TINKUY n°11
Mayo 2009

Section d'études hispaniques
Université de Montréal

TINKUY

BOLETÍN DE INVESTIGACIÓN Y DEBATE

Nº 11 – Mayo 2009

Número especial

La Enseñanza del Español como Lengua Extranjera en Quebec

Proceedings del CEDELEQ III (1-3 de mayo de 2008)

Editores del número especial

Enrique Pato
Luis Ochoa
Javier Lloro

ISSN 1913-0473

TINKUY n°11
Mayo 2009

Section d'études hispaniques
Université de Montréal

El uso de los corpus lingüísticos como herramienta pedagógica para la enseñanza y aprendizaje de ELE*

Elena F. Pitkowski y Javier Vásquez Gamarra
Université de Montréal

Introducción. Definición del concepto de *corpus*. El *Corpus Oral y Sonoro del Español Rural* (COSER). Los corpus de la Real Academia Española. *Corpus del español*. *Así hablamos*. *Jergas de Habla Hispana*. Ventajas y desventajas del uso de los corpus. Anexo

---

Resumen
En Internet encontramos un abanico muy amplio de herramientas útiles para el proceso de enseñanza-aprendizaje del español como lengua extranjera. Sin embargo, no siempre estamos al tanto de la existencia de estos instrumentos didácticos que podemos llevar a la clase. Un claro ejemplo de estos nuevos medios disponibles para la enseñanza de ELE lo constituyen los corpus lingüísticos. Este artículo ofrece un acercamiento al uso de algunos corpus informatizados que contamos hoy en día y que podemos utilizar para el estudio y la enseñanza de ELE.[3]

Résumé
Dans l'Internet nous pouvons trouver une panoplie d'outils efficaces pour le processus d'enseignement-apprentissage de l'espagnol comme langue étrangère. Par contre, nous ne sommes pas toujours au courant de l'existence de ces instruments didactiques profitables pour les cours que nous donnons. Quelques-uns de ces outils, disponibles de façon gratuite, sont les corpus linguistiques. Pour ce fait, cet article offre l'utilisation des quelques corpus informatisés dont nous disposons aujourd'hui pour l'étude et l'enseignement de l'ELE.

* Los contenidos plasmados en este trabajo son el resultado del Seminario "Instrumentos de trabajo sobre el español" que el profesor Enrique Pato impartió en el marco de la Maestría en Estudios Hispánicos de la Universidad de Montreal (otoño 2007), a quien queremos expresar nuestro agradecimiento por su entera disponibilidad y colaboración.

[3] Este artículo sintetiza los talleres titulados "Los nuevos medios disponibles en Internet para la enseñanza de ELE: los Corpus" y "El uso de corpus en el aula de ELE y como herramienta para la creación de materiales didácticos" que los autores impartieron en el Congreso CEDELEQ III.

1. Introducción

El desarrollo científico y tecnológico ha permitido el progreso de la informática conjuntamente con el acceso y la manipulación computarizada, tanto de textos escritos como de transcripciones de diálogos, con una rapidez, fiabilidad y facilidad impensables hasta hace poco tiempo. Esta revolución tecnológica de los últimos años produjo un gran avance, sobre todo en lo relacionado a la capacidad de almacenamiento de los ordenadores, dando lugar así a la creación de una herramienta innovadora: los corpus lingüísticos. Desde la década de los 60, los corpus informatizados y bases de datos textuales han contribuido significativamente en el área de la investigación lingüística. A este respecto, brinda a los investigadores la posibilidad de tener a su disposición grandes volúmenes de datos y, de este modo, estudiar la lengua integrada en el contexto discursivo, a través de ejemplos reales y precisos de uso; en contraposición al empleo de la introspección y los métodos intuitivos tradicionales como recurso para la formulación de principios lingüísticos. Si bien durante varios años el acceso a los corpus ha facilitado el trabajo de numerosas áreas de estudio del campo de la lingüística, sólo en los últimos tiempos, se ha abierto una nueva vía de uso y ha comenzado a tener auge su implementación con fines pedagógicos en el aprendizaje de lenguas extranjeras, y en la enseñanza del lenguaje en general.

El objetivo de este trabajo es dar a conocer la relevancia del empleo de los corpus lingüísticos, tanto para los estudiantes como para los profesores de lengua extranjera. Es de señalar que las aplicaciones de los corpus en la enseñanza de lenguas extranjeras ya ha sido considerado por varios investigadores (Córdoba Rodríguez 2001, San Mateo 2003, Sinclair 2004) y la didáctica de lenguas extranjeras con corpus ha sido planteado en congresos internacionales como el *TALC (Teaching and Language Corpora*).[4]

En esta ocasión, se presenta de manera abreviada los siguientes corpus: el *Corpus Oral y Sonoro del Español Rural* (COSER) de Inés Fernández-Ordóñez, el *Corpus de Referencia del Español Actual* (CREA) de la Real Academia Española, el *Corpus Diacrónico del Español* (CORDE) de la Real Academia Española, el *Corpus del español* de Mark Davies, *Así hablamos* y *Corpus de Jergas de Habla Hispana* (JHH). En el presente artículo se define, en primer lugar, la noción de corpus y se describen brevemente sus características relevantes. A continuación, de manera general, se exponen los diferentes procesadores de búsqueda utilizados en el

[4] El 8º congreso tuvo lugar en Lisboa, entre el 4 y el 8 de julio de 2008 [http://talc8.isla.pt].

taller y se especifica más en detalle las características de uso en el *Corpus del español* y en *Jergas de Habla Hispana* (JHH). Se sugieren algunas implementaciones de esta herramienta como apoyo para la creación de materiales didácticos con datos lingüísticos fiables y reales. Finalmente, se exponen ciertas ventajas y desventajas del uso de los corpus.

2. Definición del concepto de *corpus*

En general, se puede llamar *corpus* a una colección extensa de diferentes tipos de textos, orales o escritos, en formato electrónico, de varios millones de palabras que se codifican y clasifican adecuadamente. Los mismos se guardan y se procesan en medios de almacenamiento masivo y, de esta manera, permiten al usuario hacer diferentes búsquedas entre grandes cantidades de textos electrónicos.

Según el *Diccionario de la Real Academia Española* (DRAE, 22ª edición), un *corpus* es "conjunto lo más extenso y ordenado posible de datos o textos científicos, literarios, etc., que pueden servir de base a una investigación". En otras fuentes, como en el trabajo de Pérez Hernández (2002), se refiere a la definición de Leech (1992: 106): "On the face of it, a computer corpus is an unexciting phenomenon: a helluva lot of text, stored on a computer". Se hace hincapié que Leech completa la definición sobre los corpus como una colección de textos en formato magnético, recalcando que es gracias a los avances tecnológicos, a la "habilidad" de los ordenadores para buscar, recuperar, ordenar y hacer cálculos sobre cantidades masivas de texto, nos brinda la oportunidad de comprender y de explicar el contenido de estos corpus de forma que no era imaginable en la era que Leech denomina pre-computacional.

En particular, los corpus reflejan el contexto en el que se utiliza la lengua e intentan ser un modelo de la realidad lingüística, muestran el uso que sus hablantes hacen de ella. Es decir, procuran ser representativos de una lengua, o de una variedad de ella. Torruella y Llisterri (1999) señalan que para que un corpus sea representativo del funcionamiento de una lengua natural es necesario que el corpus sea neutro. Sin embargo, los autores agregan que esta neutralidad es una tendencia y no una realidad, porque "siempre dirigimos la mirada o el pensamiento hacia aquello que, consciente o inconscientemente, queremos demostrar".

Existen en la actualidad una gran variedad de corpus determinados por la finalidad u objetivo que se persigue. Podemos encontrar un corpus de determinado país, como, por ejemplo, el *Corpus del español de la*

Argentina, el *Corpus del español de Chile* o el *Corpus Histórico del Español de México* (CHEM) que representa el español de México entre los siglos XVI y XIX. Asimismo, un corpus puede registrar las producciones lingüísticas de los habitantes de una región geográfica y estar constituidos por las transcripciones de los registros de la lengua hablada, como el *Corpus Oral y Sonoro del Español Rural* (COSER).

Existen además, los corpus especializados que recogen la producción literaria de una época determinada como el Barroco, por nombrar sólo algunos tipos de corpus.[5]

De manera general, los corpus sostenidos con programas informáticos son una herramienta eficaz y rápida en la búsqueda tanto de una palabra, un conjunto o una serie de palabras en un contexto determinado. Los programas nos permiten explorar, en particular, el uso de frases y vocabulario utilizado en un área especifica, el nivel de uso, frecuencia o variación de un lema en contextos reales provenientes de diversos tipos de textos, analizar la colocación de las palabras, obtener muestras de cuestiones gramaticales, así como el uso real de una palabra o expresión en un país determinado, en la obra de un autor o en un cierto período de la historia del español, entre otros.

Por medio de los corpus, podemos enseñar la lengua a través de ejemplos reales de uso, porque esta excelente herramienta informática nos brinda el contexto cultural y situacional de los términos. Los corpus informáticos, como recurso léxico de consulta, nos ofrecen la posibilidad de interactuar en conjunto con diferentes libros de textos y otros materiales pedagógicos, convirtiéndose de esta manera en una fuente de ideas.

Con respecto a los buscadores tradicionales que podemos encontrar hoy en día en Internet, los corpus se diferencian por su respaldo académico, y además, porque nos brindan la posibilidad de seleccionar varios criterios. De esta manera, podemos combinar variables como autor, obra, año, tema, país y optimizar así la búsqueda para terminar, de alguna manera, con la inevitable pérdida que todos alguna vez tuvimos en el laberinto de Internet. Por lo tanto, nos permiten, como profesores de español, conseguir material pedagógico fiable y real del uso de la lengua que enseñamos.

Algunas sugerencias en el campo de la enseñanza:[6]

[5] Un estudio sobre la clasificación de los corpus en función de otra serie de criterios puede encontrarse en Sinclair (1996).
[6] Cf. Torruella y Llisterri (1999).

- Buscar el uso frecuente de palabras o construcciones en los libros de textos y lecturas recomendadas.
- Corregir barbarismos o malos usos lingüísticos (errores más repetidos, construcciones no normativas, léxico mal usado, grafías incorrectas, etc.).
- Recopilar corpus de producciones de estudiantes de ELE como fuente de datos.

Otras ideas:

- Contextualizar una palabra o expresión en relación a un tema específico.
- Averiguar el empleo de un prefijo o un sufijo. Obtener términos que comiencen o terminen por un determinado prefijo o sufijo, respectivamente.
- Extraer frecuencias de palabras para comprobar los usos reales.
- Consultar el empleo de ciertas expresiones idiomáticas en diferentes países.
- Comparar el uso de un vocablo entre el oral y el escrito.
- Registrar la combinación de palabras.
- Indagar los rasgos contextuales que acompañan a una palabra o expresión.
- Ante una corrección, el estudiante puede buscar por sí mismo por qué cometió un error relevante.
- Sistematizar el conocimiento intuitivo ante algo que "suene mal", pero no se sepa exactamente el motivo.
- Explorar las colocaciones oracionales para tener en cuenta la posición de los vocablos en el contexto de uso.
- Entre otros, debido a que la utilidad de los corpus como recurso lingüístico es inagotable y depende, a su vez, de la necesidad y creatividad del docente.

3. El *Corpus Oral y Sonoro del Español Rural* (COSER)

El *Corpus Oral y Sonoro del Español Rural* (COSER), dirigido desde la Universidad Autónoma de Madrid por Inés Fernández-Ordóñez (RAE), se empezó a compilar en 1990 para el estudio de la variación dialectal del español en la Península Ibérica.

Algunas de sus características generales:[7]

- Formado por grabaciones de la lengua hablada en enclaves rurales de la Península Ibérica.
- Se distingue de otros corpus orales del español en que registra el habla de individuos cuya vida transcurre en un entorno rural.
- Constituye un complemento de los materiales recolectados en los atlas lingüísticos del español peninsular.
- Informantes: Hablantes rurales, con una edad media global de 72 años (74 en los hombres y 71 en las mujeres), de escasa escolarización y naturales del lugar en que son entrevistados.
- Metodología empleada: "Ha sido la de la entrevista sociolingüística, dirigida por parte de los encuestadores hacia ciertos temas de la vida tradicional en el campo" (Fernández-Ordóñez 2004).
- Entrevistas: en 754 enclaves rurales del centro y de la mitad norte de la Península Ibérica. Consta cerca de 950 horas de grabación.
- Fenómenos dialectales: Posee muestras sonoras breves, con su transcripción correspondiente, que ilustran diferentes empleos dialectales de la gramática y del léxico del español rural.
- Novedad: A diferencia de otros corpus orales dialectales, en los que sólo se brinda al usuario breves muestras sonoras, el COSER ofrece, como muestra de las diversas formas de hablar en la Península Ibérica, las grabaciones en formato audio con sus respectivas transcripciones a disposición de cualquier usuario.

4. Los corpus de la Real Academia Española[8]

El banco de datos de la *Real Academia Española* está formado por dos grandes corpus textuales: el *Corpus de Referencia del Español Actual* (CREA, escrito y oral) y el *Corpus Diacrónico del Español* (CORDE). Entre ambos poseen alrededor de 400 millones de formas de todos los períodos del español, tanto de España como de América y, por lo tanto, constituyen uno de los recursos más importantes para el estudio del español en todo el mundo hispano. Ambos corpus permiten obtener datos reales sobre la consulta realizada, y nos facilitan la combinación de distintos criterios de selección.

Algunas búsquedas que se pueden realizar:

[7] Información y datos extraídos del *Corpus Oral y Sonoro del Español Rural* (*COSER*) dirigido por Inés Fernández-Ordóñez [www.ffil.uam.es/coser/contenido.php?es].
[8] Real Academia Española [www.rae.es].

- Combinar palabras.
- Comprobar las frecuencias de aparición de un término.
- Consultar el uso de palabras y expresiones.
- Averiguar la época o el país en el cual el empleo de una palabra o frase resulta más frecuente.

Combinación de variables:

- Autor.
- Obra.
- Año o intervalo de años.
- Área temática.
- País.

Por medio de la aplicación del CREA y del CORDE podemos obtener diferentes datos de la consulta efectuada, por ejemplo:[9]

- Número total de textos.
- Distribución cronológica.
- Ámbito de aparición.
- Sesgos temáticos.
- Referencia bibliográfica.

4.1 El Corpus de Referencia del Español Actual (CREA)[10]

Datos generales:

- Corpus escrito y oral.
- Textos fechados desde 1975 hasta 2004.
- Cuenta con algo más de 160 millones de formas (mayo de 2008).
- Clasificación temática: Más de cien materias diferentes.
- Los textos escritos corresponden a libros, revistas y periódicos.
- Los textos orales proceden de grabaciones de radio o de televisión, transcritos y codificados.
- Medio: El 90% corresponde a la lengua escrita y el 10% a la lengua oral de España y de América.
- Origen de los textos: 50% de España y 50% de Latinoamérica.

[9] Para un ejemplo de uso demostrado en el taller, véase el Anexo.
[10] Real Academia Española [www.rae.es].

La nueva versión del CREA (junio de 2008)

- Se incorporan algo más de 3,5 millones de formas, correspondientes todas ellas al período 2000-2004.
- Se añade material al bloque correspondiente de la prensa americana, sin embargo, hay también textos procedentes de libros.
- Se agregan las listas de las formas ortográficas registradas en el CREA, con sus frecuencias absolutas y normalizadas.
- Se anexan las listas de las 1000, 5000 y 10000 formas más frecuentes.
- En todas las listas, se ha anulado la diferencia entre grafías con mayúsculas y minúsculas, y a su vez, se han suprimido tanto cifras, como fechas.

4.2 El Corpus Diacrónico del Español (CORDE)[11]

Datos generales:

- Corpus textual.
- Textos fechados desde los inicios del idioma hasta el año 2004.
- Contiene 250 millones de registros (abril de 2005).
- Los textos escritos son: narrativos, líricos, dramáticos, científico-técnicos, históricos, jurídicos, religiosos, periodísticos, etc.
- El Corpus se divide en tres grandes etapas: Edad Media, Siglos de Oro y Época Contemporánea.
- Por su perspectiva diacrónica: otorga 74% para el español peninsular y un 26% para Latinoamérica.

4.3 El Corpus del siglo XXI[12]

- Proyecto de la RAE en colaboración con las veintiuna instituciones que forman la Asociación de Academias de la Lengua Española.
- El trabajo servirá como fuente para los proyectos académicos y proporcionará materiales básicos para la investigación lexicográfica y gramatical del español.
- Se ampliará el banco de datos léxicos.

[11] Real Academia Española [www.rae.es].
[12] Real Academia Española [www.rae.es].

5. *Corpus del español*[13]

El *Corpus del español* (CdE), creado por Mark Davis, almacena una colección de más 20.000 textos orales y escritos que ofrecen una buena representación del español utilizado desde el siglo XIII hasta el siglo XX. Estos textos, son una muestra del uso real de la lengua, y pone a nuestra disposición más de 100 millones de palabras para la creación de materiales didácticos y para el desarrollo de investigaciones lingüísticas. El CdE, con una interfaz en español y otra en inglés para utilizar la que más nos convenga, nos permite hacer búsquedas de *palabras exactas* o *frases* por medio de *comodines*, *etiquetas*, *lemas* o *categoría gramatical*. De igual manera, se pueden combinar todas estas herramientas para hacer una búsqueda más avanzada. Por otro lado, en el CdE podemos también, por medio de una suscripción gratuita, hacer búsquedas por las frecuencias de uso de palabras. A su vez, nos permite comparar las frecuencias entre textos orales o escritos (sean de ficción, prensa o registro académico). Otro tipo de comparación que se puede buscar en este corpus es por campos semánticos, sinónimos, etc. No cabe duda de que el *Corpus del español* es una poderosa herramienta para hacer un sinfín de investigaciones o, lo que más interesa en el ámbito pedagógico, para la realización de materiales didácticos para el aula de ELE. Presentamos, a continuación, algunas indicaciones sobre su funcionamiento, que permitirá a los profesores de ELE conocerlo e indagar más de acuerdo a sus propios intereses.

5.1 Uso del Corpus del español

El uso del CdE se puede presentar como algo complejo, pero cabe recalcar que este corpus ha tenido ciertas modificaciones (o actualizaciones) de aspecto técnico y que facilitan la investigación de los usuarios. A pesar de estos cambios, es importante mencionar que para evitar sorpresas, hay que ser cuidadoso al escribir las palabras, lemas, etiquetas y demás componentes que permiten recuperar los datos deseados. Siguiendo los pasos que les presentamos, podemos tener nociones básicas del uso de este corpus y así ver cuan útil, interesante y hasta adictivo puede ser el uso del CdE.

[13] *Corpus del español* [www.corpusdelespanol.org/x.asp].

5.2 Secciones

Para facilitar la explicación, enumeramos cada una de las secciones que se encuentran en la primera página que encontramos al comenzar una consulta:

- Mostrar.
- Buscar.
- Secciones.
- Ver opciones.
- "Sumario de resultados" (no tiene un título particular visible).
- Resultados.

Sección 1
Al pulsar en el signo de interrogación, que se encuentra al lado derecho, nos aparece la descripción de cada una de las opciones y funciones que forman esta sección. No obstante, explicamos brevemente en qué consiste cada una de ellas.

Para realizar una búsqueda, se debe seleccionar en la *sección 1* si se desea obtener los resultados en *tabla*, en *gráficos* o si se quiere obtener, por ejemplo, la *comparación de palabras*. Es de destacar que las soluciones en la *tabla* podría ser la mejor opción, porque es la que nos permite recuperar los datos con mayor facilidad. Esta opción nos muestra los resultados en la *sección 5*, clasificados desde el siglo XIII hasta el siglo XX. De igual manera, en la *sección 5*, encontramos la clasificación por textos académicos, periodísticos, de ficción (novelas) y orales.

En la opción de *gráficos*, como su nombre lo indica, encontramos la clasificación por medio de gráficos que nos facilita hacer la comparación de frecuencias de uso de la palabra que buscamos entre los diferentes siglos.

La opción *comparación de palabras*, nos permite diferenciar el uso de dos palabras cuyos significados son similares (por ejemplo, *saber* y *conocer*).

Sección 2
En esta sección se coloca, en el espacio designado para este fin, la(s) palabra(s) que deseamos buscar. En la misma, encontramos la opción que nos permite escribir el contexto en el cual aparecería dicha palabra. Es decir, buscar la palabra que consultamos acompañada de otro vocablo, preposición, artículo, etc. Para que aparezca el espacio específico que permite introducir esta información (necesario para colocar los datos que

determinan el contexto), es imprescindible pulsar en el signo de interrogación que se encuentra a la derecha de la palabra *contexto*.

La manera de colocar esta información es muy específica. Para tal fin, es necesario utilizar las *etiquetas*, es decir, la codificación diseñada para este corpus. Al elegir en el signo de interrogación, que se encuentra a la derecha de *categoría gramatical* (abreviado, *CAT GRAM*), nos aparece una lista de opciones que nos facilita seleccionar la categoría gramatical de la palabra que queremos que acompañe a la palabra que buscamos. Si volvemos a pulsar en el signo de interrogación, aparecerá en la *sección 6* la lista de *etiquetas* necesarias para la búsqueda. Junto con el espacio designado para colocar las etiquetas, aparecen dos números, que se pueden cambiar a números menores o mayores. Estos números nos permiten indicar la cantidad de palabras que pueden existir entre la palabra que buscamos y la palabra que acompaña.

Para agilizar la búsqueda, encontramos en esta sección la útil opción *Lista usuario*. Esta opción, nos permite crear un historial de las palabras que queramos almacenar durante la búsqueda, y que podremos volver a consultar en otro momento que sea necesario. Estas listas se guardan bajo un nombre particular sin la necesidad de claves; es decir, no es confidencial. Al lado de esta opción, tenemos los botones *Buscar* y *Borrar*. Obviamente, el primer botón sirve para lanzar la búsqueda con los datos indicados y el segundo para eliminar todos los datos que se hayan colocado y así empezar una nueva búsqueda desde el comienzo.

Sección 3
A continuación, encontramos la opción llamada *Secciones*. Esta nos permite delimitar nuestra búsqueda según el siglo (o los siglos) en el cual fue utilizada la palabra que consultamos. Además, encontramos la frecuencia mínima que queramos que aparezca el lema en los siglos indicados, esto es, si la palabra sólo aparece en tres textos en determinado siglo, el CdE no incluiría estos en los resultados. Por supuesto, podemos aumentar el número de frecuencia que deseamos.

Sección 4
Al seleccionar el signo de interrogación que se encuentra a la derecha de la palabra *Opciones*, aparecen en la *sección 6* cuatro opciones de preferencias que nos permiten personalizar la presentación de los resultados que se obtendrán.

Sección 5
En esta sección, los resultados de la búsqueda lanzada aparecen en una lista, ya sea por frecuencia, gráficos o por comparación de palabras, según la opción que se haya escogido en la *sección 1*. Los resultados que se muestran son enlaces que nos permiten ver en la *sección 6* un fragmento del texto en que se encuentran la palabra que hemos buscado.

Sección 6
Nos brinda la lista enumerada de los resultados de las búsquedas, indicando en números arábigos el siglo en el cual se produjo el texto; el guión y la letra que acompaña indica el tipo de texto (oral, ficción, científico, etc.). A la derecha, vemos el título del texto y posteriormente una línea del texto. Al pulsar en el número que representa el siglo o en el título, aparecerá un fragmento más extenso.

6. *Así hablamos*[14]

Es un diccionario latinoamericano informal, de fácil consulta, y en el que podemos indagar sobre la diversidad de la variación lingüística del español.

Características generales:

- Permite la búsqueda del significado de una palabra en todos los países de habla hispana, o solamente en un país en particular.
- Contiene ejemplos de uso de los diferentes términos según los países.
- El diccionario se construye con las contribuciones de los usuarios.
- Los aportes son publicados directamente y revisados posteriormente por los editores.
- Posee un foro en el cual los visitantes del sitio se pueden comunicar directamente para intercambiar ideas sobre algún tema de interés general, o sobre el uso de algún vocablo o frase.

7. *Jergas de Habla Hispana* (JHH)

Jergas de Habla Hispana es un sitio disponible en Internet que no es considerado como un corpus; no obstante, la información recolectada en este portal sirve como herramienta para la enseñanza del español.

[14] *Así hablamos* [www.asihablamos.com].

Es probable que esta base de datos no goce del mismo prestigio que otros corpus disponibles en Internet, por no contar con el respaldo académico. Como cuenta su creadora, Roxana Fitch, el sitio JHH nace gracias a la curiosidad de una amiga española por saber el significado de palabras "raras" que mencionaban en telenovelas mexicanas. Sin embargo, hoy en día existe la primera edición impresa del *Diccionario de jergas de habla hispana*. Por tanto, se debe reconocer la labor de Fitch y del grupo de colaboradores, que han hecho crecer este corpus de manera voluntaria desde 1997.

A pesar de no tener un respaldo académico, es importante resaltar que los resultados que podemos recuperar en JHH son bastante fiables. Las personas que han asumido la responsabilidad de editar este corpus se han dado la tarea de verificar si los vocablos se usan realmente en el país que se indica. A su vez, brindan ejemplos de uso con los verbos y/o preposiciones que normalmente acompañan a la palabra en cuestión.

La evolución de este sitio web ha sido notable. Hasta ahora, se encuentran casi 2000 términos disponibles para nuestras consultas, pero hay que saber que está en constante crecimiento y que se mantiene siempre activo. Además, desde 2006, se comenzó a registrar muestras de audio para permitir a los usuarios darse una idea de cómo son los acentos en los diferentes países. Esta evolución se proyecta a la futura creación de un diccionario de sinónimos con el argot de cada país. Incluso, se piensa crear una sección interactiva en la cual los usuarios podrán intercambiar sus conocimientos, experiencias y anécdotas relacionadas con las diferentes jergas.

A diferencia de muchos corpus, este permite hacer búsquedas de manera sencilla y rápida. En la primera página, en la parte inferior, encontramos un buscador y un filtro de búsqueda por país. Para hace una investigación simple, escribimos en el buscador la palabra que queremos recuperar. De esta manera el corpus nos mostrará dos listas de resultados: La primera nos ofrece los países[15] en donde se utiliza esta palabra, una breve explicación del significado y por lo menos un ejemplo de uso. La segunda, muestra otros resultados en donde aparece la palabra en cuestión o un vocablo similar. Dentro de las explicaciones y ejemplos podemos pulsar las palabras de color azul, enlace que nos lleva a la explicación y los ejemplos.

[15] Es importante recalcar que esta lista no es exhaustiva, puesto que los datos disponibles depende de la colaboración de los propios usuarios.

Si queremos limitar nuestra búsqueda a un país en particular, lo podemos realizar seleccionándolo en la lista que se encuentra disponible en la primera página. Una vez elegido el país, se pulsa en *ver* y aparece una nueva página que nos solicita la primera letra de la palabra que queremos buscar. Luego de haber escogido la letra, se selecciona en *mostrar* y nos aparecerá las palabras incluidas en el corpus que comienzan por esta letra y que se utilizan en el "país filtro". Si deseamos ver en qué otros países se utiliza la palabra que hemos filtrado, el buscador nos permite lanzar de nuevo la búsqueda. Este paso se realiza pulsando un botón que aparece en la misma página de resultado.

Además de los buscadores principales, esta base de datos nos ofrece también otras secciones disponibles a la izquierda de la página principal. Entre estas, encontramos *curiosidades jergales*, *ejemplos de oraciones en jerga*, *términos compartidos* y *canciones en jerga*. Cabe resaltar que estas secciones se encuentran en una misma barra de navegación; no obstante, el orden de aparición en la lista de secciones varía cada vez que se abre una página.

En la sección *curiosidades jergales*, se muestra un repertorio de palabras escogida por Fitch y sus colaboradores durante sus horas de trabajo. El término *curiosidades* hace referencia a que una palabra puede tener varias acepciones en diferentes países. Esta lista de palabras es corta, pero está en constante crecimiento.

Otra sección interesante es el apartado *ejemplos de oraciones en jerga*. Esta sección no está terminada, pero podemos consultarla aunque se encuentra en desarrollo. En la misma encontraremos tres frases escritas en español estándar. En la parte inferior, vemos la lista de países disponibles para la búsqueda. Al pulsar en el nombre del país que escogemos, se presentan las tres frases que expresan lo mismo que aquellas que se encuentran en la parte superior de la página, pero utilizando el argot del país seleccionado. Esto es muy útil para hacer comparaciones, aunque el objetivo de esta sección es escuchar los ejemplos de los diferentes países con sus respectivos acentos. Al lado izquierdo de cada frase encontramos uno o dos íconos que representan una pequeña bocina o altavoz, al pulsar en ella podemos escuchar la lectura de cada frase.[16]

Este corpus ofrece también una sección llamada *términos compartidos* en la cual se presenta una lista de 1936 vocablos en orden alfabético. Al ver la

[16] Nótese que para reproducir el audio es necesario tener Windows Media Player o QuickTime, ya que los dos tipos de formatos están disponibles.

serie de palabras disponibles, resulta casi imposible no consultar más de una. Además de la curiosidad que pueda generar este catálogo de palabras, su fácil uso contribuye al deseo de ver más resultados. Simplemente pulsando la palabra que deseamos averiguar nos aparecerá una lista de países en donde se utiliza el vocablo, con su respectiva frase de uso.

La siguiente y última sección se encuentra en la barra de navegación, a la izquierda de la página, separada bajo el título de *canciones en jergas*. En este apartado tenemos a nuestra disposición la transcripción de tres canciones compuestas en argot. Estas representan la jerga mexicana, argentina y española con las canciones, respectivamente, *La chilanga banda*, *Chorra* y *La sociedad es la culpable*. Al leer la letra nos daremos cuenta de que hay varias palabras que no conocemos, pero es de destacar que estos vocablos constituyen, a su vez, enlaces que nos conducen a descubrir el significado de los mismos.

Algunas sugerencias en el campo de la enseñanza para estudiantes del nivel intermedio o más avanzado:

- En equipos, escoger un país y crear una lista de palabras particulares de su argot, y de los términos compartidos con otros países para realizar una presentación oral.
- Comprobar el uso de estas voces en el CdE.
- Consultar *blogs* y foros disponibles en Internet, ver qué palabras no comprenden posteriormente y averiguar si se trata de una palabra de una jerga.
- Realizar un corpus propio de palabras nuevas, por ejemplo cuando los estudiantes viajan a un país de habla hispana.

8. Ventajas y desventajas del uso de los corpus

Los corpus pueden ser de gran utilidad para el desarrollo de materiales o para el uso en el aula de ELE. Sin embargo, hay que saber que al principio su empleo no siempre es fácil, y esto puede desanimar a los estudiantes y a los mismos profesores que preparan materiales. Aunque puedan crear ciertas incomodidades por la falta de experiencia, no se puede negar que el uso de los corpus es muy funcional y presenta más ventajas que desventajas.

A pesar de que su manejo puede ser complejo, los corpus pueden ser una manera muy interesante de conocer el uso real de la lengua española. Ya sabemos que, en los últimos años, los materiales de enseñanza han tratado de utilizar textos que se asemejen lo más posible a textos reales. En

consecuencia, el contenido de los corpus, tal y como expone Campillos Llanos (2005-2006):

> Procede directamente del uso del lenguaje, y no de la norma académica o lo gramaticalmente correcto, que puede falsear el modelo de lengua que se aprende. Frente a las afirmaciones de gramaticalidad con criterios heterogéneos y a veces contradictorios, o con consideraciones extralingüísticas, se toma como medida objetiva, relativa y graduada el uso actual de la lengua en un corpus.

Por ello, este instrumento tecnológico es una gran herramienta para la creación de materiales didácticos. Otro atributo de los corpus, que no se puede obviar, es su acceso fácil y gratuito. A su vez, el uso de los corpus nos permite la creación de actividades inductivas. De esta manera, los estudiantes serían aprendices activos en el aula y en casa, facilitándose así la motivación y la participación. Por esta razón, vale la pena invertir un poco de tiempo para explicar en las clases de ELE la utilización de los mismos. Sin duda, una buena manera es comenzar con pequeñas actividades de búsqueda, para que los estudiantes se familiaricen con ellos. Posteriormente, se pueden ofrecer actividades en las cuales ellos tengan que descifrar las reglas por medio de la consulta de los corpus.

Otra virtud que tienen los corpus, como indica Campillos Llanos, recogiendo la idea de Alcón Soler (2000) es que:

> La lengua aparece integrada en el contexto discursivo, y con respecto a la lengua oral, permite apreciar sus peculiaridades: estilísticas (uso de vacilaciones, reformulaciones, evaluaciones y comprobación del discurso, uso frecuente de la elipsis, presencia de oraciones simples, etc.), textuales (elementos de cohesión, estructura de pares adyacentes, aperturas y cierres conversacionales, etc.), o pragmáticas (funciones de los marcadores discursivos, marcas pragmáticas de cortesía, etc.).

Lo mencionado anteriormente no siempre se puede apreciar en los textos de aprendizaje, y muchas veces los profesores nos vemos limitados por la falta de ejemplos reales. Situación ésta, que nos lleva a la creación de otros ejemplos.

Por otro lado, los corpus también pueden presentar desventajas. Una de ellas suele ser de naturaleza técnica, bien por la falta de instrucción en el uso de los corpus, bien por las fallas en la conexión a Internet. No obstante, estos problemas o riesgos son comunes en el uso de cualquier herramienta

tecnológica. Incluso, debido a que parte del contenido de los corpus son transcripciones textuales de muestras orales, se pueden encontrar errores o vocabulario no deseado para el aprendizaje de la lengua. Para evitar inconvenientes, recomendamos que se les explique a los estudiantes que estos textos son para estudios lingüísticos y que, como en todo idioma, en español los hablantes nativos se expresan de una manera particular. En efecto, los corpus son un instrumento tecnológico que refleja el contexto mismo en que se utiliza la lengua e intentan ser un modelo de la realidad lingüística. Como tal, muestran el uso que sus hablantes hacen de ella con todas sus variedades.

Referencias bibliográficas

AsíHablamos.com. 2006-2008, [en línea: www.asihablamos.com/].

Campillos Llanos, L. 2005-2006. *Adaptación del corpus C-ORAL-ROM a la enseñanza de español para extranjeros*. Madrid: Universidad Autónoma de Madrid [documento en línea: www.mepsyd.es/redele/Biblioteca2007/LeonardoCampillos/Memoria.pdf].

Córdoba Rodríguez, F. 2001. "El uso de los corpus lingüísticos en la enseñanza del español". *Boletín de la Asociación de Profesores de Español de la República Checa*, [documento en línea: http://oldwww.upol.cz/res/ssup/ape/boletin2001/cordoba.htm].

Fernández-Ordóñez, I. (dir.). 2004-2009. *Corpus Oral y Sonoro del Español Rural* (COSER). Madrid: Universidad Autónoma de Madrid, [en línea: www.ffil.uam.es/coser/contenido.php?es].

Fernández-Ordóñez, I. 2004. "Nuevas perspectivas en el estudio de la variación dialectal del español: el *Corpus Oral y Sonoro del Español Rural*", en *Actes du XXIV Congrès de Linguistique et Philologie Romanes* (Aberystwyth, Wales, 2-5 August 2004), [documento en línea: http://pidweb.ii.uam.es/coser/publicaciones/ines/5_es.pdf].

Fitch, R. *Jergas de Habla Hispana* (JHH), [en línea: www.jergasdehablahispana.org/].

Leech, G. 1992. "Corpora and theories of linguistic performance", en J. Svartvik (ed.), *Directions in corpus linguistics*. Berlín/New York: Mouton de Gruyter, 105-122.

Mark, D. 2002-2008. *Corpus del español*. Provo: Brigham Young University, [en línea: www.corpusdelespanol.org].

Pérez Hernández, M. Ch. 2002. "Explotación de los córpora textuales informatizados para la creación de bases de datos terminológicas basadas en el conocimiento". *Estudios de Lingüística Española (ELiEs)* 18, [documento en línea: http://elies. rediris. es/elies18/index. html].

Real Academia Española. 2005. *Corpus Diacrónico del Español* (CORDE). Madrid: RAE.

Real Academia Española. 2005. *Corpus de Referencia del Español Actual* (CREA). Madrid: RAE.

San Mateo, A. 2003. "Los corpus lingüísticos y la enseñanza de ELE". *Frecuencia-L* 22: 52-58.

Sinclair, J. 1996. *Preliminary recommendations on Corpus Typology*. Technical report EAGLES (Expert Advisory Group on Language Engineering Standards), [documento en línea: http://www.ilc.cnr.it/EAGLES96/corpustyp/corpustyp.html].

Sinclair, J. (ed.) 2004. *How to use corpora in language teaching*. Amsterdam: John Benjamins.

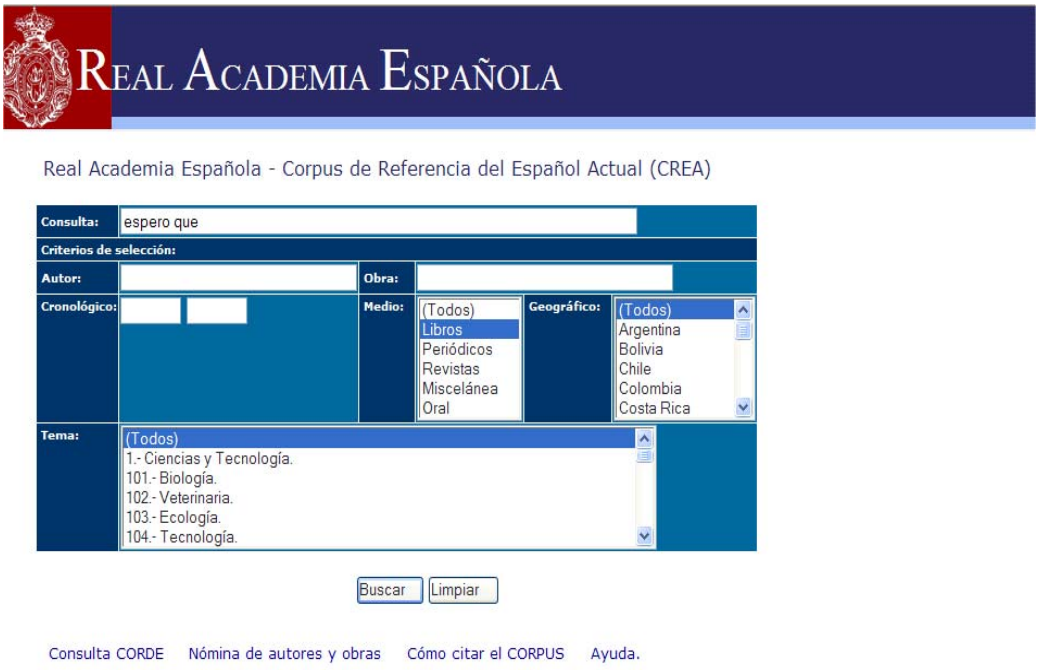
Torruella, J., J. Llisterri. 1999. "Diseño de corpus textuales y orales", J. M. Blecua et al. (eds.), *Filología e informática. Nuevas tecnologías en los estudios filológicos*. Barcelona: Universidad Autónoma de Barcelona/ Editorial Milenio, 45-77, [documento en línea: http://liceu.uab.es/~joaquim/publicacions/Torruella_Llisterri_99.pdf].

Anexo. Ejemplo de empleo del CREA: El uso del subjuntivo[17]

1. Primera ventana: Construcción del perfil de consulta

Se dispone en esta ventana principal de un apartado **[Consulta]** destinado a la redacción de la búsqueda: palabra, prefijo, sufijo, combinación de palabras, etc. A su vez, se configura el sistema de selección determinando, por ejemplo, el autor, la obra, el año o intervalo de año **[Cronológico]**, el área temática **[Tema]**, la procedencia **[Medio]** y el país **[Geográfico]** que es objeto de consulta.

Imagen 1. Ventana principal del CREA.



2. Segunda ventana: Resultados

Se ofrecen los datos estadísticos de la consulta realizada. En el caso de que el número de ejemplos supere los límites prefijados, u obstaculice el objetivo de la consulta, se dispone de un apartado **[Filtros]** que permite limitar el contenido de una consulta.

[17] Información y datos extraídos directamente de la página web de la Real Academia Española [http://corpus.rae.es/ayuda_c.htm].

La selección del apartado **[Recuperar: Concordancias]** recupera los ejemplos (concordancias) de la búsqueda realizada.

En el apartado **[Obtención de ejemplos]** seleccionando la opción **[Documentos]** y pulsando **[Recuperar]**, se obtiene la relación bibliográfica de los documentos que contienen los ejemplos.

Las opciones **[Clasificación]** permiten ordenar los ejemplos según criterios temáticos o contextuales. Por ejemplo, el apartado **[Casos]** ordena los ejemplos de acuerdo con su relevancia.

3. Tercera ventana: Concordancias

En esta sección se visualizan las *concordancias*, es decir, los ejemplos propiamente dichos en una lista de las apariciones de determinada palabra o del objetivo de consulta. Se muestra la descripción bibliográfica y el contexto relacionado con la búsqueda.

4. Referencias bibliográficas de los ejemplos

Se dispone de la descripción bibliográfica completa relacionada con la consulta.

5. Presentación estadística de los datos obtenidos

El apartado denominado **[Estadísticas]** muestra los datos estadísticos correspondientes a la consulta efectuada: ámbito de aparición, los sesgos temáticos y la distribución cronológica de los ejemplos obtenidos.