

# LOS CORPUS LINGÜÍSTICOS INFORMATIZADOS: TIPOLOGÍA Y APLICACIONES EN EL CAMPO DE LA TRADUCCIÓN

*Noelia Ramón García*  
*Universidad de León*

## 1. INTRODUCCIÓN

Los avances informáticos que se han venido produciendo en las últimas décadas, y especialmente en los últimos años, han desarrollado materiales y herramientas de diversos tipos que son de gran utilidad para los investigadores en todos los campos lingüísticos, entre otros, el de los estudios de contraste de lenguas, y el de la traducción. Las herramientas informáticas desarrolladas para uso específico de traductores son muchas y muy variadas, como por ejemplo las múltiples opciones de los procesadores de textos más potentes, los diccionarios electrónicos (en CD-ROM o en la red Internet), las bases de datos terminológicas, los corpus lingüísticos, y las más modernas memorias de traducción.

En este trabajo trataremos de una de estas herramientas, los corpus lingüísticos informatizados, en función de su utilidad en el campo de la traducción, ya sea en su aspecto teórico, en la aplicación práctica, o en la didáctica de la traducción.

## 2. ¿QUÉ SON LOS CORPUS LINGÜÍSTICOS INFORMATIZADOS?

Básicamente, un corpus es un conjunto de textos, procedentes de fuentes diversas, de los que se puede extraer todo tipo de información lingüística. Según la definición de Sinclair, un corpus es "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." (1995: 17). El tipo de información lingüística que se puede extraer de un corpus es muy variado. El análisis descriptivo de las lenguas, la lexicografía, la terminología, la lingüística contrastiva y la traducción son sólo algunas de las áreas que se pueden beneficiar del uso de corpus.

Bloomfield y sus discípulos fueron los pioneros en el uso de amplios corpus textuales en la investigación lingüística en los años 30. En esta época, los corpus estaban aún almacenados en soporte de papel. Sin embargo, hoy en día todo corpus es por definición un corpus informatizado, ya sea en soporte transportable (CD-ROM o disquete) o accesible a través de redes informáticas (Internet, Telnet). Esto facilita enormemente la tarea del investigador y le ahorra tiempo y esfuerzo.

## 3. TIPOS DE CORPUS

Dependiendo de los criterios con que se hayan compilado, y de los fines para los que se hayan diseñado, existen numerosos tipos distintos de corpus. Podemos distinguir entre las siguientes categorías principales:

1. Sincrónicos / diacrónicos: son corpus sincrónicos aquellos que recogen textos de un estadio histórico concreto de la lengua, y diacrónicos aquellos que recogen textos de diversas etapas en la evolución de la lengua.
2. Escritos / orales: existen corpus que incluyen únicamente textos escritos, corpus de textos orales transcritos, y corpus que combinan textos de las dos procedencias en diferentes porcentajes.
3. Textos completos / textos fragmentarios: algunos corpus sólo incluyen textos completos independientemente de su longitud, mientras que otros seleccionan únicamente fragmentos más o menos largos de una mayor cantidad de textos diferentes.

4. Generales / especializados: son corpus generales de la lengua los que aspiran a ser representativos de toda la lengua, y especializados aquellos que solamente recogen textos de un área temática concreta para extraer resultados aplicables exclusivamente a ese campo. Existen, por ejemplo, corpus de la industria petroquímica, de economía, de prensa, de política, etc.
5. Cerrados / abiertos: son corpus cerrados aquellos a los que ya no se continúan añadiendo nuevos textos, y abiertos aquellos que se renuevan continuamente.
6. Según variantes regionales: existen corpus que incluyen únicamente textos pertenecientes a una de las posibles variantes geográficas de la lengua (por ejemplo, inglés británico exclusivamente), y corpus que combinan textos de todas o la mayoría de las variantes existentes con el ánimo de ser más representativos.
7. Etiquetados / no etiquetados: algunos corpus tienen etiquetas en cada uno de sus elementos. Estas etiquetas pueden contener información de muchos tipos, normalmente relacionada con la categoría gramatical de la palabra, o su estatus semántico. Puesto que esto exige un gran esfuerzo, sólo los grandes corpus tienen sus palabras etiquetadas.
8. Monolingües / bilingües o multilingües: un corpus monolingüe contiene textos en una sola lengua. Un corpus bilingüe incluye textos en dos lenguas distintas, y, por último, también existen corpus con textos en tres o incluso más lenguas.

Esta primera tipología es básica y sencilla. Sin embargo, el avance imparable de la tecnología y de la informática ha permitido la creación de otros tipos de corpus mucho más complejos por medio de la combinación de varias de estas posibilidades. Estas diferentes combinaciones nacieron gracias al especial interés innovador de ciertos investigadores por crear nuevas herramientas de aplicación en el campo de la traducción. Se trata de variedades de corpus bilingües o multilingües, y de corpus de traducciones. Baker (1995) presenta una clasificación de acuerdo con el origen de los textos en estos corpus con textos en varias lenguas, que son los más útiles en el campo de la traducción. Seguiremos aquí la terminología empleada por esta autora, aunque no hay consenso entre los diferentes investigadores en este aspecto debido al desarrollo reciente de este campo.

- **Corpus paralelos:** un corpus paralelo consta de textos originales en una lengua y su traducción a la otra. Estos corpus también se denominan corpus de traducciones. "A parallel corpus consists of original, source language-texts in language A and their translated versions in language B" (Baker 1995: 230).
- **Corpus bilingües o multilingües:** son aquellos corpus que incluyen, por un lado, textos producidos originalmente en una lengua, y por otro lado, textos producidos también originalmente en otra lengua distinta (corpus bilingüe) o en varias otras lenguas distintas (corpus multilingüe). No se trata de traducciones, sino de material lingüístico auténtico recopilado con criterios similares en todas las lenguas que componen el corpus. A pesar de que no se trata de traducciones, los textos deben tratar sobre la misma temática, deben poseer una extensión parecida, y deben haber sido producidos en una fecha similar, para alcanzar un nivel de comparabilidad aceptable.
- **Corpus comparables:** "Comparable corpora consist of two separate collections of texts in the same language: one corpus consists of original texts in the language in question and the other consists of translations in that language from a given source language or languages." (Baker 1995: 234).

Los tipos de corpus mencionados aquí son los más conocidos y los más utilizados en la investigación en el campo de la traducción. No obstante, las

posibilidades tecnológicas actuales hacen posible que cualquier idea nueva pueda llevarse a cabo por medio de la combinación de alguno de estos tipos de corpus ya existentes, o incluso que puedan crearse nuevas categorías de corpus próximamente.

Un buen ejemplo de estos avances lo encontramos en dos proyectos de vanguardia en este campo. Varias universidades escandinavas (Suecia, Noruega y Finlandia) han desarrollado sistemas de combinación de un corpus paralelo, por ejemplo en Suecia, inglés-sueco, con otro corpus paralelo sueco-inglés, con lo que esta combinación permite también acceder a dos corpus comparables, uno de originales y traducciones en inglés, y otro de originales y traducciones en sueco. Hay más información en la correspondiente página de Internet de la Universidad de Lund: <http://www.englund.lu.se/research/corpus.html>.

#### **4. HERRAMIENTAS INFORMÁTICAS APLICADAS A CORPUS.**

La mayoría de las herramientas informáticas desarrolladas para ayudar a los traductores en su labor de trasvase lingüístico, especialmente los diccionarios electrónicos, o las bases de datos terminológicas, proporcionan una información muy precisa. A esta información se puede acceder de manera fácil y rápida, y está lista para ser empleada por el usuario. En cambio, los corpus son herramientas lingüísticas más complejas y requieren un manejo más especializado para obtener la información deseada. "Language corpora are more of a raw data type, and require a more sophisticated user if they are to be of any assistance. They offer information about language indirectly, concealed in the contextual entourage." (Lewandowska-Tomaszczyk 1997: 327).

Un corpus por sí sólo no es más que un conjunto de textos en soporte electrónico, y no tiene ninguna utilidad si no le aplicamos algún tipo de sistema informático de análisis. A continuación enumeramos algunas de las posibilidades de búsqueda más sencillas que se pueden aplicar a un corpus para extraer de él información lingüística. Los grandes corpus disponibles a través de algún medio electrónico suelen disponer ya de todas estas herramientas, pero también se pueden conseguir por separado para aplicarlas en cualquier texto propio o corpus recopilado por nosotros mismos.

##### **4.1. Sistema de obtención de concordancias.**

La herramienta más frecuente a la hora de extraer datos de un corpus lingüístico es la que proporciona concordancias, especialmente del tipo KWIC (Key Word In Context). Una concordancia es "a list of all the occurrences of a specified keyword or expression in the corpus, set in the middle of one line of context each." (Baker 1995: 226). La mayoría de los programas también permiten ver un contexto mayor que el de la línea, así como decidir cuántas palabras y qué palabras queremos ver a la izquierda o a la derecha de la palabra clave.

##### **4.2. Sistema de obtención de índices de frecuencias.**

Otro de los aspectos más útiles es la posibilidad de obtener porcentajes de aparición de ciertas palabras. Esta herramienta permite así determinar qué palabras son más frecuentes que otras en determinados corpus o subcorpus, y establecer estudios comparativos de frecuencia, así como ayudar a la hora de elegir un término u otro en ciertas ocasiones.

### **4.3. Sistema de obtención de colocaciones.**

Este tipo de herramienta permite conocer cuáles son las palabras junto a las que aparece más frecuentemente el elemento elegido para la búsqueda. Se pueden observar así las tendencias de colocación de esa palabra y utilizar la información para obtener una mayor exactitud y naturalidad en la lengua meta de cualquier traducción.

### **4.4. Sistema de alineación.**

Este sistema sirve para tener acceso en una misma pantalla de ordenador a un texto y su traducción al mismo tiempo. Se elige un criterio de alineación, normalmente la oración (el punto ortográfico) o el párrafo, y el sistema coloca la traducción junto a su original. Posteriormente se pueden aplicar búsquedas de concordancias en una de las dos lenguas, lo que nos permitirá ver todas las diferentes traducciones elegidas para dicho término. En general, es difícil que se consiga un alineamiento perfecto sin intervención humana, y una corrección o revisión posterior.

### **4.5. Sistema de reconocimiento de categorías gramaticales.**

En los grandes corpus, todas las palabras están etiquetadas gramaticalmente. Este hecho sirve para buscar ciertas combinaciones de una palabra concreta con diferentes categorías gramaticales, con lo que su uso redundará en una mayor exactitud de las traducciones obtenidas.

## **5. APLICACIONES DE CORPUS EN TRADUCCIÓN**

"Among many applications of machine readable corpora [...] are research themes such as contrastive analysis and error analysis, and applied topics such as human and machine (aided) translation, as well as educational applications connected with the training of translators and interpreters." (Lewandowska-Tomaszczyk 1997: 331).

Una vez definidos los tipos de corpus más frecuentes, y las herramientas más importantes que se pueden aplicar a los corpus para extraer de ellos información lingüística, nos detendremos en aquellos tipos de corpus que son de mayor utilidad en el campo de la traducción.

### **5.1. Corpus generales monolingües.**

Este tipo de corpus es el más conocido y de uso más habitual. Los grandes corpus monolingües abarcan normalmente varios cientos de millones de palabras de textos de todos los géneros, registros y tipos textuales, puesto que pretenden ser representativos de la lengua en general. Su mayor ventaja estriba en que incluyen únicamente textos producidos de manera original y espontánea por los hablantes nativos de una lengua, por lo que los datos que contienen están avalados por la fiabilidad que representa el uso real de la lengua. Dado que un traductor necesita información de las dos lenguas con las que está trabajando, podrá utilizar dos corpus monolingües, uno en cada una de las lenguas. Un corpus monolingüe en la lengua origen de la traducción tiene básicamente usos de exactitud terminológica, así como de frecuencia de uso, aunque el corpus de la lengua meta será probablemente de mayor utilidad a un traductor, ya que es la lengua en la que tiene que producir su traducción.

La información que se puede extraer de estos corpus y que es aplicable en traducción es muy variada, aunque destacan en especial dos aspectos: la obtención de concordancias, y la obtención de índices de frecuencia de aparición de ciertas palabras o grupos de palabras. Las concordancias son listas de todas las apariciones de una determinada palabra rodeada de su contexto más inmediato. Esto supone el acceso a un

amplio abanico de todas las posibles acepciones de cualquier palabra y de numerosos ejemplos de cada una de ellas.

Los grandes corpus monolingües son especialmente adecuados para extraer índices de frecuencias de apariciones de ciertas palabras o incluso estructuras sintácticas, lo que facilita al traductor la elección del término más empleado en su versión en la lengua meta. El manejo de estos corpus es una ayuda a la hora de utilizar un lenguaje más fluido y natural en la lengua meta, sin influencias de la lengua origen. Además, los índices de frecuencia han demostrado ser sumamente útiles también en la práctica y en la didáctica de la traducción, como ha demostrado Ebeling (1998) en un experimento llevado a cabo con estudiantes de traducción, un grupo de los cuales tenían acceso a información de corpus, mientras que otro grupo carecía de ella. Los resultados fueron notablemente mejores en el caso de los alumnos con acceso al corpus gracias a la información de frecuencia estadística: cuando dudaban entre dos o más palabras elegían aquélla que tuviera un porcentaje de aparición mayor en el corpus, sin dejar de tener en cuenta también las posibilidades combinatorias.

Como ejemplo, citaremos dos grandes corpus monolingües, uno en inglés y otro en español. El llamado Cobuild/Bank of English es el mayor corpus general de la lengua inglesa existente en la actualidad. Contiene aproximadamente 400 millones de palabras, e incluye textos tanto de inglés británico como de inglés americano y australiano. Hay información disponible y una demostración del funcionamiento de este corpus en la siguiente dirección de Internet: [http://titania.cobuild.collins.co.uk/boe\\_info.html](http://titania.cobuild.collins.co.uk/boe_info.html). El mayor corpus general del español es el CREA (Corpus de Referencia del Español Actual), con unos 125 millones de palabras, tanto de la variante peninsular, como de la variante hispanoamericana, repartidas a un cincuenta por ciento. El acceso es a través de la página web de la Real Academia Española, que ha desarrollado el proyecto de este corpus: <http://www.rae.es/NIVEL2/recursos.htm>.

## **5.2. Corpus paralelos o de traducciones.**

Consideramos que los corpus paralelos son los que, en principio, más aplicaciones directas pueden tener en la traducción, especialmente en su vertiente práctica y en la enseñanza de la traducción, ya que constan de textos originales en una lengua y su traducción en otra lengua.

La forma de presentación más frecuente de un corpus paralelo es la alineación. Existen ciertos programas informáticos que permiten dividir una pantalla de ordenador a la mitad y observar al mismo tiempo el original (en la parte izquierda) y su traducción (en la parte derecha). Los dos textos deben estar alineados según un criterio de formato textual, normalmente por oraciones o párrafos. Este proceso requiere a menudo revisiones manuales, ya que no siempre coinciden estos criterios formales en los dos textos: una única oración puede estar escindida en dos en la traducción, por ejemplo.

Una vez que el original y su traducción están alineados, cualquier simple programa de búsqueda léxica, similar al de cualquier procesador de textos, será capaz de encontrar todas y cada una de las apariciones de una determinada palabra en el original, localizarla en la pantalla, y permitirnos así ver cuáles han sido las distintas opciones utilizadas por el traductor para traducir esa palabra. Este proceso también puede realizarse a la inversa. Buscando una determinada palabra en el texto traducido, se podrá observar todas las distintas formas que dieron lugar a esa misma traducción.

Las aplicaciones en la práctica de la traducción son obvias, ya que proporcionan de manera inmediata numerosos ejemplos posibles de traducción, de la misma forma que podría hacerlo un diccionario, pero con mucho más acierto, con un contexto completo rodeando el término, y con la ventaja de estar tratando con traducciones

auténticas, lo que refrenda su validez y su profesionalidad, si se han seguido unos mínimos criterios de calidad a la hora de la recopilación del corpus. Si el corpus paralelo es de textos científico-técnicos, será una fuente muy valiosa de información terminológica, aunque, en general, son difíciles de conseguir o de recopilar.

Un ejemplo muy conocido de este tipo de corpus alineados, disponible en Internet gratuitamente se encuentra bajo la siguiente dirección: <http://www-rali.iro.umontreal.ca/TransSearch/TS-simple-uen.cgi>. Se trata de las sesiones del parlamento canadiense, con el texto completo (1986-1993) alineado en inglés y francés. Además, el sistema TransSearch permite la búsqueda inmediata de cualquier término.

Los corpus de traducciones, una vez alineados, tienen múltiples aplicaciones en el campo de la traducción, no sólo en su vertiente práctica, sino también en su vertiente teórica, ya que permiten estudiar las elecciones lingüísticas de los traductores. Así se puede establecer las tendencias generales de desviación en la lengua meta provocadas por determinadas estructuras en la lengua origen, así como regularidades que llevan a considerar que existen ciertas normas traductorales imperantes en determinadas épocas.

Aparte de sus aplicaciones en la práctica de la traducción y en el análisis de comportamientos traductores, se han señalado también la traducción automática y la lingüística contrastiva como dos áreas de aplicación importantes de los corpus paralelos (Bravo Gozalo & Fernández Nistal 1998: 225). En cuanto a la traducción automática, la posibilidad de tener un texto original alineado con su traducción es un gran paso adelante para hacer que un ordenador pueda reconocer equivalentes léxicos y realizar una cierta labor traductora, que después tendrá que ser revisada por un traductor humano. Y en lo que se refiere a la lingüística contrastiva, el estudio descriptivo de las estructuras que aparecen en un corpus paralelo lleva a conclusiones con aplicaciones posibles en la práctica de la traducción.

### **5.3. Corpus bilingües.**

Nos centramos ahora en los corpus bilingües, compuestos por dos subcorpus, cada uno en una lengua distinta, que contienen textos escritos originalmente en cada una de las lenguas sobre una temática similar, de una extensión parecida, y dentro de un mismo marco temporal. Los dos subcorpus también deben haber sido recopilados utilizando unos criterios similares. Este tipo de corpus también puede incluir más de dos lenguas, aunque en este caso nos centraremos sólo en los bilingües, ya que los resultados se podrán extrapolar automáticamente a los multilingües.

Si tenemos en cuenta que estos corpus contienen únicamente textos originales en ambas lenguas de un campo concreto del saber, podemos determinar que su mayor utilidad estará en el caso de traducciones de tipo técnico o científico. La principal razón es la incalculable fuente de información terminológica que representan. La traducción de este tipo de textos es altamente especializada, y el aspecto lingüístico más predominante es el de la exactitud en el trasvase de términos técnicos o científicos concretos, por lo que tener a disposición material original en ambas lenguas es de gran utilidad.

Se trata, en realidad, de dos corpus monolingües de temática específica, que están unidos de algún modo, ya sea alineados, o compartiendo una misma base de datos. Algunos corpus monolingües de temática específica se pueden comprar, y también se pueden recopilar corpus de este tipo incluyendo aquellos textos que más nos interesen.

#### **5.4. Corpus comparables.**

Los corpus comparables, como ya hemos mencionado anteriormente, incluyen dos subcorpus en una única lengua: un subcorpus consta de textos originales escritos en esa lengua, mientras que el otro consta de textos traducidos a esa lengua de una o varias lenguas distintas. Los textos de estos dos subcorpus "should cover a similar domain, variety of language and time span, and be of comparable length." (Baker 1995: 234). En consecuencia, los corpus comparables son especialmente útiles en el análisis de traducciones, así como en la investigación sobre las características típicas de la lengua tal y como aparece en textos traducidos.

S. Laviosa (1997) propone una tipología amplia y compleja de corpus comparables dependiendo del origen de los textos incluidos en cada una de las subdivisiones. Este tipo de estudios llevan a describir lo que se da en llamar *translationese*, "a generally pejorative term used to refer to TL usage which because of its obvious reliance on features of SL is perceived as unnatural, impenetrable or even comical" (Shuttleworth & Cowie 1997: 187). Se trata, en general, de ciertas construcciones lingüísticas o patrones de uso léxico, que aparecen en traducciones con una frecuencia distinta o en un contexto en el que no aparecen en textos escritos originalmente en esa lengua, con lo que se convierten en señales inequívocas del proceso de traducción del que provienen.

En general, los estudios de este tipo llevados a cabo hasta el momento revelan que la proporción de variedad léxica es inferior en textos traducidos, y que los textos traducidos emplean con una frecuencia muy superior a la habitual aquellas palabras que ya son comunes y frecuentes en la lengua a la que se traduce. Como resultado, estos estudios indican una tendencia homogeneizadora y simplificadora de las traducciones (Laviosa 1997).

Estos datos son muy útiles para los traductores, ya que llaman la atención sobre aquellos aspectos que, aun no siendo incorrectos, podrían expresarse de una manera mucho más natural en la lengua meta. Además, los patrones de uso léxico que aparecen en las traducciones pueden variar dependiendo de la lengua origen, con lo que estudios con textos procedentes de diversas lenguas afinarían mucho más las pautas a seguir por los traductores cuando se enfrentan a textos en una determinada lengua. De la misma manera, todos los datos resultantes de estos trabajos se pueden aplicar con gran provecho en la enseñanza de la traducción.

M. Baker está desarrollando en la Universidad de Manchester un proyecto denominado 'Translational English Corpus' (TEC), <http://ubatuba.ccl.umist.ac.uk/tec/>. Se trata de un auténtico corpus general compuesto exclusivamente por textos traducidos al inglés de una variedad de diferentes lenguas origen. Combinando este corpus con uno monolingüe con textos producidos por hablantes nativos de inglés, tendremos un gran corpus comparable.

#### **6. DESARROLLO DE NUEVOS CORPUS**

Por último, cabe destacar que la tecnología informática de hoy en día, además de permitirnos el acceso rápido a grandes corpus ya existentes, nos ofrece las herramientas necesarias para la construcción de nuestros propios corpus lingüísticos, en los que sólo incluiremos aquellos textos que pretendamos estudiar. Esta posibilidad no es nada desdeñable, ya que permite llevar a cabo ciertos análisis muy concretos, como el estudio de la traducción o traducciones de una obra determinada, análisis para el cual ésta es la mejor elección. También los traductores profesionales se pueden beneficiar de esta posibilidad y crear así sus propios corpus con textos traducidos o recopilados por ellos mismos, auténticos bancos de datos que van aumentando paulatinamente, y a los que se

les pueden aplicar todas las tecnologías de búsqueda de información descritas anteriormente.

En definitiva, la tecnología está al servicio del investigador, especialmente en el caso de los corpus lingüísticos. De hecho, son los propios usuarios y los investigadores en este campo los que proponen instrumentos con determinadas posibilidades que les serían de utilidad, y que después trasladan a los informáticos para que éstos desarrollen o diseñen los programas necesarios para llevar a cabo el proyecto deseado. "Cuanto mayor sea la demanda de corpus y de herramientas y programas informáticos capaces de proporcionar nuevos datos, mayor será la calidad y variedad de la oferta, y por lo tanto, mejores posibilidades tendremos de realizar investigaciones en ciencias como la traducción. Se trata de una relación recíproca de fomento y colaboración entre un método de análisis y una disciplina." (Labrador 1998: 257).

## 7. CONCLUSIONES

En el presente trabajo hemos tratado de presentar los aspectos más importantes de los tipos de corpus más destacados en el campo de la traducción. También hemos analizado las herramientas que se emplean con mayor frecuencia para extraer información lingüística de los corpus informatizados. Y finalmente, hemos analizado en detalle las posibilidades que ofrecen dentro del campo de la traducción aquellos corpus que están diseñados para facilitar estudios de contraste de lenguas o que pueden ser de alguna utilidad en este aspecto.

Los estudios de traducción han experimentado un profundo cambio a partir del desarrollo de las primeras herramientas informáticas aplicadas a la traducción. Todas ellas han sido y son de gran utilidad. No obstante, los corpus lingüísticos han marcado un antes y un después en las posibilidades de aplicación en la traducción, ya que la cantidad de información lingüística de primera mano que ofrecen es enorme y muy variada. Los corpus informatizados son útiles tanto en los aspectos prácticos como en los teóricos de la traducción, llegando incluso a desarrollar una rama propia: los estudios de traducción basados en corpus, un fenómeno conocido en inglés como *the corpus-based approach to translation studies*, y que ya está dando sus primeros frutos en forma de monografías sobre el tema, como, por ejemplo, la página web <http://www.sslmit.unibo.it/bertabs.htm>, que incluye una serie de abstracts sobre la relación entre corpus y traducción, o el número 43: 4 de la revista *Meta: Translators Journal*, dedicado íntegramente a esta nueva y provechosa rama de la traducción, a la que le espera un brillante futuro en los próximos años.

## BIBLIOGRAFÍA

- Baker, M. (1995), "Corpora in Translation Studies: An Overview and Some Suggestions for Future Research", en *Target*, 7: 2, 223-243.
- Biber, D.; CONRAD, S. & REPPEN, R. (1998), *Corpus Linguistics. Investigating Language Structures and Use*, Cambridge: Cambridge University Press.
- Bravo Gozalo, J.M. & Fernández Nistal, P. (1998), "La lingüística del corpus, las nuevas tecnologías de la información y los Estudios de Traducción en la década de 1990", en Fernández Nistal, P. & Bravo Gozalo, J.M. (eds.) (1998), *La traducción: orientaciones lingüísticas y culturales*, Valladolid: Servicio de Apoyo a la Enseñanza, Universidad de Valladolid, 205-257.
- Ebeling, J. (1998), "Contrastive Linguistics, Translation, and Parallel Corpora", en *Meta: Translator's Journal*, 43: 4, 602-615.



- Labrador de la Cruz, M<sup>a</sup>.B. (1998), "Los corpus y la traducción: un nuevo método para un disciplina", en Valero Garcés, C. & De la Cruz Cabanillas, I. (eds.) (1998), *Nuevas tendencias y aplicaciones de la traducción*, Alcalá de Henares: Servicio de Publicaciones de la Universidad de Alcalá de Henares, 253-258.
- Laviosa, S. (1997), "How Comparable Can 'Comparable Corpora' Be?", en *Target* 9:2, 289-319.
- Lewandowska-Tomaszczyk, B. (1997), "Electronic Dictionaries, Corpora and Translation", en Fleischmann, E.; Kutz, W. & Schmitt, P.A. (eds.) (1997), *Translationsdidaktik. Grundfragen der Übersetzungswissenschaft*, Tübingen: Gunter Narr Verlag, 327-335.
- Partington, A. (1998), *Patterns and Meanings*, Amsterdam/Philadelphia: John Benjamins Publishing Company.
- Pearson, J. (1998), *Terms in Context*, Amsterdam/Philadelphia: John Benjamins Publishing Company.
- Shuttleworth, M & Cowie, M. (1997), *Dictionary of Translation Studies*, Manchester: St. Jerome Publishing Company.
- Sinclair, J. (1991), *Corpus, Concordance, Collocation*, Oxford: Oxford University Press.
- \_\_\_\_\_ (1995), "Corpus Typology: A Framework for Classification", en Melchers, G. & Warren, B. (eds.) (1995), *Studies in Anglistics*, Stockholm: Almqvist and Wiksell International, 17-34.
- Thomas, J. & Short, M. (eds.) (1996), *Using Corpora for Language Research*, Londres: Longman.

## RECURSOS ELECTRÓNICOS

- Página de la Universidad de Essex con información completa de todos los corpus de cierta importancia que existen hoy en día, así como explicaciones de tipos de corpus: [http://clwww.essex.ac.uk/w3c/corpus\\_ling/content/introduction.html](http://clwww.essex.ac.uk/w3c/corpus_ling/content/introduction.html).
- Corpus general del inglés Cobuild / Bank of English. Información y prueba de funcionamiento: [http://titania.cobuild.collins.co.uk/boe\\_info.html](http://titania.cobuild.collins.co.uk/boe_info.html)
- Corpus de Referencia del Español Actual (CREA). Acceso gratuito a través de la página web de la Real Academia Española: <http://www.rae.es/NIVEL2/recursos.htm>
- Corpus paralelo inglés-francés TransSearch (actas del parlamento canadiense): <http://www-rali.iro.umontreal.ca/TransSearch/TS-simple-uen.cgi>
- Translational English Corpus (TEC), corpus comparable inglés de la Universidad de Manchester (UMIST): <http://ubatuba.ccl.umist.ac.uk/tec/>
- Información sobre el corpus paralelo inglés-sueco de la Universidad de Lund (Suecia): <http://www.englund.lu.se/~contrast/projdesc.htm>.
- Información sobre el corpus paralelo inglés-noruego de la Universidad de Oslo (Noruega), que sigue la misma línea que el de Lund (Suecia): <http://www.hf.uio.no/iba/prosjekt/>.
- Información sobre el corpus paralelo inglés-finés de la Universidad de Jyväskylä (Finlandia): [http://www.jyu.fi/tdk/hum/englanti/research\\_main.htm](http://www.jyu.fi/tdk/hum/englanti/research_main.htm).
- Información sobre el proyecto de construcción de un corpus inglés-español en el que cooperan la Universidad de Wolverhampton y la Universidad de Alicante: [http://www.wlv.ac.uk/~le1825/projects/english\\_spanish\\_anaphora\\_resolution.html](http://www.wlv.ac.uk/~le1825/projects/english_spanish_anaphora_resolution.html).

- Lista de una veintena de abstracts sobre "Corpus use and learning to translate" de los autores que más han investigado sobre estos aspectos de la traducción: <http://www.sslmit.unibo.it/bertabs.htm>.
- La Universidad de Bolonia (Italia) ofrece en esta página información sobre corpus y traducción: <http://www.sslmit.unibo.it/zanettin/cltransl.htm>.