerdo a las expectativas del dominio de especialidad en cuestión. Por último, el conocimiento de los géneros textuales se presenta como elemento indispensable para poder dominar un campo especialidad, por lo que para dominar dicha disciplina se hace necesario el aprendizaje de sus normas de elaboración.

Todo lo anterior debe sumarse a los elementos propios del aprendizaje de lenguas en general (desarrollo de competencia oral, escrita, auditiva y lectora), lo cual demuestra la complejidad de la enseñanza de LFE. Esta complejidad deberá reflejarse, necesariamente, en las metodologías empleadas para su impartición. Si bien es cierto que autores como Hutchinson y Waters (1987) postulaban en los albores de la época moderna de esta disciplina que no existía como tal una metodología de LFE, sino que se trataba de metodologías utilizadas para su enseñanza pero que se podían utilizar para la docencia de la lengua general, estudios más recientes (Robinson, 1991; Cabré y Gómez de Enterría, 2006) abogan por una especialización en las metodologías con el objetivo de satisfacer las necesidades específicas de esta disciplina.

En este sentido, Rodríguez Piñero y García Antuña (2009: 924) exponen que la didáctica de las LFE «se fundamenta en el análisis de necesidades concretas y reales que demanda el grupo, aspecto que junto con el conocimiento de la competencia comunicativa específica y de la experiencia previa en dicho campo específico conforman la base del proceso de enseñanza-aprendizaje de LFE». Bajo esta premisa recopilan las técnicas de enseñanza propias de esta disciplina (basándose en Cabré y Gómez de Enterría, 2006 o Aguirre Beltrán, 2004), que vienen determinadas por el objetivo fundamental de reflejar en la clase aquello que ocurre en situaciones reales de uso de la LFE que se esté trabajando (íbid: 924-926):

- Simulación global: reconstrucción en el aula de los elementos que configuran una situación comunicativa en un ámbito específico.
- Proyectos: tarea final amplia en la que todas las actividades están orientadas a la preparación, desarrollo y elaboración de un proyecto relacionado con la actividad profesional.
- Tareas: implica a los aprendices en la comprensión, manipulación, producción e interacción en la LE.
- Presentaciones orales: permite ejercitar la comunicación no verbal de forma natural. Puede utilizarse por sí sola o puede formar parte de uno de los enfoques anteriores.
- Estudios de casos: analizar problemas reales en un contexto profesional.

Observamos en esta recopilación una serie de actividades orientadas, efectivamente, a la práctica profesional real del alumno. En todas ellas existe un proceso de aprendizaje basado en el «learn by doing» (Schank, 1995), y en la mayoría de estas técnicas se trabajan las características distintivas de las LFE (terminología, sintaxis/gramática y géneros) a las que hemos hecho referencia. Pero abordar el estudio de todos estos elementos de forma unificada sin hacerlo de forma artificial puede suponer un problema en el aula de LFE. Por ello, en la presente propuesta abogaremos por el uso de una metodología que permite el estudio de todas estas características en contexto e investigada por numerosos expertos (cabe destacar Aston, 2000, 2001; Johns 1991, 1994, 1997, 2002; Santamaría-García, 1995, 1998, 2011; Sinclair, 1991): el aprendizaje de lenguas mediante análisis de corpus.

3. El uso de corpus en la enseñanza de lenguas

Para entender esta metodología es esencial comenzar analizando el concepto de corpus. Podemos definirlo como «a collection of pieces of language that are selected and ordered according to explicit linguistic criteria in order to be used as a sample of the language» (EAGLES, 1996: 4). El objetivo principal del análisis basado en corpus es documentar e interpretar patrones de uso de la lengua mediante la observación de la comunicación real entre los hablantes (Gozdz-Roszkowski, 2011). Al tratarse de un análisis de textos reales es considerado un método con un amplio índice de fiabilidad (Biber, Conrad y Reppen, 1998), y de ahí su utilidad en la enseñanza y aprendizaje de lenguas (Santamaría-García, 1995), más concretamente de LFE (Tolchinsky, 2014).

Muchas son las perspectivas aportadas por diversos autores a lo largo de las más de tres décadas en que la unión corpus y docencia se ha venido desarrollando, cada vez con más peso. Parece existir acuerdo en que Johns es el precursor de la exposición directa del alumno y el profesor al corpus, en lo que denominó «data-driven learning» (1991, 1994, 1997, 2002, entre otras). Desde ese momento otros autores han seguido contribuyendo en el desarrollo de esta metodología con distintos enfoques, entre los que cabe destacar el trabajo mediante corpus de muestras producidas por aprendices de una lengua extranjera (Aston, 2000; Flowerdew, 2000), el uso de las concordancias en la docencia (Cobb, 1997; Santamaría-García, 1998; Aston, 2001; Flowerdew, 2001), los estudios sobre colocaciones (Sinclair, 1991; Stubbs, 1995), la obtención de datos sobre géneros textuales a través de corpus (Henry y Roseberry, 2001) o el análisis del discurso mediante el uso de corpus oral (Santamaría-García, 2011), entre otras.

Todos estos enfoques de análisis son solo una muestra resumida de las múltiples aportaciones con que esta metodología puede contribuir a la enseñanza de lenguas. La contribución más inmediata es la posibilidad de estudiar a través del corpus los elementos básicos de la lengua, como son la gramática, el vocabulario y el discurso (Santamaría-García, 1995). Pero la investigación de estos elementos no se detiene ahí, pues puede derivar en la mejora de manuales pedagógicos (Nesselhauf, 2004; McCarten, 2010), el diseño y evaluación de planes de estudio o la creación de diccionarios y gramáticas (Aston y Burnard, 1998; Walter, 2010). Por último, no debemos olvidar los aspectos positivos que supone para el aprendizaje del alumnado la implementación del uso de corpus o material basado en él. Por un lado, su aplicación implica que el estudiante se aleje de modelos lingüísticos idealizados y artificiales, habituales en los libros de texto, para exponerse a muestras naturales del lenguaje, lo cual le proporcionará una formación más cercana a sus necesidades reales. Por otro, fomenta el autoaprendizaje más allá del aula una vez los

alumnos ya se han formado en el manejo de corpus, pues como afirman Aston y Burnard (1998: 19-20) «with appropriate training, advanced learners can use large corpora as reference tools which overcome many of the limitations of existing dictionaries and grammars by providing a much larger number of more contextualized examples». Finalmente, aunando los beneficios para enseñanza y aprendizaje, como valor añadido a su utilización se produce una simbiosis entre profesor y alumno en la cual ambos aprenden mediante la investigación del lenguaje en contexto, lo cual hace de ella una metodología tanto útil como atractiva, pues, como apunta Santamaría-García (1995: 195) es percibida por los usuarios como «[…] easy, quick and motivating».

3.1 El análisis de corpus en LFE: compilación y herramientas

Para que el aprendizaje logre cubrir las necesidades reales anteriormente mencionadas es esencial adaptar el corpus al estudiante y sus objetivos de aprendizaje (Johns, 2002). Como establecimos anteriormente en su definición, el concepto de corpus gira en torno a dos ejes: las muestras del lenguaje, es decir, una delimitación de textos concretos que se incluirán en el corpus en función de su propósito, y los criterios lingüísticos, es decir, la definición de cómo se compondrá ese corpus.

Para determinar las muestras del lenguaje que formarán parte de la compilación nos valdremos de la clasificación por géneros textuales, dado que lo que vamos a compilar para nuestros propósitos es un «topic-specific corpora» (Aston y Burnard, 1998: 12), es decir, un corpus centrado en un dominio de especialidad concreto. Swales (1990) introduce ligado a su definición de género el concepto de prototipos. Se trata de ejemplares de un género que muestran patrones estructurales, estilísticos, de contenido y receptor meta comunes. Estos prototipos suelen surgir asociados a comunidades discursivas específicas, por lo que, en el caso de la aplicación del análisis de corpus a las LFE, los géneros textuales prototípicos seleccionados vendrán determinados por la comunidad discursiva a la que se adscriba dicha LFE: juristas, médicos, etc. Una vez definida la comunidad discursiva y sus géneros prototípicos podremos seleccionar aquellos que más puedan interesar para nuestro propósito, y serán principalmente aquellos más representativos de cada dominio de especialidad. Por un lado, esto permitirá que los alumnos se familiaricen con los textos a los que más posibilidades tengan de enfrentarse en el uso profesional de la LFE y, por otro, al tratarse de los más representativos contendrán los principales elementos terminológicos y sintáctico-gramaticales del dominio de especialidad.

Una vez determinados qué géneros incluir en el corpus es momento de realizar su diseño en base a los criterios lingüísticos. Para ello, nos valdremos de la lingüística de corpus, metodología que se encarga de definir los parámetros principales para compilar un corpus. Los parámetros de mayor relevancia para nuestra propuesta serán los siguientes:

- Tipología: determinada por la selección de distintos elementos, como medio oral o escrito, temporalidad diacrónica o sincrónica, grado de especialización, extensión completa o parcial de las muestras incluidas, lengua(s) de dichas muestras, originales o traducciones, etc. (Sinclair, 1991; Kennedy, 1998; Leech, 2007; McEnery y Hardie, 2012).
- Tamaño: si bien se presupone que un corpus debe tener el mayor número posible de elementos (EAGLES, 1996), el número específico dependerá de cada estudio en particular, dados los múltiples factores que pueden influir según el caso: el propósito del corpus (Kennedy, 1998), la dificultad para encontrar los materiales incluidos en el corpus (EAGLES, 1996), que no requieran el uso de equipos costosos y que sean manejables sin que el usuario posea habilidades técnicas muy específicas (Sinclair, 2001), etc. En este caso, el tamaño se verá afectado por factores como la duración del curso, que

determinará cuántos géneros dará tiempo a analizar, el nivel del curso, que determinará qué tipo de géneros se analizarán, o la facilidad para acceder a los documentos por parte del profesor, directamente relacionada con su especialización en la materia. En cualquier caso, a modo orientativo, trabajos previos han demostrado que, en el caso del estudio de LFE, 10.000 palabras son un buen punto de partida (Bowker y Pearson, 2002), número que parece accesible dadas las restricciones expuestas anteriormente.

- Procedencia: es necesario confirmar que se trata de muestras de calidad (EAGLES, 1996) elaboradas por usuarios reales y cualificados.
- Representatividad: es importante que las fuentes del corpus sean lo suficientemente diversas para que sea una muestra real del género seleccionado. Si todos los textos provinieran del mismo autor podríamos encontrarnos con preferencias en el uso de expresiones o términos que darían como resultado unos datos sesgados.

Para aplicar esta propuesta metodológica a la enseñanza de LFE podemos realizar distintos diseños en función de las necesidades de los alumnos, su nivel de especialización o las restricciones temporales. A continuación vamos a recopilar una visión general de las distintas posibilidades que podrían ser más beneficiosas en este contexto concreto.

La compilación de un corpus oral podría ser interesante para cuestiones como comprensión y pronunciación, pero se debe tener en cuenta que, habitualmente, es complicado obtener muestras de audio originales y reales que no hayan sido elaboradas artificialmente por especialistas en la enseñanza e interpretadas por actores. Por esta razón vamos a centrarnos en la compilación de corpus escritos.

Lo ideal sería que fueran géneros prototípicos completos de entre los más representativos del dominio de especialidad, de carácter sincrónico (en uso en la actualidad) y redactados por especialistas. Con respecto a la lengua del corpus, existen dos opciones: compilar un corpus monolingüe en la lengua extranjera que se esté estudiando y utilizarlo solo para consultas dentro de una misma lengua o compilar un corpus comparable bilingüe en la lengua materna y en la extranjera para realizar un análisis contrastivo entre ambas lenguas. Esto dependerá del tiempo con que se cuente y el nivel de partida de los alumnos.

En cuanto al grado de especialización, lo ideal para un aprendizaje completo sería compilar subcorpus de distintos géneros. Así se podrá adquirir una panorámica de cómo funciona la LFE desde los términos y estructuras típicos de cada género.

Con respecto al tamaño del corpus de partida, dependerá enteramente del docente, sobre todo del tiempo que pueda dedicar a la compilación y de sus medios para encontrar el material adecuado al diseño. No consideramos que el tamaño sea de vital importancia para nuestra propuesta, dado que uno de los objetivos del empleo en el aula de esta metodología es que los propios alumnos amplíen el corpus poco a poco con sus aportaciones, incluso creando otros subcorpus de otros géneros que no aparezcan inicialmente pero sean importantes en el campo, lo cual contribuirá a autoformación del alumno a la vez que a la del profesor. Sin embargo, para la fase de análisis debería contar en inicio con un número suficiente de muestras de cada género que permita tener una visión general de su terminología y elementos sintáctico-gramaticales.

Una vez esté compilado el corpus será el momento de implementar su uso en el aula. Para ello será necesario que los alumnos aprendan a manejar la herramienta informática que les permitirá analizarlo. El software de análisis y gestión de corpus permite que su análisis sea más ágil, y además permite almacenar el conocimiento y seguir ampliándolo de forma más cómoda y rápida.

Este software se ha venido empleando desde hace años en la investigación en ámbitos lingüísticos, aplicados o no a la enseñanza. Por ello, existen numerosos programas y aplicaciones a disposición de los usuarios con distintas características y opciones de uso. Para nuestra propuesta hemos elegido AntConc, un software gratuito que permite trabajar en los principales sistemas operativos y que, además, fue creado para su uso en el aula (Anthony, 2005). Se trata de un programa sencillo pero no por ello menos útil, pues contiene herramientas suficientes para un análisis en profundidad y a distintos niveles.

Este mismo autor (ibíd.) describe las principales herramientas de AntConc, de entre las cuales consideramos que podrán sernos de especial utilidad las siguientes:

- «Concordance» (concordancias): permite realizar búsquedas de términos y observarlos en contexto (KWIC) para analizar cómo se utilizan en el corpus.
- «Concordance Plot» (gráfico de concordancias): permite observar de forma gráfica con qué frecuencia y distribución aparece un término en los distintos textos del corpus.
- «Clusters» (grupos de palabras): permite buscar unidades terminológicas polilexemáticas (formadas por más de un lexema) mostrando las palabras que acompañan al término buscado, pudiendo delimitar el número de palabras previas y posteriores y la frecuencia mínima de aparición.
- «Wordlist» (listado de palabras): permite realizar un listado alfabético o por frecuencia de las palabras de un corpus con la posibilidad de delimitar la búsqueda mediante «stoplists» en las que el usuario define qué palabras quiere o no quiere que aparezcan en el listado.

El manejo de este software puede resultar complicado dadas las innumerables variables de búsqueda y ordenación permitidas, pero debe recordarse que se va a utilizar como una herramienta de apoyo al aprendizaje de LFE. Por tanto, las actividades que se programen con su uso deberán ser sencillas y efectivas, de modo que, tras una primera toma de contacto en la que el alumno comprenda qué se pretende hacer con el programa y cómo tiene que hacerlo, pueda trabajar de forma autónoma sin demasiados problemas después de una pequeña práctica.

Las posibilidades que ofrece el uso de esta herramienta son muy variadas y, al igual que la compilación de corpus, su explotación depende principalmente de cuestiones de tiempo y nivel del alumnado y del docente. A continuación presentamos una propuesta de aplicación de esta metodología en el aula para mostrar cómo puede ayudar en la enseñanza-aprendizaje de LFE intentando tener en cuenta distintas variables. No se trata ni mucho menos de un modelo único, sino que se pretende ser una muestra para que cada profesor pueda adaptarlo a sus circunstancias.

4. Aplicación práctica de la metodología en el aula

Esta propuesta metodológica no es excluyente, es decir, su uso no debe suponer el destierro de otras metodologías ya empleadas en la actualidad, sino que el objetivo es combinarla con éstas como hilo conductor de todas ellas. Se incorporaría en una fase más temprana del conocimiento (previa a la implementación de las actividades enumeradas en el apartado 2), serviría al alumno como material de referencia a lo largo de todo el proceso de aprendizaje (como apoyo a la realización de todas estas actividades propuestas) y, posteriormente, continuaría como elemento de apoyo y extensión mediante un proceso de actualización continua.

Para esta puesta en práctica hemos decidido no limitarnos a ningún par de lenguas, dado que podría utilizarse cualquiera, sino que haremos referencia a la lengua materna (LM) y la lengua extranjera (LE) de los alumnos. Hemos dividido la propuesta en 4 fases:

4.1. Fase previa

Esta fase es aquella en la que se sientan las bases para el empleo de la metodología en el aula. Se divide, a su vez, en dos partes:

4.1.1. Compilación del corpus

Esta tarea corresponderá al docente. En función del diseño de su asignatura, el tiempo y el nivel de sus alumnos, será el momento de determinar si se compila exclusivamente un corpus comparable en la LE o un corpus dividido en dos subcorpus, uno de textos en LE y otro en LM. Siempre es recomendable compilar ambos, no sólo por si finalmente se dispone de tiempo en clase, sino también para que el docente vaya construyendo su propio conocimiento experto al trabajar sobre ellos. Independientemente de la decisión que se tome, será necesario establecer los géneros que se van a incluir en los subcorpus. Nosotros apostaremos por los más representativos, que permitan obtener una visión global del dominio de especialidad y que pertenezcan a las distintas áreas en que el dominio de especialidad se divide. Así pues, si por ejemplo estuviéramos aplicando esta metodología a un curso de lengua para los negocios, deberíamos buscar esa heterogeneidad incluyendo textos pertenecientes a las distintas áreas de ese dominio de especialidad: derecho (contrato de distribución), economía (informe macroeconómico), finanzas (letra de cambio) y comercio (correspondencia comercial) (Román, 2012). Si, por ejemplo, se tratara de un curso de inglés para medicina, deberíamos incluir textos de comunicación entre los profesionales (notas clínicas), de divulgación (folletos de salud pública) y comunicaciones «familiares» (diario de laboratorio) (López, 2000).

Si optamos por el comparable dividido en dos subcorpus sería necesario buscar los textos correspondientes en la LM, que suelen existir en su mayoría dada la actual globalización del mundo profesional. Además, gracias a la proliferación de estudios sobre los géneros textuales especializados muchos textos se pueden encontrar sin excesiva dificultad. Si el docente tuviera dificultades para encontrar textos, el corpus podría limitarse a unos pocos géneros que se fueran ampliando con ayuda de otros aportados por los estudiantes (si, por ejemplo, trabajan en ese dominio de especialidad).

4.1.2. Introducción a los alumnos

En esta fase, ya dentro del curso en sí, la labor del docente será presentar a los alumnos el corpus y explicarles la metodología que se va a seguir con su análisis. Esta introducción irá inexorablemente acompañada de una explicación paso a paso del uso del programa informático. Para que puedan familiarizarse con ella, en primer lugar se les enseñará a cargar el corpus en el programa para, sobre él, explicarles el funcionamiento de las distintas herramientas seleccionadas como primera aproximación, animándoles a realizar una búsqueda de prueba para cada una de ellas ya sobre el corpus. Sería conveniente entregar a los alumnos un manual paso a paso del uso de cada herramienta tras la demostración y, si se dispone de tiempo, dejar a los alumnos practicar el resto de la sesión.

4.2. Análisis del corpus en LE

En esta parte se comenzaría el trabajo con el corpus en LE. Lo ideal a partir de esta fase sería dedicar una parte de la clase al análisis de corpus y, otra, a las actividades complementarias, incluso alternar una

sesión completa de análisis de corpus con una siguiente sesión de actividades relacionadas con la temática trabajada. Los alumnos podrán en esta fase comenzar a detectar y aprender los términos más utilizados en el dominio de especialidad para construir su «glosario mental», forma en que hemos denominado a la relación interiorizada de los elementos léxicos más significativos del campo en cuestión, junto con los conceptos que determinan y sus patrones de uso. Los términos que componen ese glosario los podrán y necesitarán implementar en el resto de actividades simultáneas fuera del trabajo con corpus. Dichas actividades irán, preferentemente, ligadas a los textos que se estén trabajando en cada momento.

4.2.1. Extracción de términos clave

Para la construcción del mencionado «glosario mental» es de vital importancia determinar cuáles son las palabras clave del dominio de especialidad, tarea que puede resultar prácticamente imposible para un aprendiz inexperto por la falta de contexto y la amplitud léxica y conceptual que puede llegar a tener un dominio de especialidad, además de las restricciones temporales que marquen la adquisición de su aprendizaje. Ante esta tesitura, podemos encontrar en el análisis a través de corpus un gran aliado en esta tarea, pues, como apunta Flowerdew (2001: 75), «the great power of the corpus-based word list is in that the course designer can be sure that the words selected are the most useful». No obstante, dado que nuestro interés es explorar la terminología propia de los géneros especializados, antes de comenzar la extracción de esta lista sería recomendable introducir una «stoplist» genérica para cribar resultados más rápidamente. Para elaborarla se realizaría género a género una «Wordlist» ordenada por frecuencia de aparición para detectar las denominadas «high frequency words» (Kennedy, 1998: 67), palabras generales como preposiciones o artículos que no son representativas del dominio de especialidad. Estas serán la base de nuestra «stoplist», que podremos ir modificando para afinar más la búsqueda según analicemos cada género en particular. El docente puede entregársela hecha a los alumnos, pero sería muy interesante que pudieran elaborarla ellos mismos, pues así podrían implementar ese mismo proceso cuando manejen el corpus de forma autónoma en el futuro. Una vez la tengamos podríamos proceder al análisis de cada género por separado. En un primer acercamiento los alumnos elaborarían un listado de palabras de dicho género tras haber cargado la «stoplist» y así obtendrían, por frecuencia, los términos más representativos de cada uno. No obstante, sería recomendable establecer un mínimo de apariciones de un término y en un número mínimo de textos para asegurarse de que es realmente representativo del género en cuestión, criterio que dependerá del volumen del corpus con que trabajemos.

Todo este proceso de extracción conviene realizarlo en dos fases: en una primera se buscarían unidades terminológicas monolexemáticas. Una vez obtenido el listado definitivo, se pasaría a detectar mediante la herramienta «Clusters» las asociaciones frecuentes de esos términos con otras unidades, con lo que obtendríamos las unidades terminológicas polilexemáticas.

4.2.2. Elaboración del glosario

La extracción de los términos representativos en la LE no es útil por sí sola, pues los alumnos necesitan conocer también su traducción a su LM y, por tanto, el concepto subyacente, lo que les permitirá elaborar un glosario. Una vez obtenido el término en LE los alumnos deben buscar en un diccionario especializado monolingüe en la LE el concepto, es decir, la definición de dicho término. Aquellos con un mayor nivel de conocimiento experto en el dominio de especialidad podrán deducir cuál es el término equivalente en la LM, mientras que los que no puedan buscarán directamente el equivalente en español en un recurso

bilingüe y comprobarán en diccionarios monolingües especializados en LE y en LM que se trata del mismo concepto. Estas búsquedas permitirán aprender durante el proceso y, a la vez, asegurarnos de que se trata de términos propios del dominio de especialidad en cuestión. Para que el glosario sea lo más eficaz posible se puede dividir en apartados por género textual o realizar un solo apartado por orden alfabético añadiendo etiquetas que establezcan en qué género se utiliza cada término.

Si consideramos la convivencia de esta metodología con las otras actividades empleadas en la enseñanza de LFE, la primera parte de esta fase podría servir de apoyo, por ejemplo, a la realización de estudios de caso sencillos que tengan relación con la temática del género (por ejemplo sobre la historia clínica de un paciente o los estados financieros de una empresa) o combinarse con la realización de tareas en las que deban utilizar los términos clave de cada género que hayan encontrado. Posteriormente, tras la segunda parte de esta fase, los alumnos podrán enfrentarse a actividades más complejas, pues ya disponen de un mapa mental de la terminología de cada género. Podría, por ejemplo, comenzarse a realizar actividades de presentación oral o de simulación global (por ejemplo en la Junta General de una empresa o la revelación del diagnóstico a un paciente).

4.3. Análisis contrastivo

Para profundizar en el conocimiento de la LFE estudiada, en esta fase se procedería al análisis contrastivo de los subcorpus en LE y en LM.

4.3.1. Comparación de terminología

En un primer ejercicio podría repetirse el proceso de obtención de la «Wordlist» de términos clave para observar si coinciden con los conceptos y equivalentes que obtuvimos en el análisis del corpus en LE. Así determinaríamos si un término o expresión se utiliza con la misma frecuencia en ambas lenguas y, si no, podríamos intentar determinar cuál se utiliza en su lugar si obtuvimos alguno con un significado similar. La herramienta «Concordance Plot» nos permitiría en este caso observar frecuencia y distribución de los términos por texto.

4.3.2. Comparación de macroestructura

Una de las utilidades de poder contrastar géneros textuales en distintas lenguas es que se puede comparar si la macroestructura es equivalente en ambas o si, por el contrario, existen variaciones entre lenguas o no coincide, tanto en contenidos como en forma. Por ejemplo, ¿encontramos las mismas cláusulas en un contrato de distribución en LE y LM? ¿Las partes de dicho contrato se distribuyen de la misma manera? (encabezado, mayúsculas, títulos, etc.). Esto puede resultar extremadamente útil a la hora de comprender, redactar o traducir dichos documentos entre las dos lenguas de trabajo. Mediante la herramienta «Concordance» podemos observar en contexto un término o frase concretos (por ejemplo, el título de una cláusula en un contrato o de un apartado del informe anual) en todos los textos en que aparezca y seleccionar el que nos interese para que se nos muestre el texto en que aparece en su totalidad y así comparar si siempre encontramos los mismos apartados en los mismos lugares.

4.3.3. Comparación sintáctico-gramatical

La más compleja de las posibilidades que ofrece el análisis de corpus es la comparación de elementos sintáctico-gramaticales. Se requiere un nivel elevado de la LE y su análisis está orientado fundamentalmente a la comprensión y redacción textual, por lo que se pospone a la fase más avanzada del aprendizaje. Se podría proceder al estudio de los textos completos de un género concreto en el propio

programa para extraer características sintáctico-gramaticales propias de ese género (oraciones pasivas, uso de gerundios, etc.) o emplear el corpus para buscar directamente los elementos deseados y estudiar entonces su funcionamiento en exclusiva, género a género o en conjunto (introduciendo en la herramienta «Concordance», por ejemplo, las marcas propias de pasiva o gerundio de las lenguas de trabajo).

Dado el nivel de profundidad al que se trabaja en esta fase, el docente podría combinarla con la realización de cualquiera de las actividades distintas a nuestra metodología descritas en la fase anterior, pudiendo aumentar el nivel de dificultad. El alumno ya estaría preparado, por ejemplo, para tareas complejas como la redacción de textos del ámbito profesional en la LE (informe de auditoría o folleto sanitario).

4.4. Fase de ampliación

Esta última fase corresponde al alumno. En ella podrá ampliar el corpus con más textos de los géneros que lo componen o bien incluir subcorpus con géneros nuevos para poder analizarlos de igual manera. Puede adelantarse a las fases anteriores si el docente necesitara ayuda para completar el corpus inicial, siendo en este caso el alumno partícipe de la fase previa al empleo de la metodología. Puede realizarse a lo largo del curso o una vez terminado.

5. Resultados esperados y conclusiones

La metodología basada en el uso de corpus aquí propuesta pretende servir como apoyo al aprendizaje de LFE a lo largo de todas sus fases y niveles como complemento a las actividades y metodologías empleadas más habitualmente. Sumar esta metodología, avalada por un amplio número de investigadores, a las ya implementadas permitirá que el alumno no sólo adquiera y practique las habilidades propias de la enseñanza de lenguas (oral, escrita, auditiva y lectora), sino que además podrá adquirir los siguientes conocimientos:

- Conocimiento de los términos clave del dominio de especialidad utilizados en contexto. Esto les podrá ser de utilidad a la hora de desenvolverse en la LFE estudiada, no sólo en el entorno profesional sino también en clase con el resto de actividades complementarias que se realicen, como hemos ido sugiriendo.
- Metodología de elaboración de un glosario, que podrán ampliar y utilizar para futuras consultas.
- Familiarización con la macroestructura de los géneros más habituales en el dominio de especialidad, siendo capaces de comprenderlos, traducirlos e incluso redactarlos, dependiendo del nivel del alumno y del grado de profundización del curso.
- Uso de una herramienta informática sencilla que les pueda ser de utilidad en su futuro profesional, pudiendo ampliar el corpus con otros géneros que necesiten en su puesto de trabajo y así ampliar conocimientos, ampliar el glosario o elaborar otros nuevos, etc.

Como todo método docente, por supuesto, esta propuesta puede tener pros y contras, que dependen fundamentalmente de la idiosincrasia de cada curso, docente o grupo de alumnos. Entre los posibles contras que pueden surgir podemos señalar el hecho de que el uso de las TIC puede ralentizar el proceso de enseñanza-aprendizaje: no todos los alumnos tienen la misma destreza en el uso de herramientas informáticas, si bien es cierto que la elegida para esta propuesta, así como el conjunto de actividades planteadas, son bastante sencillas. Además, la utilidad de esta metodología es mayor en cursos de extensión media o larga, o en programas con cursos sucesivos, puesto que si el número de horas de que se dispone es muy limitado no daría tiempo a que el alumno indague de forma provechosa en el corpus y su

elaboración. Por último, somos conscientes de que las posibles limitaciones de acceso a los textos, especialmente del docente, pueden restringir en gran medida su utilización.

Sin embargo, la aplicación de esta metodología también cuenta, en nuestra opinión, con una serie de aspectos positivos más numerosos que los negativos, de los que querríamos destacar los siguientes: el uso de herramientas TIC en el aprendizaje añade atractivo para el alumno. Esta herramienta, además, puede almacenarse para su uso continuo y posterior al curso, fomentando así el aprendizaje permanente y el autoaprendizaje, pues a lo largo de toda su implementación el alumno toma parte en su propia formación, lo cual es un aliciente más. Esta autoformación, que además también lo es para el profesor, aparece combinada con el trasvase de conocimientos entre alumno y profesor que, en lugar de realizarse de la forma convencional del segundo al primero, en este caso se produce en ambas direcciones, suponiendo una ventaja para ambos. Por último, pero no menos importante, debemos resaltar que con esta metodología el alumno aprende empleando material de uso en la vida profesional real y no adaptado o creado para el entorno docente, lo que le hará estar más preparado en un futuro.

Por todas estas razones consideramos que la inclusión del análisis de corpus mediante herramientas informáticas en la enseñanza de LFE como metodología complementaria a las ya existentes puede ser altamente beneficiosa a lo largo de todo el proceso de enseñanza-aprendizaje, así como para todos sus actores.

References

Aguado, G. (2002). "Lenguas para fines específicos y terminología: algunos aspectos teóricos y prácticos". En Bravo, S. (Ed.), *Actas del I Encuentro de Estudios de Filología Moderna y Traducción, Departamento de Filología Moderna de la ULPGC, Las Palmas, 29 y 30 de noviembre y de diciembre de 2001)*. Las Palmas de Gran Canaria: Universidad de Las Palmas de Gran Canaria (pp. 15-27).

Aguirre Beltrán, B. (2004). "La enseñanza del español con fines profesionales". En Sánchez Lobato, J. e I. Santos Gargallo (dirs.), *Vademécum para la formación de profesores*, Madrid: SGEL, (pp. 1107-1302).

Anthony, L. (2005). "AntConc: A Learner and Classroom Friendly, Multi-Platform Corpus Analysis Toolkit". En *Proceedings of IWLeL 2004: An Interactive Workshop on Language e-Learning*. Tokyo: Waseda University, 7-13.

Aston, G. (2000). "Corpora and language teaching". En Burnard, L. y McEnery, T. (Eds.), *Rethinking language pedagogy from a corpus perspective* (pp. 7-17). Frankfurt: Peter Lang.

Aston, G. (2001). *Learning with corpora*. Houston: Athelstan.

Aston, G. y Burnard, L. (1998). *The BNC Handbook. Exploring the British National Corpus with SARA*. Edinburgh: Edinburgh University Press.

Biber, D., Conrad, S. y Reppen, R. (1998). *Corpus Linguistics: Investigating Language Structure and Use.* Cambridge: Cambridge University Press.

Bowker, L. y Pearson, J. (2002). *Working with Specialized Language: a practical guide to using corpora*. London: Routledge.

Cabré, M. T. (2005). *La terminología: representación y comunicación. Elementos para una teoría de base comunicativa y otros artículos*. Barcelona: Institut Universitari de Lingüística Aplicada.

Cabré, M. T. y Gómez de Enterría, J. (2006). *La enseñanza de los lenguajes de especialidad. La simulación global*. Madrid: Gredos.

Cobb, T. (1997). "Is there any measurable learning from hands-on concordancing?". *System*, 25, pp. 301-315.

EAGLES. (1996). "Text Corpora Working Group Reading Guide". Documento Eagles (Expert Advisory Group on Language Engineering) EAG-TCWG-FR-2.

Eckkrammer, E. M. (2009). *La comparación en los lenguajes de especialidad*. Berlín: Frank&Timme.

Flowerdew, J. (2001). "Concordancing as a tool in course design". En Ghadessy, M., Henry, A. y Roseberry, R. L. (Eds.), *Small Corpus Studies and ELT* (pp. 71-92). Amsterdam/Filadelfia: John Benjamins.

Flowerdew, L. (2000). “Investigating referential and pragmatic errors in a learner corpus”. En Burnard, L. y McEnery, T. (Eds.), *Rethinking language pedagogy from a corpus perspective* (pp. 145-154). Frankfurt: Peter Lang.

García Laborda, J. y Litzler, M. F. (2015). “Current Perspectives in Teaching English for Specific Purposes”. *Onomázein*, 31, pp. 38-51.

Gozdz-Roszkowski, S. (2011). *Patterns of Linguistic Variation in American Legal English. A Corpus-Based Study*. Frankfurt am Main: Peter Lang.

Henry, A. y Roseberry, R. L. (2001). “Using a small corpus to obtain data for teaching a genre”. En Ghadessy, M., Henry, A. y Roseberry, R. L. (Eds.), *Small Corpus Studies and ELT* (pp. 71-92). Amsterdam/Filadelfia: John Benjamins.

Hutchinson, T. y Waters, A. (1987). *English for Specific Purposes. A learning-centered approach*. Cambridge: Cambridge University Press.

Johns, T. F. (1991). “Should you be persuaded - Two samples of data-driven learning materials”. En Johns, T. F. y King, P. (Eds.), *Classroom concordancing. ELR Journal*, núm. 4, pp. 1-16.

Johns, T. F. (1994). “From printout to handout: Grammar and vocabulary teaching in the context of data-driven learning”. En Odlin, T. (Ed.), *Perspectives on pedagogical grammar* (pp. 27-45). Cambridge: Cambridge University Press.

Johns, T. F. (1997). “Contexts: The background, development and trialling of a concordance-based CALL program”. En Wichmann, A., Fligelstone, S., McEnery, T. y Knowles, G. (Eds.), *Teaching and language corpora* (pp. 100-115). London: Longman.

Johns, T. F. (2002). “Data-driven learning: The perpetual challenge”. En Kettemann, B. y Marko, G. (Eds.), *Teaching and learning by doing corpus analysis* (pp. 107-117). Amsterdam, The Netherlands: Rodopi.

Kennedy, G. (1998). *An introduction to corpus linguistics*. London: Longman.

Leech, G. (2007). “New resources, or just better old ones? The Holy Grail of representativeness”. En Hundt, M, Nesselhauf, N. y Biewer, C. (Eds.), *Corpus Linguistics and the Web* (pp. 133-49). Amsterdam: Rodopi.

López, C. I. (2000). “Tipologías textuales y géneros en la normalización terminológica y ortotipográfica de la traducción médica”. *Terminologie et traduction*, 3, 95-115.

McCarten, J. (2010). “Corpus-informed course book design”. En O’Keeffe, A. y McCarthy, M. (Eds.), *The Routledge Handbook of Corpus Linguistics* (pp. 413-427). London: Routledge.

McEnery, T. y Hardie, A. (2012). *Corpus Linguistics: Method, theory and practice*. Cambridge: Cambridge University Press.

Nesselhauf, N. (2004). “Learner corpora and their potential for language teaching”. En Sinclair, J. McH. (Ed.), *How to Use Corpora in Language Teaching*, (pp. 125-155). Amsterdam/Filadelfia: John Benjamins.

Robinson, P. (1991). *ESP Today: A Practitioner’s Guide*. London: Prentice Hall International.

Rodríguez Piñero, A.I. y García Antuña, M. (2009). “Lenguas de especialidad y lenguas para fines específicos: precisiones conceptuales y terminológicas e implicaciones didácticas”. En Vera, A. y Martínez, I. (Eds.), *El español en contextos específicos: enseñanza e investigación* (pp. 907-932). Comillas: Fundación Comillas.

Román, V. (2012). “Algunas peculiaridades de la traducción inglés-español del discurso de los negocios”. En Martino Alba, P. & Lebsanft, Ch. (Eds.), *Telar de traducción especializada* (pp. 91-103). Madrid: Dykinson.

Santamaría-García, C. (1995). “Corpora for English language teaching and learning”. *Encuentro: revista de investigación e innovación en la clase de idiomas*, núm. 8, pp. 190-198

Santamaría-García, C. (1998). “Corpora y Programas de Concordancias en la Formación de Maestros Especialistas en Lengua Inglesa”. En Rodríguez Marcos, A., Sanz Lobo, E. y Sotomayor Sáez, M. V. *La formación de los maestros en los países de la Unión Europea* (pp. 281-287). Madrid: Narcea.

Santamaría-García, C. (2011). “Bricolage assembling: CL, CA and DA to explore agreement”. *International journal of corpus linguistics*, vol. 16, núm. 3, pp. 345-370.

Schank, R. C. (1995). *What we learn when we learn by doing*. Evaston, IL: Northwestern University Press.

Sinclair, J. M. (1991). *Corpus, Concordance, Collocation*. Oxford: OUP

Sinclair, J. M. (2001). “Preface”. En Ghadessy, M., Henry, A. y Roseberry, R. L. (Eds.), *Small Corpus Studies and ELT* (pp. vii-xv). Amsterdam/Filadelfia: John Benjamins.

Stubbs, M. (1995). Collocations and semantic profiles: On the cause of the trouble with quantitative methods. En *Function of* Language, *2/1,* pp. 1-33.

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

Tolchinsky, L. (2014). “El uso de corpus lingüísticos como herramienta pedagógica”. *Textos de Didáctica de la Lengua y de la Literatura*, núm. 65, pp. 9-17.

Walter, E. (2010). “Using corpora to write dictionaries”. En O’Keeffe, A. y McCarthy, M. (Eds.), *The Routledge Handbook of Corpus Linguistics* (pp. 428-443). London: Routledge.

Leticia Moreno Pérez es profesora de inglés para fines específicos y su traducción en la Universidad de Valladolid. Especializada en el campo jurídico-económico, su investigación se centra en el lenguaje especializado y su traducción desde una perspectiva funcional a través de los estudios basados en corpus. Ha publicado y participado como ponente en congresos internacionales sobre este campo, en el cual se encuadra, además, su Tesis Doctoral. Es miembro del grupo ACTRES (Análisis Contrastivo y Traducción Inglés-Español) en el que colaboran investigadores de distintas universidades españolas e internacionales.

Recibido: 16–5–2016
Aceptado: 3–1–2017