

Los corpus como recursos didácticos para la enseñanza de las variedades diatópicas del español

Sonia Bailini^A

Università Cattolica del Sacro Cuore (Milán, Italia)

RESUMEN: Este artículo se plantea analizar el potencial de algunos corpus del español –tanto orales como escritos– para la enseñanza de las variedades diatópicas. El marco teórico de referencia es el de la lingüística de corpus aplicada a la enseñanza de lenguas desde la perspectiva del Data-Driven Learning. Se ejemplifica el uso de los corpus como recursos para llevar a cabo lo que se podría definir una didáctica panhispánica de ELE, cuyo objetivo no es tanto que el aprendiente llegue a utilizar una u otra variedad, sino que aprenda a reconocerlas y apreciarlas como ejemplos de riqueza cultural.

Los corpus en la enseñanza de la lengua extranjera

La lingüística de corpus aplicada a la enseñanza de lenguas extranjeras (Aston, 2001; Sinclair, 2004; Aijmer, 2009; Cruz Piñol, 2012) es una línea de investigación relativamente reciente cuya aplicación en el aula, como señalan varios autores (Granath, 2009; Römer, 2009; Boulton y Tyne, 2013), todavía no recibe la atención que se merecería, en parte por razones de tipo práctico y en parte por el cambio de actitud que este enfoque conlleva tanto en el profesor como en el aprendiente. Si por un lado los programas de consulta requieren algunos conocimientos básicos de lingüística computacional, por otro todavía existe cierto escepticismo entre los profesores, relacionado con la falta de formación técnica y didáctica acerca de las posibilidades de explotación de estas herramientas y con una escasa confianza en su eficacia para la enseñanza (Gilquin y Granger, 2010: 366-367).

La utilización de los corpus en el aula de lengua extranjera conlleva que el profesor acepte la idea de que ya no es la única fuente de conocimiento disponible y se considere más bien “a facilitator of the learning process, helping the learners to interpret the data, and giving them advice on how best to search the corpus and analyse their search results” (Chambers, 2010: 355). Esto supone adoptar una actitud descriptiva frente a la lengua antes que prescriptiva, pues los corpus ofrecen muestras de lengua en contextos auténticos y a menudo hacen evidente la *distancia* entre la norma y el uso. Los corpus, por tanto, deberían verse como herramientas de apoyo, ya que son recopilaciones de textos informatizados *representativos*, en la

^A **Sonia Bailini** es doctora en Lingüística Aplicada por la Universidad de Salamanca y profesora de Lengua y Lingüística Española en el Departamento de Ciencias Lingüísticas y Literaturas Extranjeras de la Universidad Católica de Milán. La dirección de su correo electrónico es: sonia.bailini@unicatt.it

medida de lo posible, de una lengua o de una variedad de lengua (Sinclair, 2004). Leech puntualiza que “más que afirmar que el corpus debe ser representativo, [...] ha de ser *aceptado como representativo* por sus numerosos y variados usuarios, incluidos los especialistas académicos, profesores y miembros de la comunidad lingüística en general” (2002: 1). Trabajar con los corpus conlleva, tanto para el profesor como para el aprendiente, adoptar la actitud de *investigador* de la lengua, como sostiene el Data-Driven-Learning (DDL) (Johns, 1991a, 1991b, 2000) o la de *explorador* (Bernardini, 2001 y 2004). El DDL es un enfoque que se centra en el uso de las herramientas y las técnicas de la lingüística de corpus con fines didácticos a través de procedimientos inductivos o deductivos (Gilquin y Granger, 2010: 359). En el primer caso se trata, para el aprendiente, de ir descubriendo las características de la lengua extranjera a través de la observación (*noticing*) de *input* auténticos en distintos contextos lingüísticos para inferir las normas que regulan su funcionamiento o fijarse en algunos aspectos de su uso. En el segundo caso los corpus se utilizan como soporte para la práctica y consolidación de lo aprendido en el aula, por ejemplo para clasificar los distintos usos de una preposición a partir de una serie de enunciados en los que aparece.

El DDL es compatible con todos aquellos enfoques que se centran en la atención a la forma y en la importancia de promover el desarrollo de la conciencia lingüística y metacognitiva así como la autonomía del aprendiente. Asimismo, se compagina también con enfoques comunicativos (Bernardini, 2001), ya que los ítems lingüísticos extraídos de los corpus se pueden utilizar para crear actividades de *problem-solving* en parejas o en grupos. Por último, el DDL encaja también con los enfoques constructivistas, ya que el análisis de los datos puede promover la interacción entre pares y entre éstos y el profesor atribuyendo al aprendiente un papel activo y un control cada vez mayor sobre su propio proceso de aprendizaje. Gabrielatos subraya que el uso de los corpus en el aula “is not meant to replace existing teaching methodologies, but to enrich and enhance them” (2005: 23) y Boulton y Tyne destacan que “the learner-as-a-researcher approach inherent in DDL may rather be viewed as an extension of ordinary practice insofar as it builds on many aspects of existing language teaching and learning methodology” (2013: 107). Básicamente hay dos modalidades de introducir el DDL en el aula:

- el profesor utiliza los corpus como fuentes de recursos para extraer *input* sobre cierto aspecto de la lengua en el que desea llamar la atención de sus alumnos;
- los aprendientes consultan directamente los corpus a partir o bien de una tarea propuesta por el profesor o bien para solucionar alguna duda o curiosidad lingüística.
- Gavioli (2001) destaca la importancia de entrenar gradualmente a los aprendientes al uso de los corpus y recomienda empezar con tareas en que el profesor selecciona de antemano una cantidad reducida de ítems lingüísticos y les pide a los alumnos que observen determinadas características recurrentes a partir de la frecuencia de un verbo o de un sustantivo o que seleccionen las colocaciones más frecuentes en cierto género textual (por ejemplo, en los anuncios de trabajo en los periódicos); también sugiere tareas que ayuden a los aprendientes a familiarizarse con los programas de consulta de los corpus con que van a trabajar. Dependiendo de los objetivos del curso, del contexto de enseñanza y de los destinatarios se pueden utilizar distintos tipos de corpus y nada impide que el profesor o los aprendientes creen su(s) propio(s) corpus según unos criterios fijados de antemano. Cabe destacar, además, que los corpus ofrecen la oportunidad de estar inmersos en una gran variedad de *input* y esto es aún más útil en contextos de enseñanza/aprendizaje de una lengua extranjera –en este caso el español– fuera de un contexto hispanohablante.

Los corpus como recursos para una didáctica panhispánica

Cuando la enseñanza del español se da en un contexto no hispanohablante –Italia, en este caso– donde, sin embargo, hay cada vez más hablantes nativos de español que no proceden de España sino de distintas comunidades hispanas de América a raíz de los masivos flujos migratorios de los últimos veinte años, y cuando los aprendientes son alumnos universitarios de una facultad denominada de “mediación lingüística” que se están formando para desempeñarse en actividades profesionales de enseñanza del español como lengua extranjera, de relaciones internacionales o de administración de empresa, no podemos evitar plantearnos la siguiente pregunta: ¿con qué variedad(es) de español entrarán en contacto? ¿Qué español escuchan en

boca de los hispanohablantes que viven en su país? Se hace necesario, por consiguiente, exponerlos a la mayor cantidad posible de variedades del español no tanto con el objetivo de que aprendan a utilizarlas sino más bien para que tomen conciencia de la diversidad en la unidad y de la riqueza que esto conlleva, ya que la inmensa comunidad idiomática de los hispanohablantes reúne a una multiplicidad de comunidades de habla con perfiles propios y distintivos (Moreno Fernández, 2000: 55). Como apunta Baralo, se trata de hacer hincapié en la variedad diatópica para desarrollar esa competencia pluricultural y plurilingüística que el mestizaje aporta dentro de la misma comunidad hispanohablante (Baralo, 2003: 163). Si bien en los últimos diez años los geolectos del español han empezado a recibir cierta atención en los manuales de ELE, reflejar su riqueza y variedad sigue siendo un reto para autores y editores, como destaca Moreno Fernández:

afortunadamente los medios electrónicos y la (relativa) facilidad del intercambio de productos culturales proporcionan en la actualidad acceso a multitud de muestras del español hablado en muchos lugares del mundo hispánico, aunque, de hecho, ni las editoriales ni los equipos docentes suelen contar con muestras de este tipo especialmente dispuestas para la enseñanza. Al no existir, hoy por hoy, una batería de muestras geolingüísticas bien organizadas y disponibles, lo que realmente encontramos en la práctica de la enseñanza es un manejo poco menos que aleatorio, por no decir errático, de los materiales que reflejan las variedades del español” (Moreno Fernández, 2010: 161-162).

Los corpus, por supuesto, no son la panacea, pero permiten llevar al aula muestras de hablas hispanas entendidas como la manifestación concreta de hablantes de distintas procedencias que tienen como referencia común una única lengua, el español (o “castellano”: como se prefiera llamarlo). Son recursos adaptables a distintos contextos de enseñanza/aprendizaje y a las necesidades de los aprendientes, pero no hay que olvidar que son, ante todo, herramientas estructuradas y diseñadas para la investigación lingüística y no para la enseñanza.

La bibliografía crítica relacionada con la explotación didáctica de los corpus en el aula de lengua extranjera se centra sobre todo en el análisis de las concordancias y las colocaciones pero, si se quiere llamar la atención sobre la variación diatópica, cabe recordar que la fonología y la prosodia son los rasgos más perceptibles, seguidos por el léxico, la pragmática y la sintaxis. Aunque en la actualidad hay muchos tipos de corpus en español, sólo algunos de ellos están estructurados para que se puedan extraer fácilmente datos relacionados con sus geolectos. En este

trabajo vamos a centrarnos en el potencial didáctico de tres de ellos: el *Atlas interactivo de la entonación del español*, el *Corpus PRESEEA* y el *Corpes XXI*.

El *Atlas interactivo de la entonación del español* coordinado por P. Nieto y P. Roseano de la Universidad Pompeu Fabra de Barcelona (<http://prosodia.upf.edu/atlasentonacion/index.html>) cuenta con 24 equipos de investigación en España y en América y “tiene como objetivo la presentación sistemática de una serie de materiales en formato audio y vídeo para el estudio de la prosodia y la entonación de los dialectos del español”¹. Recopila 17 geolectos divididos en 5 macroáreas (peninsular, andino, canario, caribeño, mexicano) y clasificados en 3 tipologías de textos: 1) enunciados con diferentes contornos entonativos cada uno de los cuales expresa un matiz semántico distinto²; 2) conversaciones semidirigidas recogidas con el método del Map-Task y 3) muestras de habla espontánea en forma de entrevistas en vídeo.

El *Corpus PRESEEA, Proyecto para el Estudio Sociolingüístico del Español de España y de América* (<http://preseea.linguas.net/>), coordinado por F. Moreno Fernández de la Universidad de Alcalá (Moreno Fernández, 2005 y 2006), “pretende ofrecer muestras de habla de un número elevado de comunidades de todas las latitudes hispánicas” y entre sus objetivos está el de “reunir un gran corpus oral, sociolingüísticamente representativo y técnicamente adecuado” (Moreno Fernández, 2006: 17-18). En el proyecto están involucrados 40 equipos de investigación en distintas áreas hispanohablantes y la previsión es que cada equipo allegue un corpus oral anotado de entre 300.000 y 500.000 palabras. El número de informantes (seleccionados según las variables de sexo, edad y grado de instrucción) varía dependiendo del tamaño de la ciudad en que se recogen los datos y va de un mínimo de 54 informantes a un máximo de 108. El corpus consta de entrevistas semidirigidas, de una duración entre 45 minutos y hora y media, cuyos temas se centran en módulos temáticos³ que permiten acceder a diversos tipos de discurso (Moreno Fernández, 2006: 20). El *Corpus PRESEEA* está en construcción y en la actualidad están disponibles los corpus del habla de Alcalá de Henares, Caracas, La Habana, Lima, Madrid, Medellín, Monterrey, Montevideo y Valencia⁴.

El *Corpus del Español del Siglo XXI (CORPES XXI)*, coordinado por la RAE y la Asociación de Academias de la Lengua Española (<http://web.frl.es/CORPES/view/inicioExterno.view>), es un corpus de referencia

semiabierto “que constará de textos de todos los tipos y todos los países con un volumen de 25 millones de formas por año”⁵, del período comprendido entre 2001 y 2012 por un total de 300 millones de formas con una distribución general del 70% de textos americanos y el 30% de textos españoles. El 90% de los textos corresponde a la lengua escrita y el 10% a la oral; los materiales escritos proceden de libros (40%), publicaciones periódicas (40%), material de Internet (7,5%) y miscelánea (2,5%). Las variantes diatópicas están clasificadas en 10 áreas: Andes, Antillas, Caribe continental, Chile, España, Estados Unidos, México y Centroamérica, Río de la Plata, Filipinas y Guinea Ecuatorial. La distribución correspondiente a las grandes áreas lingüísticas del mundo hispánico se establece mediante el cruce de criterios diferentes, entre los que figura la población, el volumen de publicaciones y la cantidad de ediciones digitales de periódicos y revistas.⁶

Algunas propuestas de explotación didáctica

En este apartado se presentan algunas propuestas de explotación didáctica para la enseñanza/aprendizaje de las variedades diatópicas del español a partir de los corpus mencionados arriba. Se pueden adaptar fácilmente a cualquier método, en distintos contextos de aprendizaje, teniendo en cuenta los destinatarios y su nivel de competencia. La idea básica es que en cualquier clase de ELE se puede trabajar con los geolectos para ir sensibilizando poco a poco a los aprendientes sobre la riqueza cultural y lingüística del idioma que están aprendiendo. Estas microtarefas han de entenderse, por lo tanto, como breves *incursiones* en la variación lingüística del mundo hispanohablante y *extensiones* de la didáctica cotidiana del aula.

El *Atlas interactivo de la entonación del español* se puede utilizar con alumnos universitarios de niveles intermedios y avanzados (del B1 en adelante) para actividades centradas en la discriminación de los alófonos y los rasgos prosódicos en distintos entornos hispanohablantes. Una primera actividad consiste en seleccionar del menú Map-Task dos textos representativos de dos geolectos del español en los cuales los interlocutores hablan del mismo tema. El Map-Task es un método de recolección de datos en que a dos informantes se les da el mismo mapa de un lugar: en uno de los dos hay un recorrido marcado y en el otro no. El interlocutor que tiene el mapa con el recorrido tiene que darle indicaciones al otro para que éste llegue a trazar el mismo trayecto en el suyo. En este caso se han

seleccionado el geolecto de Castilla (Madrid) y el de Andalucía occidental (Jerez de la Frontera). En ambos casos se trata de habla oral semiespontánea en contextos informales y, por la información que se puede extraer del corpus, podemos inferir que se trata de hablantes con un grado de instrucción medio-alto⁷. Tras escuchar dos o tres veces los dos audios en secuencia, se pide a los alumnos que apunten los rasgos fonológicos que caracterizan las dos hablas. Dependiendo del nivel de competencia, esta actividad se puede hacer de manera más libre, a través de preguntas abiertas, o más guiada, presentando un cuadro en que se indican unos rasgos fonético-fonológicos y se pide a los aprendientes que vayan marcando los que oyen en los dos audios.

Opción 1.

Vas a escuchar dos audios en que dos hablantes de origen distinto explican el recorrido para ir de un sitio a otro. Intenta contestar a las siguientes preguntas:

1. ¿Qué rasgos de la pronunciación te han llamado más la atención?
2. ¿Hay algún sonido que te suena raro/distinto con respecto a los que conoces?
3. ¿Qué diferencias observas en la entonación?

Opción 2.

Vas a escuchar dos audios en que dos hablantes de origen distinto explican el recorrido para ir de un sitio a otro.

Marca los rasgos de pronunciación que reconozcas:

Rasgo fonético-fonológico	Audio 1		Audio 2	
	Sí	No	Sí	No
Pérdida o aspiración de /s/ en final de sílaba o palabra				
Mantenimiento de /s/ en final de palabra				
Pronunciación de /θ/ como si fuera /s/ (seseo)				
Discriminación entre /s/ y /θ/				
Pronunciación de /s/ como si fuera /θ/ (ceceo)				
Abertura de vocales en final de palabra				
Pronunciación de /t/ como /tʃ/				
Pronunciación de /tʃ/ como /s/				
Pronunciación marcada de /x/				
Pronunciación aspirada de /x/				

Cuadro 1. Actividad 1: ¿Hacia dónde voy? - Atlas interactivo de la entonación en español

Este corpus también brinda la oportunidad de trabajar con la entonación del mismo tipo de oración en distintos contextos hispanohablantes. Desde la pestaña “búsqueda por frases” es posible seleccionar la modalidad oracional, el tipo de

oración y el dialecto (<http://prosodia.upf.edu/atlasentonacion/cerca/index.php>). Por ejemplo, se pueden extraer, entre los enunciados imperativos, las órdenes en todos los geolectos y seleccionar uno o dos por cada una de las cinco variedades recopiladas en el corpus (peninsular, andino, canario, caribeño, mexicano). En la ficha de cada enunciado aparece la situación en que se emite, el archivo audio, la transcripción ortográfica, el contorno de frecuencia fundamental y la descripción de la curva melódica (véase cuadro 2). Al margen de estos datos técnicos, que pueden ser más o menos relevantes dependiendo del contexto de aprendizaje en que se utilicen, se pueden hacer ejercicios de discriminación de las diferencias de entonación. Por ejemplo, se pueden copiar en desorden en una hoja las curvas melódicas, pedir a los alumnos que, durante la escucha, las asocien con el enunciado correspondiente y después mostrarles a qué área corresponden e invitarlos a escuchar más ejemplos de cada una de ellas haciéndolos trabajar directamente con el corpus⁸.



Cuadro 2. Actividad 2: ¡Ven aquí, por favor! – Atlas interactivo de la entonación en español

Después se pueden agrupar las transcripciones de los enunciados y observar las formas en que están contruidos para ver las diferencias gramaticales, pragmáticas y léxicas⁹.

*Esta es la transcripción de los enunciados que acabas de escuchar.
Teniendo en cuenta que en todos los casos
la relación entre los dos hablantes es informal ¿qué observas? Fíjate ahora
en el léxico: ¿qué similitudes y diferencias notas?*

Área hispanohablante	Tipo de enunciado: orden
Bogotá	¡María, por favor, venga para acá!
Buenos Aires	¡Natalia, vení para acá!
Las Palmas de Gran Canaria	¡Ven aquí ahora mismo!
Madrid	¡Ven aquí, por favor!
México DF	¡Ven aquí ahorita mismo!

Cuadro 3. Actividad 3: ¡Ven aquí, por favor! - Atlas interactivo de la entonación en español

La interfaz de consulta del *Corpus PRESEEA* (<http://preseea.linguas.net/Corpus.aspx>) permite seleccionar los textos, completos de la transcripción ortográfica y de la de los rasgos de oralidad según los criterios del Grupo Val.es.co¹⁰, teniendo en cuenta las variables de género, edad¹¹ y grado de estudios. La metodología con que se recopilaron los datos consta de una parrilla de preguntas organizadas por temas, que pueden ser de gran ayuda a la hora de seleccionar material para el aula de ELE. A continuación se transcribe un fragmento relacionado con el módulo temático “lugar donde vive el informante”:

¿Dónde vives? ¿Cómo es tu casa? Descríbemela un poco. ¿Qué es lo que más te gusta de ella? ¿Por qué? ¿Ha estado siempre así o has hecho reformas? ¿Cómo era antes? ¿Llevas mucho tiempo viviendo allí? ¿Dónde vivías antes? ¿Te gusta vivir aquí? ¿Dónde te gustaría vivir? ¿Por qué? ¿Conoces a mucha gente por aquí? ¿Qué tal te llevas con ella? ¿Cómo te gustaría que fueran tus vecinos? ¿Qué relación te gustaría tener con ellos? ¿Qué haces tú para mantener o mejorar esa relación?

Cuadro 4. Preguntas sobre el tema ‘lugar donde vive el informante’ - Corpus PRESEEA¹²

Siguiendo un criterio temático se pueden seleccionar dos o tres fragmentos de entrevistas a informantes de procedencia distinta que hablan del mismo tema y pedir a los alumnos que observen los rasgos de la oralidad característicos de una u otra

variedad y luego extraer algunos pasajes de texto en que aparezcan e invitar a un análisis más detallado. Este corpus también brinda la oportunidad de llevar al aula aspectos pragmáticos del mundo hispano por boca de sus propios hablantes, ya que cada entrevista se abre con un breve intercambio sobre el uso de los tratamientos formales e informales en la comunidad de habla del sujeto entrevistado, lo que se presta a la elaboración de tareas centradas tanto en el contenido como en la forma.

1. SALUDOS - Presentación

¿Cómo quiere que le trate, de tú o de usted? La verdad es que es un problema esto del tratamiento, nunca sabes cómo tratar a las personas ¿verdad? Por ejemplo, ¿tú cómo tratas a tus amigos, de tú o de usted? ¿Y si son personas mayores? ¿Si se trata de alguien joven que no conoces? ¿Y si es una persona mayor, hombre o mujer, al que, por ejemplo, le preguntas por una calle? ¿Y a tu médico? ¿A los vecinos con los que no tienes mucho contacto? ¿Y cómo te/le gusta que te/le traten a ti/usted? Si una persona más joven te/le trata de tú qué te/le parece? (Preguntar por si tratan de evitar el trato asimétrico o en qué situaciones les parece mejor). Es un problema, yo a veces no sé qué hacer. Bueno, a mí trátame de tú ¿le/te parece?

Cuadro 5. Preguntas iniciales sobre el uso de los tratamientos - Corpus PRESEEA¹³

Vas a escuchar cuatro conversaciones en que dos españoles y dos uruguayos hablan de los tratamientos que se suelen utilizar en su país en las relaciones formales e informales. Escucha los textos varias veces y trata de completar la siguiente tabla:

Tratamientos	Español peninsular	Español rioplatense
¿Cuál es el tratamiento favorito?		
¿Cuáles son los criterios que permiten decidir si utilizar un tratamiento formal o informal?		
¿En cuál de las dos variedades de español es más frecuente el uso de usted?		
¿Qué tipo de relación expresa el uso de 'tú' en estas dos variedades de español?		

Cuadro 6. Actividad 4: ¿Tú, vos o usted? – Corpus PRESEEA

Observa la transcripción de algunos fragmentos de las conversaciones de los dos hablantes uruguayos que acabas de escuchar y fíjate en qué cambios conlleva el uso del 'vos':

E.: [...] Bueno / lo primero que quiero plantearte es el asunto del tratamiento.... yo generalmente hablo de vos así que te voy a tratar de vos, pero yo te pregunto formalmente ¿a vos cómo te gusta que te traten? ¿de usted, de vos, de tú?

I.: de vos [...]

E.: de vos, ¿y voseás el verbo? o sea hacés el verbo agudo ¿vos tenés?

I.: sí

E.: ... y ¿si son personas mayores?

I.: si son personas mayores depende de la relación que tengan conmigo

E.: a ver, explicame un poquito eso [...]

E.: ¿si encontrás a alguien en la calle y tenés que preguntarle alguna cosa? ¿la hora / una dirección? [...]

E.: ... y a vos en general te gusta que te traten de vos

I.: sí

E.: [...] ¿y no te sentís mal por eso? ¿nunca?

I.: no

E.: es lo natural

I.: sí

(MONV_M13_003 (Corpus Preseea))

E.: [...] ¿vos cómo tratás a tus amigos?

I.: y... a mis amigos de vos

E.: de vos

I.: sí, yo creo que tiene que ver con la edad, más que nada eso ¿no?

E.: puede ser... ¿por qué decís....? A ver...

I.: y... porque a una persona mayor capaz que la tratás de usted, o también con el cargo capaz que una persona que tiene un cargo un poco importante, eh o sea a un jefe tuyo por ejemplo o algo así...

E.: ahí está [...] ¿y a vos te parece que el usted es una muestra de mayor respeto que vos? [...] Eeh vos me tratás de che a mí, ¿no? ¿me decís vos?

I.: sí

E.: está bien, ningún problema... por ejemplo ¿a tu médico? ¿has ido al médico?

I.: sí, dejame pensar... hace tiempo que no voy pero... como que lo veo de las dos maneras

E.: ¿sí?

I.: de vos, creo que de vos [...]

E.: ¿vos encontrás alguna diferencia entre decir 'decime la hora' o... dime la hora ooo...?

I.: sí, yo nunca uso tú... nunca

E.: nunca... y el verbo ¿conjugado así, como 'vos sabés'?

I.: claro, claro

(MONV_H12_006, Corpus Preseea)

Intenta resumir los cambios relacionados con la conjugación de algunos tiempos verbales y compáralos con el paradigma verbal 'tuteante'.

Cuadro 7. Actividad 5: ¡Voseame, che! – Corpus PRESEEA

El *Corpes XXI* (<http://web.frl.es/CORPES/view/inicioExterno.view>) se puede aprovechar sobre todo para el análisis de la variación diatópica en textos escritos y se presta a actividades centradas en el léxico y la morfosintaxis. La interfaz de búsqueda permite identificar, al lado de cada concordancia, el geolecto del que procede el texto, el año, el tipo de texto y la indicación de la categoría gramatical a la que pertenece el lema que se ha buscado. Además, el sistema permite buscar la frecuencia de un lema por “zona” geográfica y por “país”. Esto permite planear tareas que estimulen a los aprendientes a observar cómo se manifiestan la geosinonimia y la polisemia de algunas palabras en distintas comunidades hispanohablantes y a *explorar* los múltiples contextos en que aparecen, además de promover el trabajo autónomo con los corpus.

Las siguientes palabras se usan en distintas áreas hispanohablantes para referirse a los niños: ‘nene’, ‘guagua’, ‘crío’, ‘chiquitín’, ‘chaval’, ‘patojo’, ‘pibe’. Busca en el Corpes XXI algunos ejemplos de uso con la función ‘concordancia’ y, observando el contexto en que aparecen, comprueba si el significado de todos ellos corresponde siempre sólo al de ‘niño’. Después, utilizando la función ‘estadísticas’ busca su frecuencia en los distintos países y elabora un cuadro en que se muestre qué palabras son más frecuentes y dónde.

Cuadro 8. Actividad 6: ¡Cuántos niños! – Corpes XXI

Esta actividad invita a los aprendientes a consultar directamente el corpus y a *investigar* sobre la lengua. Los resultados de la búsqueda, que se resumen en el cuadro 9, muestran la frecuencia absoluta de los geosinónimos en las distintas áreas según la clasificación del *Corpes XXI*.

Área	‘nene’ 2502 c/ 836 doc.	‘guagua’ 668 c/ 256 doc.	‘crío’ 913 c/ 252 doc.	‘chiquitín’ 61 c/ 27 doc.	‘chaval’ 1792 c/ 758 doc.	‘patojo’ 67 c/ 25 doc.	‘pibe’ 1054 c/ 465 doc.
Andes	55	14	14	1	3	1	14
Antillas	186	220*	11	3	14	0	2
Caribe	135	3*	18	3	14	4	28

Cont.							
Chile	29	318	24	2	9	0	13
España	417	40*	754	16	1590	0	91
EEUU	8	2*	1	0	1	0	2
Filipinas	3	0	0	0	0	0	0
Guinea Ecuat.	57	15*	3	0	6	0	0
Méx y C. Amér.	246	20*	41	30	107	62	28
Río de la Plata	1366	8	47	6	48	0	876

Cuadro 9. Síntesis de los resultados de la búsqueda de los geosinónimos de ‘niño’ en el *Corpes XXI*¹⁴

Sintetizando los resultados se puede inferir que “nene” es la palabra más compartida en los distintos geolectos y la más usada en el Río de la Plata, mientras que en España son más frecuentes “chaval” y “crío”. También se observa que, en el Río de la Plata, el sinónimo más frecuente de “nene” es “pibe”, mientras que en México y Centroamérica es “chaval”. Los datos muestran, además, que “patojo” se usa casi exclusivamente en esta zona, por lo tanto es interesante refinar la búsqueda “por país” para descubrir su distribución: de los 67 casos encontrados 59 se dan en el español de Guatemala, 4 en el de Colombia, 3 en el de México y 1 en el de Ecuador, por lo que se puede afirmar que es una forma regional bien delimitada.

El análisis de las concordancias de la palabra “guagua” deja claro que mientras en algunos países es sinónimo de “niño” en otros lo es de “autobús” (cuadro 10). Esto llevará a los aprendientes a refinar la búsqueda “por país” y no sólo “por zona” para sacar los datos relacionados con la frecuencia de esta palabra con ese significado. De este modo descubrirán, por ejemplo, que de los 668 casos de “guagua” encontrados en 256 documentos, 300 (marcados con * en el cuadro 9) se refieren a su uso para indicar el “autobús”, que 340 son sinónimo de “niño” y que en Chile es donde más se usa con este significado. Además, resulta que en Colombia (9 casos) la “guagua” es un roedor (*Dinomys branickii*) y en Ecuador en 11 de 21 casos se utiliza como nombre propio de un volcán (Guagua Pichincha); por último, los 40 casos registrados en España se refieren al geolecto canario.

The screenshot shows a web browser window with the URL `web.frl.es/CORPES/org/publico/pages/consulta/entradaCompleja.view`. The search interface includes a navigation bar with options like 'Concordancias', 'Coapariciones', 'Configuración', 'Ayuda', 'Modo de cita', and 'Sugerencias'. The search parameters are: Lema: 'guagua', Forma: 'guagua*', Clase de palabra: 'sustantivo'. The results show 668 cases in 256 documents. A table lists results with columns for REF. (Classification, country), the concordance text, and the word 'guagua' highlighted in yellow. A tooltip is visible over one of the results, showing the definition: 'guagua - sustantivo masculino plural común'. The bottom of the browser shows the taskbar with icons for Windows, Firefox, and other applications, and the system clock indicating 18:15 on 20/06/2014.

Cuadro 10. Resultados de la búsqueda de concordancias del lema ‘guagua’ = ‘niño’ en el *Corpes XXI*

Al buscar las concordancias del lema *crío* los aprendientes tendrán que fijarse en que puede ser también una forma verbal y, por consiguiente, seleccionar en la interfaz la opción “clase de palabras” que les permite buscar sólo el sustantivo. Observando los varios contextos en que se encuentra, entenderán que se suele utilizar también para indicar a “los pequeños” de varios seres vivientes (cuadro 11).

The screenshot shows the same web browser window with the search parameters changed to Lema: 'crío', Forma: 'crío*', and Clase de palabra: 'sustantivo'. The results show 2852 cases in 1251 documents. The table lists results with columns for REF. (Classification, country), the concordance text, and the word 'crío' highlighted in yellow. The table has columns for REF. (Clasificación, país), CONCORDANCIA, and Ordenar por: Año ascendente. The bottom of the browser shows the taskbar with icons for Windows, Firefox, and other applications, and the system clock indicating 17:09 on 20/06/2014.

Cuadro 11. Resultados de la búsqueda de las concordancias del lema ‘crío’ en el *Corpes XXI*

Este tipo de actividades colocan al aprendiente en el papel de investigador y analista del idioma y le permiten familiarizarse con el uso de los corpus como herramientas de investigación lingüística, además de promover el desarrollo de la autonomía y estimular la curiosidad por explorar las distintas manifestaciones del español en una dimensión panhispánica.

Conclusiones

En este trabajo se han querido enseñar las potencialidades de algunos corpus como recursos para sensibilizar a los aprendientes de ELE acerca de las variedades diatópicas del español. El objetivo de las actividades que se proponen no es que los alumnos aprendan a utilizar uno u otro geolecto sino que tomen conciencia de la riqueza de matices que conforma el mosaico de la inmensa comunidad hispanohablante. No se trata de promover la atención a los regionalismos, sino más bien de llevar al aula algunos principios básicos del panhispanismo, que valoriza la diversidad en la unidad del idioma: desde esta perspectiva la forma en que cada comunidad de habla hispana se expresa, a través de una única lengua, intercomprensible, vehicula una parte de la(s) cultura(s) que en ella se identifican. Si queremos que nuestros alumnos sepan relacionarse de manera respetuosa con las distintas culturas hispanohablantes, es importante que aprendan a reconocer y apreciar sus geolectos. Los profesores –nativos y no nativos–, solos, no podemos con este reto. Los corpus no son la panacea pero son un recurso valioso, una herramienta más para llevar a cabo aquella didáctica panhispánica muy anhelada y todavía poco practicada.

Bibliografía

- Aijmer, Karin (ed.) (2009): *Corpora and language teaching*. Amsterdam/Philadelphia: John Benjamins.
- Aston, Guy (2001): *Learning with Corpora*. Bolonia: CLUEB.
- Baralo, Marta (2003): "Mestizaje e interculturalidad en la variación diatópica y su incidencia en el español/LE". En Pérez Gutiérrez, Manuel; Coloma Maestre, José (eds.): *El español, lengua del mestizaje y la interculturalidad. Actas del XIII Congreso Internacional de ASELE*. Murcia: ASELE, pp. 152-164.
- Bernardini, Silvia (2001): 'Spoilt for choice': a learner explores general language corpora". En Aston, Guy (ed.): *Learning with Corpora*. Bologna: CLUEP, pp. 220-249.

- (2004): "Corpora in the classroom. An overview and some reflections on future developments". En Sinclair, John M.: *How to use corpora in Language Teaching*. Amsterdam: John Benjamin Publishing, pp. 15-36.
- Boulton, Alex & Henri Tyne (2013): "Corpus linguistics and data-driven learning: a critical overview". En *Bulletin VALS-ASLA* 97, pp. 97-118.
- Chambers, Angela (2010): "What is data-driven learning?". En O'Keeffe, Anne; McCarthy, Michael (eds.): *The Routledge Handbook of Corpus Linguistics*. Routledge: London/New York, pp. 345-358.
- Cruz Piñol, Mar (2012): *Lingüística de corpus y enseñanza del español como 2/L*. Madrid: Arco Libros.
- Gabrielatos, Costas (2005): "Corpora and Language Teaching: Just a fling or wedding bells?". En *Teaching Language as a Second or Foreign Language*, 8, 4, 1-35. Accesible desde <<http://www.tesl-ej.org/ej32/a1.html>> 15/09/2014.
- Gavioli, Laura (2001): "The learner as a researcher: introducing corpus concordancing in the classroom". En Aston, Guy (ed.): *Learning with Corpora*. Bologna: CLUEP, pp. 108-137.
- Gilquin Gaëtanelle & Sylviane Granger (2010): "How can data-driven learning be used in language teaching?". En O'Keeffe, Anne; McCarthy, Michael (eds.): *The Routledge Handbook of Corpus Linguistics*. Routledge: London/New York, pp. 359-370.
- Granath, Solveig (2009): "Who benefits from learning how to use corpora?". En Aijmer, Karin (ed.): *Corpora and language teaching*. Amsterdam/Philadelphia: John Benjamins, pp. 47-65.
- Johns, Tim (1991a): "Should you be persuaded: Two examples of data-driven learning". En *ELR Journal*, 4, pp. 1-16.
- (1991b): "From printout to handout: Grammar and vocabulary learning in the context of data-driven learning". En *ELR Journal*, 4, pp. 27-45.
- (2000): *Tim Johns Data-driven Learning Page*. Accesible desde <http://archive.ecml.at/projects/voll/our_resources/graz_2002/ddrivenlrning/whatisddl/resources/tim_ddl_learning_page.htm> 15/09/2014.
- Leech, Geoffrey (2002): "Sobre la importancia de los corpus de referencia". En *Donosít*, 24-25, pp. 1-3.
- Moreno Fernández, Francisco (2000): *Qué español enseñar*. Madrid: Arco Libros.
- (2005): "Corpus para el estudio del español en su variación geográfica y social: el corpus PRESEEA". En *Oralia: Análisis del discurso oral*, 8, pp. 123-140.
- (2006): "El español en su variación geográfica y social: informe sobre el corpus PRESEEA". En Cestero Mancera Ana M., Molina Martos, Isabel Paredes García, Francisco (eds.): *Estudios sociolingüísticos del español de España y América*. Madrid: Arco Libros, pp. 15-34.
- (2010): *Las variedades de la lengua española y su enseñanza*. Madrid: Arco Libros.
- O'Keeffe, Anne & Michael McCarthy (2010): *The Routledge handbook of corpus linguistics*. Routledge: London/New York.
- Römer, Ute (2009): "Corpus research and practice: What help do teachers need and what can we offer?". En: Aijmer, Karin (ed.) *Corpora and Language Teaching*. Amsterdam/Philadelphia: John Benjamins Publishing Company, pp.83-98
- Sinclair, John (2004): *Developing Linguistic Corpora: a Guide to Good Practice Corpus and Text — Basic Principles*. Accesible desde: <<http://www.ahds.ac.uk/creating/guides/linguistic-corpora/chapter1.htm#ection12>> 15/09/2014.

¹ *Atlas interactivo de la entonación en español*, <http://prosodia.upf.edu/atlasentonacion/index.html> [Fecha de consulta: 21/06/2014].

² Los tipos de enunciados evocados son: declarativos, interrogativos (absolutos, parciales y reiterativos), ruegos y órdenes, vocativos. El corpus consta de 70 enunciados por cada uno de los 17 geolectos.

³ Los módulos temáticos se refieren a: saludos, el tiempo, el lugar donde vive el informante, su familia y amistades, costumbres de su país, anécdotas importantes de su vida y aspiraciones y deseos de mejora económica.

⁴ Datos extraídos de la página web <http://preseea.linguas.net/> [Fecha de consulta: 05/09/2014].

⁵ *Corpes XXI*, <http://www.rae.es/recursos/banco-de-datos/corpes-xxi> [Fecha de consulta: 21/06/2014].

⁶ *Corpes XXI*, http://www.rae.es/sites/default/files/CORPES_Sistema_de_codificacion2014.pdf, p. 8 [Fecha de consulta: 21/06/2014].

⁷ Audio 1: <http://prosodia.upf.edu/atlasentonacion/cerca/index.php> Variedad peninsular (Madrid, 5); edad participantes: 25 y 26 años; profesión: administrativas, duración: 3 minutos 53' [Fecha de consulta: 21/06/2014].

Audio 2: <http://prosodia.upf.edu/atlasentonacion/cerca/index.php> - Variedad peninsular (Jerez de la Frontera, 1); edad participantes: 41 y 46 años; profesión: empresarios; duración: 3 minutos 15' [Fecha de consulta: 21/06/2014].

⁸ <http://prosodia.upf.edu/atlasentonacion/cerca/index.php> - Enunciado: orden - Dialecto: Bogotá; Buenos Aires; Las Palmas de Gran Canaria; Madrid; México DF [Fecha de consulta: 05/09/2014].

⁹ En este caso el uso de usted para las relaciones informales en el español andino; el uso del voseo en la variante rioplatense; aquí/acá; ahora/ahorita.

¹⁰ http://www.uv.es/uvweb/departamento_filologia_espanola/es/investigacion/grupos- - Grupo de Investigación VAL.ES.CO.

¹¹ Grupo 1: de 20 a 34 años; Grupo 2: de 35 a 54 años; Grupo 3: de 55 años en adelante [Fecha de consulta: 20/09/2014].

¹² *Corpus PRESEEA*, file:///C:/Users/Sonia/Downloads/METODOLOG%C3%8DA%20PRESEEA.pdf [Fecha de consulta: 14/06/2014].

¹³ *Ibidem*.

¹⁴ <http://web.frl.es/CORPES/view/inicioExterno.view> [Fecha de consulta: 10/06/2014].