

Integración de los corpus como herramienta de apoyo en la enseñanza de ESP

VICTORIA LÓPEZ SANJUÁN

UNED, Madrid

Recibido: 19 abril 2007 / Aceptado: 17 febrero 2008

ISSN: 1697-7467

ABSTRACT: Recently, the application of corpora to language learning has been studied, due to their condition of real texts. However, in order to consider the utility of this application, it's needed to show it from two different perspectives: the teacher and the learner, because they are the potential users in the learning process. Moreover, here the integration of corpora in the learning process is seen as an assistance tool, and an easy way to compile an ESP corpus from the Internet.

Key words: corpora, ESP, Internet, learning.

RESUMEN: En los últimos años se ha estudiado la aplicación de los corpus a la enseñanza de lenguas, dada su condición de textos reales. Sin embargo, para ver la validez de la aplicación de los corpus a la enseñanza de lenguas es necesario mostrarlo desde dos perspectivas diferentes: la de los aprendientes y la de los docentes que son quienes van a hacer uso de ellos en el proceso de aprendizaje. Además, aquí se propone su integración en el proceso de aprendizaje como herramienta de apoyo y se muestra un modo sencillo de compilar un corpus de ESP a través de Internet.

Palabras clave: corpus, fines específicos, Internet, enseñanza.

1. DEFINICIÓN DE CORPUS

Un corpus es un conjunto de textos recopilados con una finalidad lingüística, que sirve como fuente de información para demostrar algún aspecto concreto de la lengua a la que pretende representar y que facilita que se hagan generalizaciones a partir de los datos que hay en su contenido (López Sanjuán, 2006:31). Así, un corpus es una colección de textos *representativos* de otra entidad lingüística igual o mayor que tiene como finalidad la investigación científica sobre la lengua o alguno de sus aspectos frente al mero carácter informativo e ilustrativo de una antología. Además, en la actualidad entendemos por corpus una colección de textos *reales electrónicos* susceptibles de ser procesados electrónicamente por un programa de ordenador. Es importante determinar el formato electrónico del corpus, pues sería imposible aplicar cualquier tipo de herramienta informática a un corpus en papel.

Aston (1998:205) precisa que el término *corpus* se refiere a los que están disponibles públicamente o los que los docentes o investigadores construyen *ad hoc* para fines particulares. Esta explicación no es imprescindible, pues siempre que cumplan los requisitos especificados en la definición, se trata de un corpus independientemente de cual sea su fuente

original. Incluso una antología se podría utilizar como un corpus, por ejemplo, en una antología de poemas de W.B. Yeats adecuadamente seleccionada se podría hacer un estudio de las colocaciones, la frecuencia de alguna palabra o estructura sintáctica, etc. para caracterizar la lengua del autor.

Entre las ventajas que ofrecen los corpus destacan:

- * La exactitud del contenido, pues representan datos reales, no datos inventados.
- * La facilidad de procesamiento, potencia y eficacia (sobre todo si los comparamos con los estudios manuales).
- * La obtención de afirmaciones más objetivas que las que se consiguen con la introspección.
- * La aportación de evidencia objetiva y de un modo útil de analizarla.
- * Son fuentes de información esenciales para áreas de la lingüística aplicada como la enseñanza de lenguas, traducción automática, reconocimiento del habla, correctores gramaticales o de estilo y un largo etcétera.
- * Sirven de asistencia a hablantes no nativos de una lengua.

2. UTILIDAD DE LOS CORPUS EN LA ENSEÑANZA DE LENGUAS

Los corpus son valiosos en la enseñanza de lenguas por su condición de conjuntos de textos *reales* de una lengua, aplicables, por ejemplo, a la práctica con ejercicios extraídos directamente de ellos.

Braun (2005:47) apunta que:

The use of corpora in the second-language learning context requires the availability of corpora which are pedagogically relevant with regard to choice of discourse, choice of media, annotation and size.

Esta afirmación pone de manifiesto algunas de las características de los corpus, como son la representatividad y el tamaño. Kennedy (1991:366) señala cuatro características cruciales de la lengua de estudio que han de estar presentes en los textos que seleccionemos como muestra de dicha lengua: dominio, tema, registro y medio de expresión. Así, por ejemplo, sólo un conjunto de informes meteorológicos en inglés puede servir como corpus para el estudio de informes meteorológicos en inglés.

En palabras de Barnbrook (1996:24), «*the question of representativeness underlies the whole question of corpus*», señalando así la representatividad como uno de los factores cruciales a tener en cuenta al diseñar un corpus lingüístico. Un corpus es representativo cuando los resultados obtenidos de su contenido se pueden generalizar a una entidad lingüística mayor (Leech 1991:27). Teniendo presente esto, para formar un corpus de una variedad lingüística cualquiera necesitamos seguir estrictamente los mismos parámetros o criterios de selección acordes con que la lengua que se pretende que represente. Por ejemplo, si el objeto de estudio es la lengua de los adolescentes varones de Newcastle, ¿por qué incluir textos de mujeres, o incluso de varones de más edad? Además, deberíamos cubrir proporcionalmente toda el área geográfica, el período de años a estudiar, los estamentos sociales y el nivel educativo, etc., en suma, todos los grupos que sean relevantes a la investigación con porcen-

tajes adecuados según existan en la comunidad lingüística, asegurando así la fiabilidad y representatividad del corpus.

El tamaño de un corpus está muy relacionado con su capacidad de representatividad y los objetivos investigadores o la utilidad que se le quiera dar al corpus, es decir, su adecuación. Por ejemplo, si queremos examinar la co-ocurrencia de una palabra determinada en el sublenguaje médico en inglés, de poco serviría tener un corpus muy amplio, pero general de la lengua inglesa, en el que no estuviese representado este sublenguaje. En algunos casos, corpus de 20.000 palabras han sido suficientes para obtener resultados estadísticamente completos (por ejemplo, en el caso del corpus COPLA [López Sanjuán 2003], ofreciendo resultados de presencia, frecuencia y distribución constantes, pero incluso a veces, estos corpus son demasiado pequeños sublenguajes reducidos, si se pretenden examinar interacciones con otros sublenguajes o fenómenos no muy comunes.

Respecto a cifras exactas, las únicas a las que se hace referencia en la literatura es el *número mínimo* de palabras que ha de tener un texto. Así pues, Lehrberger y Bourbeau (1988:141) apuntan que para un sublenguaje completamente cerrado (en cuanto a su vocabulario), 10.000 palabras es el número mínimo aceptable.

Desde principios de la década de 1990, los corpus han sido claramente consultados por docentes de lenguas, pero también por aprendientes (Chambers, 2005), al menos en cursos dirigidos por investigadores. Esto ha creado la necesidad de que la investigación apuntale este nuevo desarrollo, enfocado en aspectos como el tipo de corpus que se consulta, pequeño o grande, generales o específicos de un dominio (por ejemplo, para la enseñanza de ESP), anotados o no anotados.

Godwing-Jones (2001:7) apunta que, entre otros usos, las listas de vocabulario que han sido generadas por los corpus y el recuento de palabras extraído del análisis de un corpus ayudan a definir los objetivos de la adquisición del vocabulario. Esto está muy presente en la enseñanza de lenguas para fines específicos (ESP), por ejemplo, en inglés de negocios o de turismo, dado que los corpus de este tipo facilitan al docente la tarea de identificación de los términos específicos que tienen que ser enseñados y su secuenciación aproximada. Obviamente, esto se debe a que los dominios bien definidos y acotados, cuyas descripciones, terminología, etc. son sistemáticos y limitados facilitan el aprendizaje.

Los corpus que no contienen ningún tipo de anotación ilustran el uso real de la lengua, mostrando las ocurrencias de determinadas estructuras o palabras. Estos corpus se pueden utilizar para realizar análisis cuantitativos y obtener, por ejemplo, cuáles son las palabras más frecuentes en un tipo de textos, por ejemplo, de inglés de marketing, las colocaciones o para buscar estructuras que pueden ir separadas y sirve de criterio de diferenciación entre los verbos frasales y los preposicionales y, por lo tanto, se podría realizar una búsqueda de la co-ocurrencia de *turn* y de *off* por ejemplo, y descubrir cuando *off* es la partícula del verbo frasal (y, por lo tanto, puede llevar en el medio un sintagma nominal como *the cold tap* o una preposición unida siempre al verbo: **turn the road off*). Por su parte, los corpus anotados, ya sea morfológica o sintácticamente, son de más utilidad sólo para los docentes que están familiarizados con la terminología y los análisis lingüísticos y saben interpretarlos. En la didáctica y práctica de la traducción también destaca el uso de los corpus paralelos¹, que

1. Los corpus paralelos contienen textos con sus equivalencias en otras lenguas (véase López Sanjuán, 2007).

permiten acceder y reflexionar sobre las diferencias entre las equivalencias de las palabras o construcciones, teniendo que ser evidentemente una de las lenguas la nativa del aprendiente.

También hay que tener en cuenta que los corpus diseñados con muestras de hablantes nativos de una lengua contribuyen a que se pueda llevar a cabo una descripción más precisa de la lengua que representan, y, permiten la creación de gramáticas y diccionarios (Hunston y Francis, 1999). Además, si se estudia la lengua del aprendiente se obtiene información para crear gramáticas y diccionarios específicos orientados a éste (por ejemplo, *Cambridge Advance Learner's Dictionary* o *Oxford Dictionary of Business English for Learners of English*).

Los libros de texto permiten que se estudie la lengua a la que están expuestos los aprendientes, y cuando se comparan con un corpus de L1 (lengua nativa), facilitan el desarrollo de materiales pedagógicos más efectivos.

Fligestone (1993:103) señala que son tres básicamente las etapas en el proceso de aplicación de los corpus a la enseñanza de lenguas por parte del docente (o del aprendiente en un entorno de educación a distancia o de autoaprendizaje):

- * **Enseñar sobre los propios corpus:** los corpus como fuente de referencia para mostrar la lengua «real». El docente hace uso del corpus para mostrar a los alumnos ejemplos reales. Si, por ejemplo, se ha recopilado un corpus con folletos de lugares turísticos de vacaciones en inglés y español, el docente puede mostrar a los alumnos ejemplos de las palabras más frecuentes que aparecen en este tipo textual.
- * **Enseñar a explotar los corpus:** en esta etapa se enseña la metodología necesaria para poder estudiar y analizar un corpus, así como las herramientas que se pueden utilizar en este proceso. El docente aprende cuáles son las herramientas computacionales que se pueden utilizar para obtener, por ejemplo, las colocaciones que son propias, si es que las hay, de las reservas de hotel hechas por teléfono o a través de Internet.
- * **Explotar para enseñar:** una vez que ya se han aprendido las técnicas y métodos que se pueden aplicar a los corpus, se procede a su explotación para, por ejemplo, encontrar muestras o ejemplos de una determinada construcción y asistir al docente, si fuese necesario, a discernir los casos de aplicación de la misma. Por ejemplo, el docente utiliza el corpus y las herramientas para encontrar concordancias de la palabra *book* (que en el inglés general es *libro*), pero que dentro del contexto de los folletos turísticos suele referirse a reservar. Por ejemplo, en muchos casos los estudiantes de turismo confunden el término e indican «*I have a book*».

Cabe matizar lo expuesto por Fligestone, ya que, en primer lugar, los docentes tienen que ser entrenados en el uso de los corpus, porque solo de ese modo se puede esperar que introduzcan los métodos aplicados al estudio de los corpus en la clase y conseguir que los aprendientes se involucren en las actividades basadas en corpus. El objetivo que se persigue con este proceso es conseguir que los aprendientes trabajen de un modo autónomo con los corpus. Esto se puede observar en la figura 1:

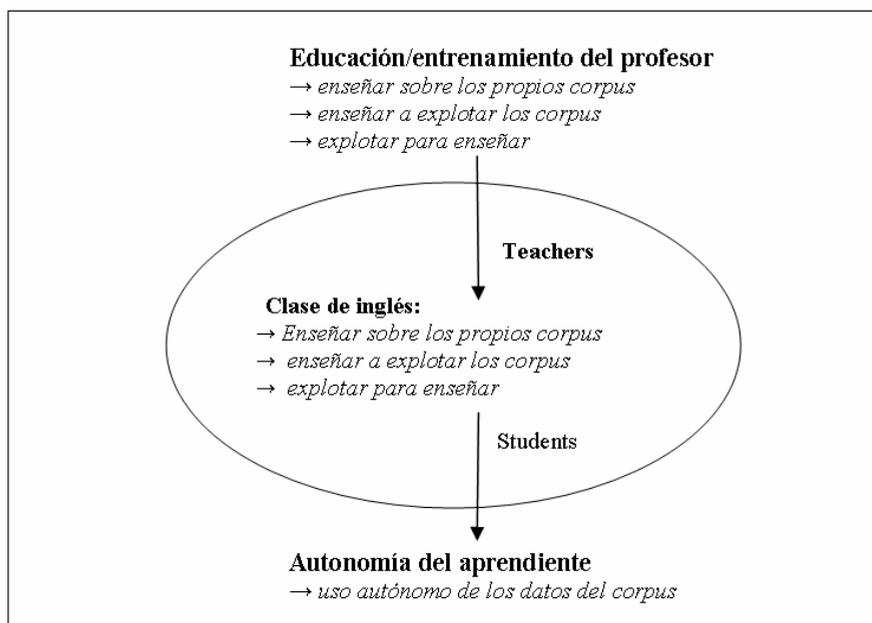


Figura 1. Aplicación de los corpus a ELT (adaptado de Mukherjee, 2004: 245).

2.1. Dos enfoques: docentes y aprendientes

Para ver la validez de la aplicación de los corpus a la enseñanza de lenguas es necesario mostrarlo desde dos perspectivas diferentes, la de **los aprendientes** y **la de los docentes** que, en realidad, son quienes van a hacer uso de ellos en el proceso de aprendizaje. Sin embargo, cabe destacar que desde cualquiera de las dos perspectivas, los corpus son vistos igualmente como una fuente de información.

Los corpus se enfocan hacia el *producto* del análisis (Aston, 1998:207). El conocimiento extraído o derivado de ellos es, en términos cognitivos, *declarativo*, es decir, que expone una serie de hechos sobre la lengua y la cultura. Este conocimiento puede no ser exacto, pues dependerá de la adecuación de los datos del corpus y de la habilidad del usuario para analizarlo y hacer generalizaciones a partir de los datos obtenidos. Este conocimiento declarativo implica en casi todos los casos *simplificación*, ya que la adquisición de una segunda lengua efectivamente implica aproximaciones sucesivas al sistema de destino.

Sin embargo, observando el uso de los corpus por parte de los aprendientes y de los docentes o investigadores, se debería considerar un segundo aspecto en este uso enfocado en el *proceso analítico*. El trabajo con corpus, además, puede proporcionar un medio de desarrollar el conocimiento *procedural*, en el sentido que de la habilidad para utilizar una lengua puede llevar a solucionar problemas operativos.

Por último, como indica St. John (2001:185) la utilización de corpus y un programa de concordancias puede tener un efecto motivador y de recompensa no sólo para los docentes,

sino también para los aprendientes. Para el docente, estos programas pueden proporcionar ejemplos contextualizados para resolver cuestiones léxicas dudosas. Además, el aprendiente puede desarrollar una habilidad de «aprender como aprender» (Johns, 1991:1) permitiéndole asumir el papel de un explorador. Esto viene avalado por Barlow (1996:2) quien afirma que «*one of the roles the language learner plays when using corpora is that of a language researcher*» y explica porqué «*a suitable research environment must be provided*» (Barlow, 1996:45).

2.1.1. Utilización de los corpus por parte de los docentes

Los corpus pueden ser utilizados por los docentes, entre otras cosas, como fuente de comprobación de la materia que están enseñando dado que, como señala Godwin-Jones (2001:7): «*corpora are, after all, huge storehouse of real language use*».

Entre algunos de los usos que los corpus pueden ofrecer a los docentes, destacan:

- * **Ilustración de la lengua en un contexto real:** con un corpus se pueden buscar ejemplos concretos de una palabra, construcción o estructura sintáctica, por ejemplo, utilizando programas de concordancias. (Barlow, 1992; Scott, 2004; Scott y Tribble, 2006 y Pérez Guerra, 1998). Según una afirmación de Oakes (1998:150), «*concordances are the oldest application of computer in the humanities*», debido, sin duda, a la simplicidad algorítmica que conlleva su desarrollo informático. De este modo, se puede definir un programa de concordancias como un programa que «*searches a corpus for a selected word or word phrase and present every instance of that word or phrase in the centre of the computer screen, with the words that come before and after it to the left and right*». (Hunston, 2002:39).
- * **Creación de ejercicios:** utilizando un corpus en inglés, se pueden realizar ejercicios de construcción de vocabulario y de exploración de rasgos gramaticales y discursivos de un texto. Por ejemplo, a partir de corpus anotados a nivel de categoría se puede elaborar un ejercicio de *fill in the gaps* de preposiciones. Más información, principalmente aplicada a la enseñanza del inglés como lengua extranjera, se puede obtener en Mindt (1997) y Flowerdew (1996).
- * **Detección de errores:** un corpus de errores permite estudiar y consecuentemente descubrir y clasificar los errores cometidos por los aprendientes nativos en el estudio de una segunda lengua. En un corpus de errores paralelo se pueden observar las diferencias entre equivalencias y descubrir, por ejemplo, si los errores son «falsos amigos». Si el docente crea un corpus con, por ejemplo, redacciones de sus estudiantes, podrá observar donde cometen más errores. Por ejemplo, en un corpus obtenido de estudiantes de turismo, se podría apreciar que tienen dificultades a la hora de utilizar correctamente las preposiciones, por ejemplo, *in the second floor* en vez de *on the second floor*. También es importante mencionar que se han elaborado corpus específicos orientados a la enseñanza de lenguas, recopilados desde el principio con un fin didáctico, que permiten, por ejemplo, descubrir y analizar los errores sistemáticos

de los aprendientes de una segunda lengua, extrayendo conclusiones para la mejora del método de enseñanza que se ha estado utilizando en el aula hasta ese momento. Son los llamados «*learner corpora*» (Granger, 2003), que recopilan textos de hablantes no nativos de una lengua. Un ejemplo de corpus de este tipo es el Corpus SULEC (*Santiago University Learner of English Corpus*)², creado por el Departamento de Filología Inglesa de la Universidad de Santiago de Compostela.

- * **Elaboración de materiales didácticos:** no sólo se pueden realizar ejercicios con los corpus para su utilización en el aula, sino también, por ejemplo, glosarios o breves diccionarios (sobre todo si se trata de lenguajes especializados, como el del turismo), así como programas docentes para cursos de idiomas y libros de texto (además de cuadernos de ejercicios). En este sentido, cabe señalar que algunos de los diccionarios de referencia del inglés más utilizados están elaborados a partir de corpus, por ejemplo, *Collins COBUILD English Dictionary*, *Oxford Advanced Learner's Dictionary* o *Longman Dictionary of Contemporary English* y también alguna de las gramáticas más recientes, por ejemplo, *Longman grammar of spoken and written English*.

2.1.2. Utilización de los corpus por parte de los aprendientes

La utilización de los corpus por parte de los estudiantes de lenguas presenta una variación importante en relación con otros métodos de enseñanza. Como apuntan McEnery y Wilson (1997:6) «*The corpus encourages the student to act as the producer of research, rather than its passive receptacle*».

De este modo, el aprendiente pasa a ser un agente importante en su proceso de aprendizaje y, en vez de que adquiera el conocimiento de un modo pasivo, con el uso de un corpus puede elaborar y desarrollar su propio aprendizaje. El objetivo de la utilización de corpus en el proceso de aprendizaje por parte de los aprendientes, es que consigan ver los corpus como cualquier otro recurso de apoyo en la enseñanza de lenguas, por ejemplo, como el libro de texto, el diccionario o la gramática de consulta. Se trata de que descubran que los corpus, por ejemplo, pueden ser tan útiles como un diccionario, al que acuden cuando tienen alguna duda de vocabulario.

Chambers (2005) realizó un estudio para descubrir la opinión de los estudiantes sobre el uso de los corpus y los programas de concordancias. En algún caso, se valoró positivamente el hecho de que los corpus contuviesen muestras y ejemplos *reales*, en contraste con los inventados de los libros de texto y gramáticas, los cuales algunos estudiantes consideran irreales y a veces «estúpidos». Otro de los rasgos considerados positivos por los estudiantes es el hecho de que los corpus recogen, en muchos casos, información actualizada, por ejemplo, de noticias recientes. Por estos dos motivos (carácter real e información más o menos conocida), los lleva a considerar en algún caso que los términos y ejemplos son más fáciles de memorizar. Además, también se valoró de un modo positivo el gran número de ejemplos que pueden proporcionar los corpus, también en contraste con el número limitado que ofrecen libros de texto, diccionarios y gramáticas. Para esta autora, «*these positive reactions suggest*

2. Para más información véase el sitio web del corpus: <http://www.usc.es/ia303/SULEC>

that corpora and concordancing certainly have their place in a language-learning environment focusing on learner autonomy and discovery learning». Como ella misma señala, la utilización de corpus favorece la autonomía del aprendiente y el descubrimiento del aprendizaje.

Se pueden identificar, entre otros, algunos de los usos que los corpus pueden ofrecer a los estudiantes de lenguas, por ejemplo:

- * **Descubrimiento de las características de la lengua:** a través de los corpus, los estudiantes pueden observar y descubrir por sí mismos las características propias de la lengua. Por ejemplo, a partir de un corpus anotado sintácticamente se puede ver, por ejemplo, cuando un gerundio actúa como sujeto de una oración, como modificador nominal o como núcleo de una cláusula subordinada. También permite observar las circunstancias en que ciertos adjetivos pueden ir en posición predicativa y cuáles en atributiva.
- * **Estudio contextualizado del vocabulario:** los corpus proporcionan a los estudiantes múltiples contextos para aprender nuevo vocabulario. Por ejemplo, en el estudio de vocabulario de inglés de negocios, un corpus serviría para observar, por ejemplo, con qué verbos se combinan los términos de fenómenos bursátiles y así aprender a expresarse como un experto nativo en este dominio.

2.2. Aplicación actual de los corpus en el proceso de aprendizaje

Tribble (2000: 31) admite que *«not many teachers seem to be using corpora in their classrooms.»* Mukherjee (2004:239) comparte esta opinión y añade que la influencia de la aplicación de la investigación sobre corpus en la enseñanza del inglés es todavía limitada. Gabrielatos (2005) considera que el término ‘corpus’ se ha convertido en un slogan o palabra de moda en la enseñanza de lenguas y que tanto los docentes como los aprendientes de lenguas reconocen de inmediato, puesto que son consumidores de estos productos (muchos diccionarios o gramáticas basados en corpus) pero no saben de qué modo han sido elaborados y qué son los corpus ni para que sirven. Aunque estos diccionarios o gramáticas se supone que muestran ejemplos reales del uso de la lengua, ni docentes ni aprendientes son conscientes de que parten del análisis cuantitativo de un corpus representativo. Gabrielatos añade que la investigación y enseñanza basada en corpus se ha llevado a cabo predominantemente a nivel universitario, lo que puede haberlos llevado a pensar que los corpus no son relevantes en su ámbito didáctico o que al no disponer de los conocimientos, destrezas o incluso de los medios técnicos necesarios, están fuera de su alcance. Además, considera que al no tener conocimientos previos sobre corpus esto dificulta su difusión y que cualquier esfuerzo para popularizar el uso de los corpus en la enseñanza de lenguas debe ir enfocado a las ideas preconcebidas y a las necesidades de esos docentes.

En lo que se refiere a los aprendientes, éstos incluso utilizan en menor medida los corpus que los docentes. Realmente, es la consecuencia lógica de lo explicado anteriormente. Si los docentes no integran los corpus en sus clases y les explican a sus alumnos qué son y cómo utilizarlos, éstos no pueden tener constancia de para qué sirven, las posibilidades que les ofrecen o cómo procesarlos para obtener datos o ejemplos que apoyen su aprendizaje.

3. PRIMEROS PASOS EN EL USO DE LOS CORPUS

Como se ha indicado anteriormente, los docentes y los aprendientes no son conscientes de las posibles aplicaciones que los corpus pueden tener tanto en el proceso de aprendizaje de los aprendientes, como en el propio proceso pedagógico, sirviéndoles a ambos como herramienta de apoyo.

Un docente que pretende hacer uso de un corpus o bien para elaborar materiales, buscar ejemplos o integrarlo en el aula, y habiéndose familiarizado y asumido lo que es un corpus, necesita saber cuáles son sus opciones (acceder a uno disponible, recopilar el suyo propio para adaptarlo mejor a sus necesidades) y seleccionar la mejor dentro de sus posibilidades, recursos y, sobre todo, el más adecuado a su objetivo.

En esta sección se muestran algunos puntos de partida a los docentes (o, por qué no, también a los aprendientes), así como una serie de herramientas de análisis disponibles para que puedan empezar a dar sus primeros pasos y adentrarse en el mundo de los corpus.

3.1. Corpus

Aunque las opciones a la hora de obtener un corpus son diferentes y, en muchos casos, dependen de los recursos de los que disponga el docente, aquí se van a plantear tres diferentes: la utilización de un corpus ya creado y que esté disponible; la utilización de la web como un corpus y, por último, la creación de su propio corpus a partir de textos que ya están en formato electrónico (CD-ROM, sitios de archivos textuales o Internet). No hay que confundir entre la utilización de la web como un corpus y su utilización como fuente de información para obtener textos y recopilar un corpus. Son dos usos diferentes que se explicarán a continuación.

3.1.1. *Corpus disponibles*

El docente puede hacer uso de corpus que ya están disponibles para llevar a cabo sus investigaciones. Hay que indicar que, aunque un corpus sea de libre disposición, eso no significa que sea gratuito. Cabe señalar, además, que muchos de ellos sí lo son para fines educativos, sin ningún tipo de propósito comercial. Como ya se ha indicado anteriormente, los docentes pueden utilizarlos para elaborar sus materiales, preparar sus clases o incluso llevarlos al aula para que los aprendientes puedan hacer uso de ellos.

Alguno de los corpus más importantes recogidos hasta el momento (monolingües en inglés y multilingües) son:

- * *The British National Corpus (BNC)*: contiene 4124 textos con alrededor de 100 millones de palabras de inglés británico, de las que 90 millones proceden de material escrito y el 10% restante de material oral. Este corpus tiene un servicio de búsqueda en línea. La dirección de su web es: www.hcu.ox.ac.uk/BNC.
- * *The Bank of English*: es la colección de textos de la Universidad de Birmingham y de la editorial COBUILD, que a comienzos de los 90 empezaron a recopilar mues-

tras de inglés moderno y hoy contiene alrededor de 450 millones de palabras de inglés oral y escrito de diferentes registros y sigue aumentando sin cesar. Está anotado y disponible en CD-ROM. Más información puede encontrarse en la página web del proyecto: http://titania.cobuild.collins.co.uk/boe_info.html.

- * *The Longman-Lancaster Corpus*: alrededor de 30 millones de palabras escritas de diferentes variedades del inglés recopiladas en diferentes fechas y tipos de textos. Fue construido cuidadosamente para ser representativo de la lengua escrita del siglo XX. Consta sólo de texto ortográfico, sin anotaciones. También se puede encontrar información general sobre este corpus en la dirección: www.longman-etl.com/dictionaries/corpus/lclonlan.html.
- * *CRATER*: es un corpus trilingüe de alrededor de un millón de palabras de inglés, español y francés. Está compuesto por textos del dominio de las telecomunicaciones y etiquetado gramaticalmente en las tres lenguas. Se puede acceder a él en línea y para más información se puede contactar con Tony McEnery en la dirección: mcenery@comp.lancs.ac.uk.
- * *Corpus multilingüe de la Universidad Pompeu Fabra*: se trata de textos en catalán, español, francés, inglés y alemán agrupados según áreas de especialización: ley, medio ambiente, medicina, economía e informática. Se puede obtener más información en: <http://www.iula.upf.es/corpus/corpus.htm>.

Algunos de estos corpus tienen un servicio de búsqueda en línea que permite descargar citas de él desde un ordenador con conexión a Internet. La figura 2 muestra la interfaz del servicio de búsqueda del BNC.

The screenshot shows the British National Corpus (BNC) website. At the top, the title 'BRITISH NATIONAL CORPUS' is displayed in large, stylized letters. Below the title is a navigation menu with links: Home, The Corpus, Using, Obtaining, Xaira/SARA, FAQ, Archive, Contact Us, and A-Z. On the left side, there is a sidebar with two main sections: 'About' and 'Using the BNC'. The 'About' section contains links for 'What is the BNC?', 'Creating the BNC', 'BNC Products', 'BNC XML Edition', 'Copyright', 'Contact Us', and 'Contents A-Z'. The 'Using the BNC' section contains a link for 'What can I do with the BNC?'. The main content area is divided into two columns. The left column has a section titled 'About the BNC' which states: 'The British National Corpus (BNC) is a 100 million word collection of samples of written and spoken language from a wide range of sources, designed to represent a wide cross-section of current British English, both spoken and written. [more]'. Below this is a section titled 'Search the Corpus' which says: 'Type a word or phrase in the search box and press the Return key on your keyboard to see up to 50 random hits from the corpus.' There is a search box labeled 'Look up:' followed by an empty input field. Below the search box, it says: 'You can search for a single word or a phrase, restrict searches by part of speech, search in parts of the corpus only and much more.' At the bottom of this section, it says: 'The search result will show the total frequency in the corpus and up to 50 examples. [more information]'. The right column shows a preview of search results, displaying XML-style markup for words like 'issue' and 'should'.

Figura 2. Interfaz del British Nacional Corpus en Internet.

3.1.2. *Utilización de la web como un corpus*

Algunos autores (Fletcher, 2001; Meyer, 2006) apoyan la utilización de la web como si fuese un corpus. Basan su decisión en la cantidad de documentos que hay en Internet, lo que permite que el número de ejemplos encontrado sea mucho mayor que en el corpus, por su tamaño finito. Además, permite la búsqueda en diferentes lenguas y géneros, mientras que un corpus es más limitado. Otra ventaja a tener en cuenta es la ingente cantidad de nuevas páginas que se crean a diario y que permite mantener la información actualizada, mientras que en un corpus ya creado la información no se actualiza.

Fletcher (2001) indica que usa la web para enseñar gramática y vocabulario, buscando ejemplos que muestren como funciona una lengua. También le permite obtener indicios y evidencias de la presencia y distribución de posibles usos (por ejemplo, de variedades dialectales) y de palabras que están presentes en la lengua, pero no en diccionarios (posibles neologismos o anglicismos, por ejemplo).

Por su parte (Meyer 2006) la Web puede servir de recurso para el aprendizaje de la modalidad en inglés por futuros profesores de inglés para segunda lengua, inglés como lengua extranjera e inglés para fines específicos.

Salzman and Mills (2001) en sproponen el uso de los navegadores (Internet Explorer, Netscape, Opera, Mozilla, etc.), y de un motor de búsqueda (Google, Yahoo, etc.) para que sus alumnos de la Universidad de Illinois busquen ejemplos de las estructuras que están aprendiendo. Los denominan «Grammar Safaris». Los alumnos buscan ejemplos en documentos o sitios web, que posteriormente copian y pegan en un documento creado por un procesador de textos y lo llevan a la clase para su puesta en común. Estos autores consideran que este enfoque les proporciona mucho material para comentar en el aula.

3.1.3. *Fuentes para la creación de un corpus propio*

Algo de lo que no hay que olvidarse son los permisos legales que se necesitan para la utilización de algunos textos. Es necesario conocer y reconocer la propiedad intelectual y los derechos de autor que tienen los textos que se van a utilizar y solicitar por escrito a su propietario el permiso necesario para su utilización y tratamiento computacional y, eventualmente, para la divulgación de resultados, especificando la utilidad y difusión que se les va a dar.

1. CD-ROM

Existe mucho material escrito actualmente disponible en CD-ROM tanto de manera comercial como gratuita, distribuida, por ejemplo, a través de periódicos o revistas de divulgación. Por ejemplo, si se quisiera hacer algún tipo de estudio sobre el lenguaje jurídico, se podría utilizar una copia de la Constitución Española de 1978 y otras legislaciones que están disponibles en este formato. Algunos CD-ROM se adquieren de forma progresiva, al distribuirse periódicamente actualizaciones de los mismos, lo que ayuda a mantener vigente su valor durante cierto tiempo.

2. Sitios textuales

Se han creado en los últimos años una serie de bibliotecas de corpus o archivos virtuales en donde se puede encontrar gran cantidad de material de corpus. Destacan entre otros:

- * *The European Language Resources Association* (ELRA), es una organización sin ánimo de lucro, que fomenta y apoya la creación de recursos lingüísticos para las tecnologías lingüísticas. Más información disponible en su sitio web: <http://www.icp.grenet.fr/ELRA/home.html>.
- * *The International Computer Archive of Modern and Medieval English* (ICAME) es **una organización formada por lingüistas e informáticos cuyo objetivo es recoger y distribuir corpus electrónicos en lengua inglesa y poner este material a disposición de los centros de investigación**. Más información disponible en su sitio web: <http://www.hit.uib.no/icame.html>.
- * *Linguistic Data Consortium* (LDC) es **un consorcio abierto de universidades, empresas y laboratorios dependientes del gobierno, que crea, recoge y distribuye recursos lingüísticos, entre otros, datos, herramientas y estándares**. Más información disponible en su sitio web: <http://www ldc.upenn.edu>.
- * *Oxford Text Archive* (OTA), es un servicio de *Oxford University Computing Service*, que proporciona a la comunidad académica textos electrónicos y corpus lingüísticos en diferentes lenguas. Más información disponible en su sitio web: www.ota.ahds.ac.uk.

3. Internet

Es una fuente inagotable de material para formar prácticamente cualquier tipo de corpus. En la red se encuentra una ingente cantidad de páginas disponibles que contienen material de todos los temas. De hecho, se encuentran páginas especializadas, que garantizan una alta fiabilidad de los contenidos y de su representatividad.

Un inconveniente se puede encontrar en que los textos presentan el formato HTML y puede ser necesaria una post-edición manual, ya que los documentos en formato HTML pueden contener gráficos y tablas y, además, las etiquetas de HTML podrían interferir en el funcionamiento de las herramientas de procesamiento a utilizar y alterar los resultados, contabilizándose las propias etiquetas como palabras del texto.

3.2. Herramientas

A continuación se enumeran algunas herramientas de análisis de corpus, especialmente aquéllas que contienen programas de concordancias. Cabe señalar que no se han incluido etiquetadores morfológicos o analizadores sintácticos, puesto que en este artículo no se ha profundizado en el uso de estas herramientas y sus posibilidades.

* TACT

Tipo de herramienta: Herramienta de análisis textual que proporciona análisis cuantitativos.

Disponibilidad: Gratuito para fines académicos. Se puede descargar desde la página web: <http://www.gla.ac.uk/Inter/Computerpast/ctich/WebFTP/tact.html>.

Descripción: TACT es un programa de análisis textual que está compuesto por una serie de programas, quince en total, que son independientes en funcionamiento. La última versión es la 2.1 del año 1995, a pesar de lo cual es todavía popular en la actualidad por su robustez y fácil uso. Sólo se puede ejecutar bajo el sistema operativo MSDOS. Hay una versión en la web (<http://tactweb.humanities.mcmaster.ca/>) que permite el acceso a las bases de datos de TACT. El sistema puede realizar análisis cuantitativos simples, como listas de frecuencias, búsqueda de concordancias y colocaciones o modelos de palabras. Además, permite que se puedan realizar análisis de textos en lenguas diferentes.

* **WORDCRUNCHER**

Tipo de herramienta: Herramienta de análisis textual.

Disponibilidad: Gratuito. Se puede obtener a través de ICAME: www.hit.uib.no/icame.cd

Descripción: Wordcruncher es un programa de análisis textual que tiene una versión para Windows y una anterior para MS DOS. Está formado por dos programas: WCView y WCIndex, que realizan la búsqueda de palabras, sintagmas u oraciones y producen listas de frecuencias y colocaciones. Permite que se pueda trabajar con textos paralelos y lenguas diferentes al inglés. Es una herramienta de análisis fácil de utilizar y también muy rápida.

* **WORDSMITH TOOLS 4.0**

Tipo de herramienta: Herramienta de análisis textual que proporciona análisis cuantitativos.

Disponibilidad: Se puede descargar una versión de prueba de la página web del autor. A cargo de su comercialización se ocupa Oxford University Press. Se puede obtener más información en la página web: <http://www1.oup.co.uk/cite/oup/elt/software/wsmith>.

Descripción: se trata de un programa para Windows que contiene seis herramientas que permiten generar listas de frecuencias y de palabras en orden alfabético, así como buscar las colocaciones de distintas palabras en el texto, entre otras funciones. Con WordSmith Tools se pueden alinear dos textos paralelos al nivel de la oración, permitiendo compararlos, calcular estadísticas de las apariciones o analizar las diferencias en los dos textos.

* **MONOCON**

Tipo de herramienta: Programa de concordancias.

Disponibilidad: No es gratuito, pero se puede descargar una versión de prueba desde la página web: <ftp://ftp.nol.net/pub/users/athel/Win/monocon.a>

Descripción: MonoCon es un programa de concordancias para Windows y Macintosh que está diseñado para la investigación basada en corpus. Está dirigido especialmente a docentes y aprendientes, a los que permite buscar palabras, sintagmas o partes de palabras en uno o más textos. Se puede llevar a cabo también la búsqueda de

frecuencias o colocaciones y todas las opciones se pueden realizar sobre cualquier lengua.

* **OXFORD CONCORDANCE PROGRAM**

Tipo de herramienta: Analizador textual.

Disponibilidad: No es gratuito. El distribuidor es Oxford University Press.

Descripción: Oxford Concordance Programa es un programa de análisis textual que permite que se realicen listados de palabras, índices y se busquen concordancias. También proporciona estadísticas sobre sus resultados. Los textos que procesa pueden estar en cualquier lengua y alfabeto.

4. CREACIÓN DE UN CORPUS DE ESP A PARTIR DE INTERNET

En esta sección se muestra la creación de un pequeño corpus de turismo que pueda ser utilizado como apoyo al proceso de aprendizaje de esta variedad de ESP.

4.1. Proceso de creación del corpus de turismo

A priori es necesario establecer los criterios necesarios de los textos que van a formar el corpus:

- Su formato debía ser electrónico para facilitar su almacenamiento y su procesamiento por herramientas de análisis y a ser posible texto plano sin gráficos ni tablas, para evitar «ruidos» y una laboriosa postedición.
- debían proceder de folletos turísticos, para garantizar una alta fiabilidad de los contenidos y su representatividad.
- debían haber sido creados para un medio escrito para garantizar la homogeneidad de los textos seleccionados.
- Los folletos debían estar destinados a un público general, no necesariamente especializado, para no encontrar trabas de comprensión.
- El objetivo de estos textos debía ser meramente informativo, no interactivo, dado que interesa el carácter científico de los textos.

Se seleccionaron los textos procedentes de los folletos turísticos debido principalmente a que no son excesivamente especializados y, por tanto, son comprensibles para los usuarios no expertos, como pueden ser los alumnos de turismo. Además, es fácil acceder a ellos y muy económico (pues no intervienen derechos de autor al ser utilizados con fines educativos y no comerciales).

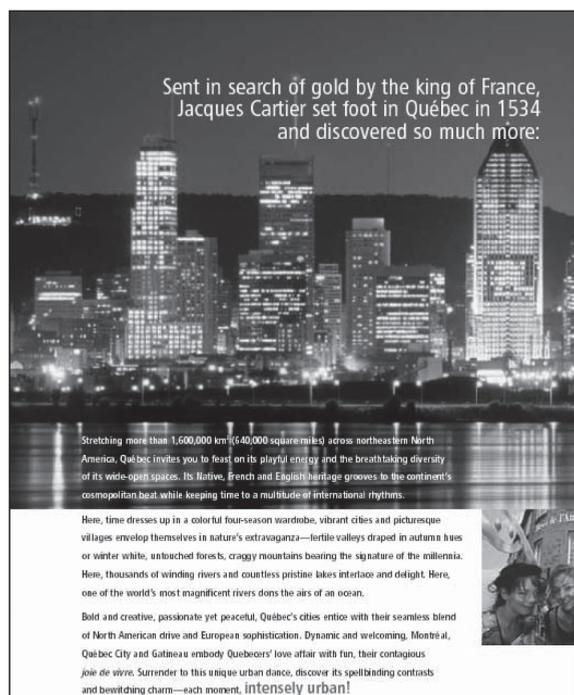


Figura 3. Ejemplo de un folleto turístico utilizado en el corpus.

A partir de estos criterios, se realizó una búsqueda en Internet de textos de este tipo para crear el corpus. Se seleccionó Internet por la gran cantidad de textos disponibles y porque, actualmente, es una de las fuentes de información más utilizada por las personas que quieren viajar. Asimismo, presentan la ventaja de que los textos ya se encontraban en formato electrónico. Se buscaron folletos turísticos publicados en Internet por ayuntamientos, gobiernos locales y nacionales, empresas turísticas, etc., pues son idóneos debido principalmente a que los textos los escriben profesionales, no sufriendo manipulación alguna anterior ni posterior. También resultaba adecuado el formato en el que se encontraban disponibles (en formato pdf), lo que tras una labor de edición facilitaba su procesamiento y ahorra tiempo en su preparación.

Se recopilaron folletos turísticos procedentes de 24 oficinas de turismo diferentes. Todos ellos tienen orígenes distintos, pues pertenecen a lugares tan diversos como Tanzania, Londres, Helsinki, Canadá o Australia. Para que el corpus sea más válido y representativo, ha sido tomado de fuentes oficiales, pero debido a las diferentes fuentes es más heterogéneo y la autoría es ilimitada. Además se han tomado de varias direcciones de Internet, no de una sola. Los múltiples orígenes de estos textos evitan una excesiva homogeneización que ponga en duda la distancia entre la regularidad propia de los sublenguajes y las reiteraciones de un idiolecto.

Definido el formato de los textos que formaban el corpus y que serían almacenados, se procedió a examinar su contenido.

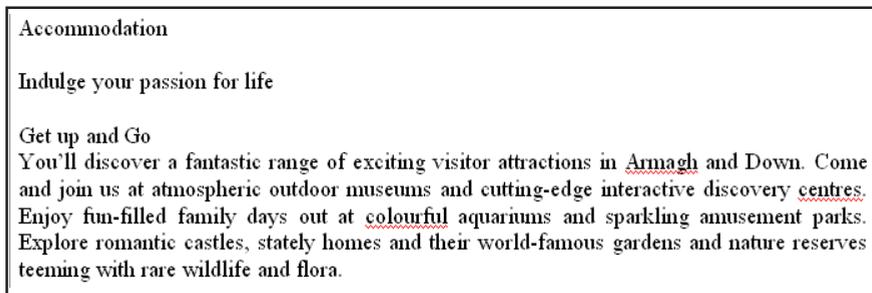


Figura 4. Extracto del corpus TOB.

En la figura 5 se muestra un extracto del corpus TOB, correspondiente al texto 3, un folleto turístico de *Armagh and Down*. En ella se puede ver como se describen algunas de los recursos que ofrece este lugar para atraer a los visitantes, tales como museos, acuarios o parques, además de castillos o jardines.

El corpus se ha denominado TOB (*Corpus of Tourist Brochures*) y se debe caracterizar como:

- textual
- con fines específicos (turismo)
- monolingüe (inglés)
- completo (los textos no están cortados)
- sin anotar

Como ya se indicó anteriormente, el tamaño de un corpus varía también en función de las necesidades de la investigación. Por eso el TOB es un corpus pequeño, pero, en principio, suficiente para el fin con el que fue creado: asistir en la elaboración y preparación de materiales didácticos para la clase de inglés de turismo. Además, se ajusta sobradamente a los mínimos establecidos por Lehrberger y Bourbeau.

Tabla 1. Tamaño de los corpus.

CORPUS	Nº DE PALABRAS	Nº DE TEXTOS
TOB	203.339	24

A propósito del tamaño, éste está directamente relacionado con la representatividad. Kittredge (2003:436) señala que basar la descripción de un corpus en una sola fuente puede ser suficiente para «*prototyping, or even building, a simple processing application to serve only that source, but does not normally give a good perspective of the whole sublanguage, as used by a wider community*». Por este motivo, los textos seleccionados proceden de diferentes fuentes. Además, la elaboración del TOB se realizó de un modo riguroso y objetivo.

El criterio de selección estuvo claro desde un principio: se pretendía recopilar una muestra de folletos de turismo lo suficientemente amplia y representativa y una vez seleccionados los textos, no se ha realizado ninguna manipulación en los datos recogidos.

En la siguiente figura se muestran los diferentes textos que componen el corpus TOB:

Tabla 2. Tamaño de los textos de TOB.

Nº PALABRAS		Nº PALABRAS	
Texto 1	26180	Texto 13	806
Texto 2	2498	Texto 14	2168
Texto 3	34105	Texto 15	2178
Texto 4	652	Texto 16	7929
Texto 5	2532	Texto 17	1765
Texto 6	5033	Texto 18	936
Texto 7	1944	Texto 19	354
Texto 8	1544	Texto 20	19132
Texto 9	6573	Texto 21	46800
Texto 10	536	Texto 22	799
Texto 11	2758	Texto 23	7813
Texto 12	20310	Texto 24	7994
		TOTAL 203.433	

La diferencia en el número de palabras de cada textos es válida, porque así no se ha tenido que cortar los textos para que sean iguales de tamaño, lo que garantiza que trabajar con material real.

4.2. Búsqueda de concordancias: book.

A través de un programa de concordancias MonoCon Pro, se ha realizado en el TOB una búsqueda simple de la palabra *book*. Los resultados se muestran en la figura 5.

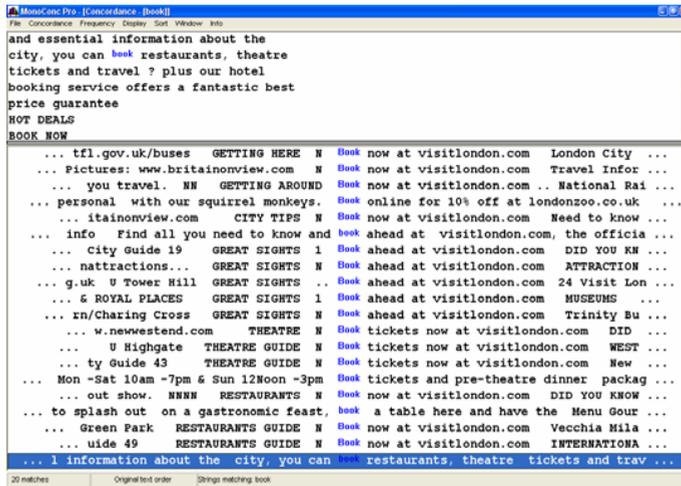


Figura 5. Concordancias de book en el TOB.

Así, se puede observar las ocurrencias de esta palabra listadas en el corpus y el contexto en el que aparece. Además, para orientar mejor los resultados, se puede obtener las palabras con las que más frecuentemente coocurre:

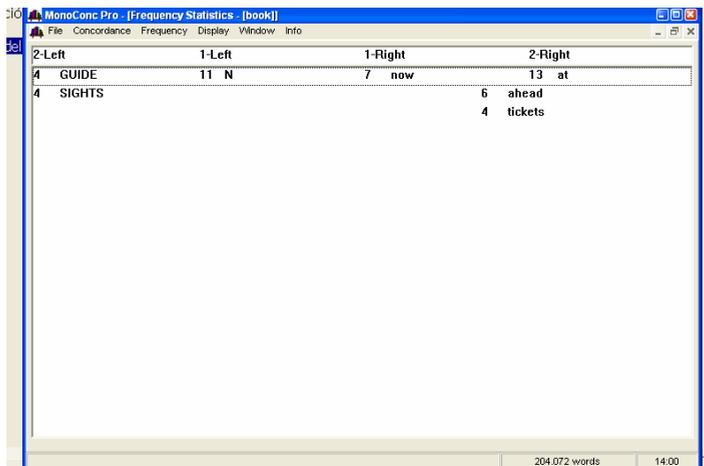
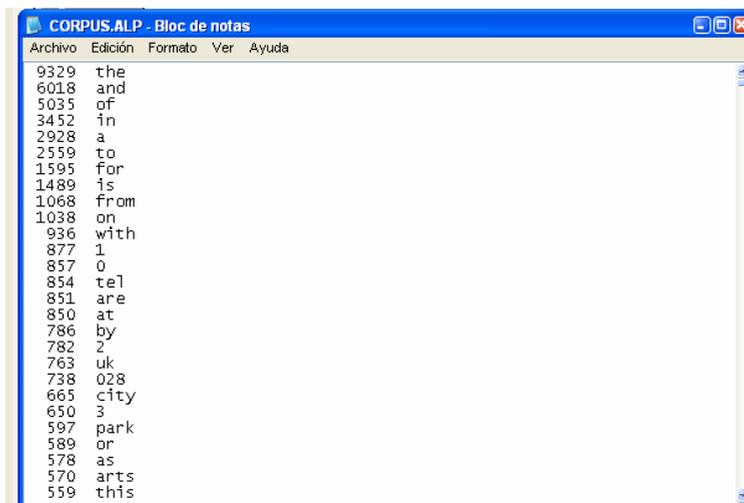


Figura 6. Palabras que coocurren con book en el TOB.



Frecuencia	Palabra
9329	the
6018	and
5035	of
3452	in
2928	a
2559	to
1595	for
1489	is
1068	from
1038	on
936	with
877	1
857	0
854	tel
851	are
850	at
786	by
782	2
763	uk
738	028
665	city
650	3
597	park
589	or
578	as
570	arts
559	this

Figura 8. Lista de frecuencia de las palabras del TOB obtenida con TACT.

Esta es solo una muestra de la información que los datos pueden proporcionar a los profesores y estudiantes de lenguas, pero las herramientas de análisis presentan opciones más complejas que permiten hacer búsquedas refinadas y obtener información más precisa.

5. CONCLUSIÓN

En este artículo se han mostrado algunas de las aportaciones que los corpus pueden hacer a la enseñanza de lenguas, especialmente por su condición de textos reales para su posible integración en el proceso de aprendizaje. Su validez queda demostrada desde dos enfoques diferentes: el del los aprendientes y el de los profesores, que son los actores del proceso de aprendizaje. El uso de los corpus permite a los aprendientes llevar a cabo un aprendizaje autónomo, por lo que puede reforzar el aprendizaje en un entorno de autoaprendizaje o de enseñanza a distancia, pero también se ha visto que puede servir de apoyo en el aula. Asimismo, los profesores pueden encontrar en los corpus una herramienta de gran utilidad para la creación de materiales o para mostrar ejemplos reales. Aunque como bien se ha indicado, por el momento, la integración de los corpus en la enseñanza de lenguas es todavía limitada, sobre todo en niveles no universitarios.

Para conseguir esta integración de los corpus en el proceso de aprendizaje, se ha dado el primer paso y se han mostrado algunas de las opciones disponibles para demostrar que los corpus no son tan inaccesibles. Se han enumerado algunos de los corpus disponibles más importantes y se han explicado algunas de las opciones que ofrece Internet como corpus o como fuente de información para crear un corpus. El artículo se cierra con un ejemplo del proceso de creación de un corpus de ESP para la utilización por parte de aprendientes y de profesores.

6. BIBLIOGRAFÍA

- Aston, G. (1998). «What corpora for ESP?». En M. Pavesi y G. Bernini (ed.), *L'apprendimento linguistico all'università: le lingue speciali*. Roma: Bulzoni., 205-226.
- Barlow, M. (1992). «Using Concordance Software en Language Teaching and Research», en W. Shinjo (eds.) *Proceedings of the Second International Conference on Foreign Language Education and Technology*. Kasugai, Japan: LLAJ & IALL.
- Barlow, M. (1996). «Corpora for Theory and Practice», en *International Journal of Corpus Linguistics*, Vol. 1 1: 1-38.
- Barnbrook, G. (1996). *Language and Computers: a Practical Introduction to the Computer Analysis of Language*. Edimburgo: Edimburgh University Press.
- Braun, S. (2005). «From pedagogically relevant corpora to authentic language learning contents», en *ReCALL* 17, 1: 47-64.
- Chambers, A. (2005). «Integrating Corpus Consultation in Language Studies Language Learning & Technology», Volume 9, 2:111-125.
- Fligelstone, S. (1993). «Some Reflections on Teaching, from a Corpus Linguistics Perspective», en *ICAME Journal* 17: 97-109.
- Flowerdew, J. (1996). Concordancing in language learning. In M. Pennington (Ed.), *The power of call*. Houston, TX: Athelstan., 97-113
- Gabrielatos, C. (2005). Corpora and language teaching: Just a fling, or wedding bells? *TESL-EJ*, 8(4), Disponible en la página web: <http://www.tesl-ej.org/ej32/a1.html>. [Consultado 10 de marzo 2007].
- Godwin-Jones, R. (2001). «Emerging Technology: Tools and Trends in Corpora Use for Teaching and Learning». *Language Learning –Technology*. 5:3. Págs. 7-12
- Granger, S. (2003). «The international corpus of learner English: a new resource for foreign language learning and teaching and second language acquisition research», en *TESOL Quarterly* 37,3: 538-545.
- Hunston, S y Francis, G. (1999). *Pattern Grammar: A corpus-driven approach to the lexical grammar of English*. Amsterdam/Philadelphia: John Benjamins.
- Hunston, S. (2003). *Corpora in Applied Linguistics*. Cambridge: Cambridge University Press.
- Johns, T. (1991). «From printout to handout: Grammar and vocabulary learning in the context of datadriven learning», en *ELR Journal* 4:27-45.
- Kennedy, G. (1992). «Preferred Ways Of Putting Things», en J. Svartvik (ed.) *Directions in Corpus Linguistics: Proceedings of Nobel Symposium 82*. Berlín: Mouton de Gruyter, 335-373.
- Kennedy, G. (1998). *An Introduction to Corpus Linguistics*. Essex: Addison Wesley Longman.
- Kittredge, R. (2003). «Sublanguages and Controlled Languages», en R. Mitkov (ed.). *The Oxford Handbook of Computational Linguistics*. Oxford: Oxford University Press, 430-448.
- Leech, G. (1991). «The state of the Art in Corpus Linguistics», en K. Aijmer y B. Altenberg (eds.). *English Corpus Linguistics. Studies in Honour of Jan Svartvik*. Londres: Longman, 8-29.

- Lehrberger, J. y Bourbeau, L. (1988). *Machine Translation. Linguistic characteristics of MT systems and general methodology of evaluation. Linguisticae Investogationes: Supplementa* 15. Amsterdam: John Benjamins.
- López Sanjuán, V. (2003). Tratamiento e identificación de sublenguajes a través de corpus electrónicos en lengua inglesa y de herramientas estándar de análisis con fines computacionales. Sin publicar.
- López Sanjuán, V. (2006). *La identificación de sublenguajes con fines computacionales basada en corpus electrónicos en lengua inglesa y en la aplicación de herramientas de análisis y su comparación a partir del Modelo Multidimensional de Biber*. Tesis doctoral. Sin publicar.
- Mcenery, T y Wilson, A. (1997). «Teaching and Language Corpora TALC» en *ReCALL* 9,7:5-14.
- Meyer, C. (2006). «Corpus Linguistics, the World Wide Web, and English Language Teaching», en *Ibérica* 12: 9-21.
- Mindt, D. (1997). «Corpora and the teaching of English in Germany», en A. Wichmann, S. Fligelstone, T. McEnery, y G. Knowles (eds.), *Teaching and language corpora*. New York: Longman, 40-50.
- Mukherjee, J. (2004). «Bridging the Gap Between Applied Corpus Linguistics and the Reality of English Language Teaching in Germany», en *Applied Corpus Linguistics: A Multidimensional Perspective*, U. Connor y T. Upton (ed.). Amsterdam: Rodopi, 239-250.
- Oakes, M. (1998). *Statistics for Corpus Linguistics*. Edimburgo: Edinburgh University Press.
- Pérez Guerra, J. (1998). *Análisis computarizado de textos. Una introducción a TACT*. Vigo: Universidade de Vigo.
- Salzmann, A. y Mills, D. (2000). *LinguaCenter Grammar Safari*. Disponible en la página web: <http://deil.lang.uiuc.edu/resources/web/pages/grammarsafari.html>. [Consultado el 20 de julio de 2000].
- Scott, M. & Tribble, C. (2006). *Textual Patterns: keyword and corpus analysis in language education*, Amsterdam: Benjamins.
- Scott, M., (2004). *WordSmith Tools* version 4, Oxford: Oxford University Press.
- St. John, E. (2001). A Case for Using a Parallel Corpus and Concordancer for Beginners of a Foreign Language. *Language Learning & Technology* 5:3, 185-203.
- Tribble, C. (1997). «Improvising corpora for ELT: Quick-and-dirty ways of developing corpora for Language teaching», en B. Lewandowska-Tomaszczyk y P. J. Melia (eds.), *Practical applications in language corpora*. Lodz, Poland: Lodz University Press, 106-117.
- Tribble, C. (2000). «Practical uses for language corpora in ELT», en P. Brett y G. Motteram (eds.), *A Special interest in computers: Learning and teaching with information and communications technologies*, Whitstable, Kent, UK: IATEFL, 31-41.