

Aprendizaje basado en datos en español como lengua extranjera: ampliando el ámbito

MAGDALENA ABAD CASTELLÓ

Instituto Cervantes Mánchester

Magdalena.Abad@cervantes.es

ANTONIO ÁLVAREZ BAZ

Universidad de Granada

antxon@ugr.es

Resumen

A pesar de todas las ventajas que ofrecen los corpus lingüísticos y de los avances tecnológicos que han mejorado su accesibilidad y manejo, su uso no se ha extendido entre alumnos y profesores de lenguas extranjeras. Aprendizaje basado en datos (ABD) es un enfoque centrado en el alumno que promueve el aprendizaje por descubrimiento mediante el análisis de datos provenientes de corpus lingüísticos. En el campo del español como lengua extranjera (ELE) este tipo de aprendizaje sigue siendo un gran desconocido. Este artículo quiere mostrar cómo los principios de ABD se pueden aplicar a exploraciones sobre datos provenientes de otras herramientas de consulta y de la propia Web. En este trabajo se presentan en un primer momento los principios teóricos que apuntalan este enfoque y los retos a los que se enfrenta. A continuación, se revisa el estado de la cuestión en ELE, con un ejemplo de aplicación del enfoque utilizando dos corpus lingüísticos: CORPES XXI y el Corpus del Español. Para finalizar, se examinan nuevos recursos híbridos (diccionario más listados de ejemplos) en español con datos que pueden ser investigados mediante este tipo de enfoque y se ejemplifica con una secuencia de búsqueda en la que se utilizan dos de estas herramientas: Linguee y WordAndPhrase.

Palabras clave

Aprendizaje basado en datos (ABD); español como lengua extranjera (ELE); corpus lingüísticos; recursos de consulta.

Data-Driven Learning in Spanish as a Foreign Language: Expanding the Scope

Abstract

Despite all its benefits and current technological advances, the use of corpora is not widespread among learners and educators of foreign languages. Data-driven learning (DDL) is a learner-centred approach that promotes discovery learning by analysing data from linguistic corpora. In the field of Spanish as a foreign language (SFL) this kind of learning is still a great unknown. This paper seeks to show how DDL principles can be applied to queries on data coming from other search tools including the Web. First, the article presents the theories that underpin this approach and the challenges this methodology faces nowadays. Next, we provide an overview of the current state of DDL in Spanish, and we show an example of implementation using two corpora: CORPES XXI and Corpus del Español. Finally, the paper looks into new hybrid resources (on-line dictionaries plus lists of examples) in Spanish containing data that can be examined using this approach and we show an example of search using two of these tools: Linguee and WordAndPhrase.

Key words

Data-Driven Learning (DDL); Spanish as foreign language (SFL); corpora; reference resources.

Recibido el 02/02/2021

Aceptado el 10/11/2021

1. Introducción

Aprendizaje basado en datos (ABD) es un enfoque de aprendizaje de lenguas en el que el alumno investiga datos provenientes de corpus lingüísticos y descubre patrones para obtener respuestas sobre diferentes aspectos lingüísticos y no lingüísticos de estos datos. Aprendizaje basado en datos es la traducción (Asención-Delaney, Collentine, Collentine, Colmenares, y Plonsky, 2015) de Data Driven Learning (DDL), término acuñado por Tim Johns, uno de los primeros autores en percibir el potencial de los corpus lingüísticos para los aprendientes de lenguas y considerado el padre de este enfoque. Johns lo define como «the attempt to cut out the middleman as far as possible and to give the learner direct access to the data» (Johns, 1991: 30). Tal y como afirman Pérez-Paredes y Zapata-Ros (2018: 8), «mediante ABD, los estudiantes pueden acceder a un conocimiento sobre el uso de la L2 que, por diversos motivos, no suele formar parte del currículo oficial en la enseñanza de lenguas».

Son muchas las ventajas que ofrece este enfoque. Los corpus lingüísticos ofrecen datos muy útiles entre los que destacamos los numerosos ejemplos genuinos de lengua y sus contextos y los datos estadísticos de frecuencia tanto general como según periodo, género textual o región. Por otro lado, en este aprendizaje por descubrimiento, las actividades realizadas con estos datos requieren un proceso de reflexión y razonamiento que para O'Sullivan (2007) se traduce en destrezas cognitivas: entre otras, captar, analizar, hacer inferencias (inductivas o deductivas), reflexionar, interpretar, teorizar y verificar. Siguiendo a Balunda (2009) y Boulton (2009), ABD se adhiere al enfoque comunicativo ya que fomenta el aprendizaje inductivo de lengua en contexto, la enseñanza centrada en el alumno y la autonomía en el aprendizaje, puesto que las actividades con ABD capacitan para trabajar de manera autónoma.

Los estudios empíricos sobre DDL han florecido en los últimos años y se han sucedido revisiones y metaanálisis (Pérez-Paredes, 2019; Lee, Warschauer, y Lee, 2019; Boulton y Cobb, 2017) que, en su mayoría, han probado su eficacia para el aprendizaje de lenguas tanto en su uso directo de corpus como en el indirecto, en el que el profesor crea materiales con datos extraídos de los corpus. Entre los numerosos objetos de las investigaciones los más numerosos son los estudios sobre combinatoria léxica, patrones gramaticales y léxico-gramaticales, frecuencia, y competencia escrita. Los estudios empíricos se han realizado mayormente en el campo del inglés como lengua extranjera. En español como lengua extranjera los estudios empíricos son mucho más escasos (Benavides, 2015; Yao, 2019; Marcos Miguel, 2020).

Paralelamente, tal y como muestra Buyse (2017), los avances en la tecnología hacen más fácil la utilización de los corpus de referencia como el Corpus del Español del Siglo XXI (CORPES XXI)¹ y el Corpus del Español de Mark Davies (CdeIE)². Con todo, al mismo tiempo han ido surgiendo recursos mucho más accesibles y manejables que usan herramientas que combinan las prestaciones de los diccionarios con listados de ejemplos similares a las concordancias de los corpus (Linguee, Reverso, SpanishDict). Asimismo, además del propio Google, se puede utilizar WebCorp, un conjunto de herramientas de búsqueda que permite

¹ <https://www.rae.es/banco-de-datos/corpes-xxi>

² <https://www.corpusdelespanol.org/>

usar la Web como corpus. Por último, disponemos de WordAndPhrase, que es también un recurso híbrido pues, además de proporcionar definiciones, sinónimos y antónimos como un diccionario, ofrece líneas de concordancia e información sobre frecuencia, género textual y región. Todas estas herramientas son más fáciles de usar que los grandes corpus de referencia y proporcionan también datos analizables siguiendo el enfoque ABD.

En este artículo proponemos que ABD no se restrinja a usar los datos de los corpus, como tradicionalmente se ha venido haciendo, sino que también se aplique a datos provenientes de diferentes recursos y herramientas como las descritas anteriormente. ABD ha llegado tarde al campo de ELE, por lo que para ganar la aceptación de docentes y estudiantes debe adaptarse constantemente a los últimos avances tecnológicos.

2. Aprendizaje basado en datos: fundamentación, aplicación y desafíos

2.1. La fundamentación teórica del Aprendizaje basado en datos

Siguiendo a Flowerdew (2015), son tres las teorías que fundamentan el aprendizaje basado en datos: la hipótesis de la captación, el aprendizaje constructivista y las teorías socioculturales de Vygostky.

Según la «hipótesis de la captación» (Noticing Hypothesis), de Schmidt (1990) para que se produzca adquisición se tienen que percibir en primer lugar las formas lingüísticas. Lo que el estudiante, intencionadamente o no, observa y capta en el *input* es lo que se convierte en *intake*, una condición ineludible para el aprendizaje. El *input* está constituido por todas las muestras de lengua a las que está expuesto el estudiante tanto en el aula como en su aprendizaje autónomo y el *intake* es el elemento lingüístico adquirido en el proceso de aprendizaje. A partir de esa primera percepción el aprendiente reflexiona sobre la misma esforzándose por entender su significado hasta alcanzar su conocimiento. Esta teoría ha sido criticada por considerar la adquisición como un proceso consciente, cuando Krashen (1988) dejó claro que existen muchos factores en la lengua que no se pueden aprender de una manera consciente. No obstante, Schmidt (1990) incide en la importancia de esta toma de conciencia lingüística para el procesamiento efectivo y la adquisición de la lengua. El ABD (Sepehri, 2015) puede incrementar la prominencia de las formas meta mediante el formato de las líneas de concordancia (con la palabra destacada en el centro y en negrita) y con la repetición de las palabras clave en múltiples líneas de concordancia.

Flowerdew (2015) define el constructivismo como una teoría educacional que percibe la adquisición del conocimiento como un proceso dinámico en el que los estudiantes son los conductores de su aprendizaje. Este enfoque promueve la participación activa del alumno, que construye su conocimiento conectando la información nueva con su conocimiento previo mediante diferentes procesos cognitivos como hacer hipótesis e inferencias. El constructivismo apuntala claramente muchas de las actividades del ABD. El alumno tiene un papel activo al consultar directamente los corpus y extraer patrones y, a partir de estos, formular sus propias reglas. El estudiante se implica y gradualmente se vuelve más autónomo en su aprendizaje. Por ende, el profesor no es la fuente de conocimiento sino un facilitador que ayuda a los estudiantes a interpretar los datos y les da consejos acerca de cómo buscar la información y cómo analizar los resultados.

El proceso de construcción del conocimiento puede exigir un esfuerzo enorme al alumno, por lo que va a necesitar ayuda y esta la puede encontrar en el trabajo grupal con sus compañeros. Para Vygotsky (1978) el conocimiento se construye a través de la interacción entre aprendientes, un diálogo colaborativo en la búsqueda de conocimiento. Es lo que también se denomina «andamiaje» pues es guiado o apoyado por el profesor o por los propios alumnos. El trabajo colaborativo es una práctica normalizada en las clases de lengua y el andamiaje se percibe como un factor clave en el aprendizaje. Tal y como sostiene Bernardini (2016), si se fomentan las actividades de investigación grupal para solucionar problemas lingüísticos se crea un contexto social, colaborativo, en el que se construye el conocimiento. De igual modo, todas las actividades de investigación de datos (de corpus lingüísticos o de otros recursos) son susceptibles para la práctica grupal. Gracias a este contexto social, colaborativo, estas actividades pueden, de hecho, generar más conocimiento que las actividades individuales.

El proceso de investigación de las muestras de lengua se ha ido refinando durante los últimos años. McCarthy y Carter (1995) proponían una secuenciación a la que denominaban las tres íes: ilustración (examinar datos reales), interacción (discutir y compartir opiniones y observaciones) e inducción (el estudiante crea sus propias reglas). Kennedy y Micelli (2001) plantean una estrategia de búsqueda con cuatro pasos: formular la pregunta de indagación, concebir una estrategia de búsqueda, observar los ejemplos y seleccionar los más relevantes, y sacar conclusiones. Pérez Paredes et al. (2011) proponen a su vez un proceso de búsqueda en cuatro etapas: inicio de la búsqueda, fase de interpretación, fase de consolidación de la hipótesis y fase de formulación final de los hallazgos realizados. La diferencia con respecto a la secuencia de Kennedy y Micelli radica en el tercer paso, cuando el estudiante, después de formular una primera hipótesis, vuelve al corpus a buscar más datos para sustentar su primera interpretación. Si esta hipótesis se confirma, se llega a una conclusión final, de la que se informa al resto del grupo. La figura 1 siguiente muestra esta secuenciación según Sinclair (2003) y Pérez-Paredes et al. (2011).

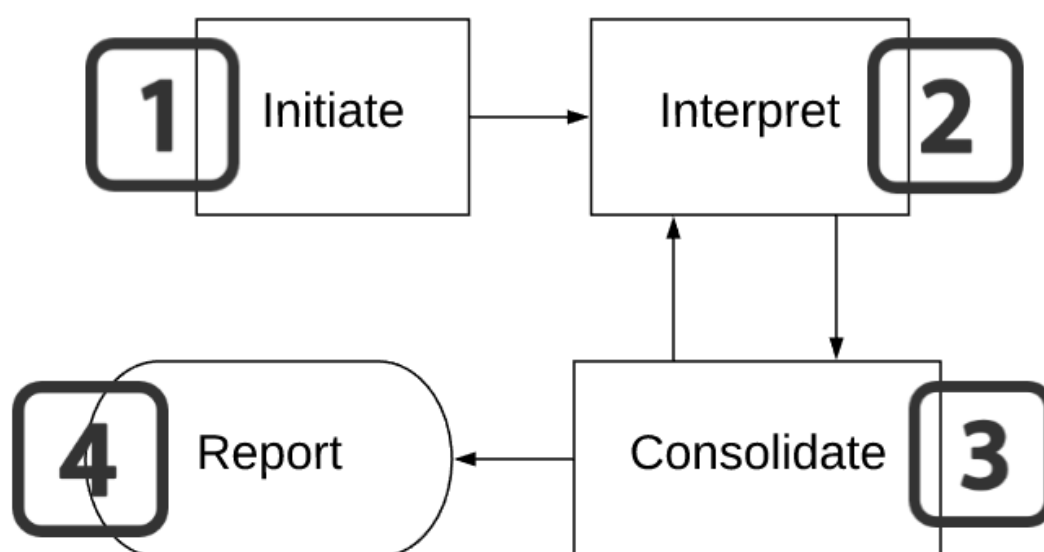


Figura 1. Proceso de búsqueda en ABD. Sinclair (2003) y Pérez-Paredes et al. (2011)

2.2. Desafíos del Aprendizaje Basado en Datos

El uso de corpus lingüísticos en la enseñanza o aprendizaje de lenguas plantea retos importantes para estudiantes y profesores. Ante todo, docentes y alumnos necesitan una formación especializada, tanto en destrezas técnicas como analíticas, para poder percibir mucha de la información que puede ofrecer un corpus. Asención-Delaney et al. (2015: 142) afirman a este respecto: «Una actividad del ABD debe lograr dirigir al aprendiz hacia el vocablo, lo cual requiere un nivel alto de recursos cognitivos de atención y consciencia metalingüística de las propiedades morfológicas del vocablo meta». Boulton resume muy acertadamente las dificultades que el ABD presenta a los estudiantes.

It is unsurprising that learners find it difficult to get to grips with new material (the corpora), new technology (the software) and a new approach (DDL) all at once – especially at lower levels of language ability. (Boulton, 2008: 39)

Por lo tanto, con esta metodología los estudiantes se enfrentan a un triple desafío. En primer lugar, el estudiante necesita familiarizarse con un material muy diferente: líneas de concordancia, listados de frecuencia o estadísticas. En segundo lugar, el propio corpus (y su sistema de búsqueda) es una nueva tecnología para la que se requiere entrenamiento. Por último, la nueva metodología, que otorga un papel central y activo al estudiante, puede ser un reto, sobre todo para los estudiantes con una creencia tradicional de la enseñanza y el aprendizaje, en el que el profesor es la figura central y de autoridad.

Aunque la interfaz de los corpus se ha simplificado y es mucho más fácil de usar, las búsquedas en corpus lingüísticos siguen llevando más tiempo que las búsquedas en otras herramientas de consulta como son los diccionarios en línea. El número de ejemplos y la complejidad lingüística de muchos de ellos plantean otro gran reto para los aprendientes. La otra gran limitación es el dispositivo, ya que la mayoría de estudiantes hoy día utiliza móviles inteligentes para acceder a la Web y a los diccionarios. Aunque se pueda acceder a ellos, la búsqueda en corpus con móviles o tabletas es tediosa si se compara con otras aplicaciones; parece necesario crear aplicaciones adaptadas a los dispositivos móviles. Paredes et al. (2019) diseñaron una aplicación móvil que probaron con tres cohortes de estudiantes universitarios de inglés como lengua extranjera en Reino Unido, Bélgica y España. Tras probar la aplicación, los estudiantes pusieron de manifiesto sus ventajas como la información proporcionada, el *feedback* inmediato y el acceso de 24 horas. Por otro lado, aunque la tecnología funcionó, los alumnos tuvieron dificultades a la hora de interpretar los resultados. Estos datos confirman la necesidad de entrenar a los estudiantes en habilidades de análisis y formulación de hipótesis.

3. Aprendizaje basado en datos en español: ampliar el ámbito

3.1. Estudios sobre ABD en ELE

En el campo del español como L2 o LE abundan las experiencias docentes sobre usos pedagógicos de los corpus lingüísticos, descritas en artículos, actas de encuentros o congresos, libros o tesinas. Si bien algunas de ellas muestran un gran potencial para futuras

investigaciones (Paz, 2013; Uriel Domínguez, 2015; Rodríguez Castellano, 2017; Peláez Pino, 2019), la mayoría se queda en propuestas o ideas sin resultados experimentales.

Por lo que se refiere a estudios empíricos en ELE, son muy escasos y sobre campos y temáticas muy diferentes. Buyse y Verlinde (2013) compararon en un estudio a pequeña escala los usos de los diccionarios tradicionales con el uso de Linguee. Los estudiantes que usaron datos de corpus tuvieron mejores resultados, particularmente en el apartado de vocabulario. Los estudiantes que usaron varias herramientas mostraron un gran nivel de satisfacción con la calidad de las herramientas, aunque no con el tiempo que estas demandaban. Los estudiantes de Benavides (2015) realizaron búsquedas en el corpus de concordancias con ocurrencias de unidades léxicas y con ciertos tipos de colocaciones. El objetivo era comprender mejor varios contenidos gramaticales: usos del imperfecto y pretérito indefinido (o perfecto simple) y usos de ser y estar. Aunque los resultados cuantitativos no mostraron una medida clara de mejora en el aprendizaje, los resultados cualitativos proporcionaron evidencia de que el corpus era una herramienta de aprendizaje eficaz. Yao (2019) comparó el uso de materiales ABD con definiciones de diccionarios en el aprendizaje de vocabulario. Los resultados del grupo experimental fueron significativamente mejores que los del grupo de control, mientras que las respuestas de los cuestionarios mostraron que los estudiantes favorecían el ABD y mostraban una actitud positiva hacia una futura aplicación de este método. La investigación de Contreras Izquierdo (2019) se centró en la percepción por parte de los estudiantes del uso de recursos lexicográficos para estudiar las variedades del español. A pesar de que los trabajos de los estudiantes fueron de alta calidad, se observó un uso reducido de los corpus lingüísticos en comparación con los otros recursos. El estudio más reciente sobre ABD en ELE a la hora de la publicación de este artículo es el de Marcos Miguel (2020), quien analizó las interacciones de sus estudiantes cuando realizaban tareas que requerían el uso de herramientas de corpus. Los estudiantes experimentaron algunas dificultades a la hora de formular hipótesis, pero la investigadora concluyó que las tareas basadas en corpus facilitan la exploración y el uso de la lengua.

3.2. Ejemplo de aplicación didáctica de ABD en ELE

Veamos un ejemplo práctico de actividad de investigación de ABD en ELE usando los dos corpus lingüísticos más importantes del español: CORPES XXI y C.del.E. En una actividad lúdica, un concurso, un grupo de estudiantes B2 (según MCER) está haciendo hipótesis sobre la edad de unos famosos. Para ello, la profesora les ha recomendado cuatro frases con las que deberían estar familiarizados pues se introducen en el nivel B1 (*Puede que, Es probable que, Quizás y A lo mejor*).

Los estudiantes no usan la estructura «A lo mejor» y, al preguntarles por la razón, muestran su inseguridad sobre su uso. La profesora les propone entonces una pequeña investigación. En pequeños grupos pequeños de 2 o 3 estudiantes realizan tres búsquedas diferentes. En primer lugar, los estudiantes buscan la frase en un diccionario en línea, WordReference.

The screenshot shows the WordReference dictionary entry for the Spanish phrase "a lo mejor". At the top, there is a search bar with the text "a lo mejor" and a speaker icon labeled "ESCUCHAR:" followed by a dropdown menu set to "MÉXICO". Below the search bar are links for "Spanish definition", "Spanish synonyms", "Gramática", "Conjugación [ES]", "Conjugator [EN]", "in context", and "images". There are two tabs: "WordReference" (selected) and "WR Reverse (3)". Below the tabs, it says "WordReference English-Spanish Dictionary © 2021:". The main section is titled "Principal Translations" and is divided into two columns: "Spanish" and "English".

Spanish	English
a lo mejor <i>loc adv</i>	(tal vez, es posible) (indicative) maybe, perhaps <i>adv</i> might ⇒, may ⇒ <i>v aux</i>
	Si mañana hace buen tiempo a lo mejor iré a nadar. En otras circunstancias a lo mejor hubiera dicho la verdad. Maybe I'll go swimming tomorrow if the weather's OK. ⓘ I might (or: may) go for a walk later if I don't feel too tired.
a lo mejor, a lo mejor se haga algo <i>loc adv</i>	<i>colloquial</i> (podría ser que) (subjunctive) maybe, perhaps <i>adv</i> might ⇒, may ⇒ <i>v aux</i>
Note: Se construye con subjuntivo.	
En otras circunstancias a lo mejor hubiera dicho la verdad. Under other circumstances maybe (or: perhaps) he would have told the truth.	

Figura 2. Búsqueda de «a lo mejor» en Word Reference

La búsqueda les ayuda parcialmente a comprender el significado de la expresión, pero les confunde en el uso de indicativo y subjuntivo y tampoco están seguros de cómo poder usarla en contexto. Realizan entonces una búsqueda de concordancias en el Corpus del Siglo XXI (CORPES XXI) de la Real Academia Española. Para reducir el número de resultados y para centrarse en la variedad que más se usa en el aula, los alumnos restringen la búsqueda a España.

5.335 casos en 2.192 documentos.

EF. (Clasificación, país)	CONCORDANCIA	Ordenar por: Año ascendente	sin criterio
2001 Esp.	juego de hombres de pura raza. Tienes que confiar en Surcos. Mira, si no me fallas	a lo mejor	te licencio. Y hasta te rapo. Todos hemos pasado por esto alguna vez. A ver, déja
2001 Esp.	-¡Baja, tonta! Vamos a buscar los gusanos de seda,	a lo mejor	ya han terminado de tejer los capullos y ya podemos sacar el hilo para hacer los
2001 Esp.	Anduve cabizbajo hasta la playa -pensaba en mis gusanos de seda que	a lo mejor	estaban retorciéndose sobre el calor de las planchas-, despacio porque el miedo
2001 Esp.	-Dicen en el pueblo que me tiene que ver un médico -dice sin inmutarse-, que	a lo mejor	es por el cambio de clima, al venir yo de tierras más calientes. Más que nada por
2001 Esp.	-Pero	a lo mejor	necesita ayuda. A lo mejor necesita beber un poco de leche para luego amamantar
2001 Esp.	-Pero a lo mejor necesita ayuda.	A lo mejor	necesita beber un poco de leche para luego amamantar a sus crías -y quedó mirando
2001 Esp.	alojado en una casita en la cima, próxima al monasterio de las monjas, en el frío...	a lo mejor	es que usted también vive por ahí.
2001 Esp.	le queda mucho, y Alfredo que sí que le quedaba, le queda más de lo que supones,	a lo mejor	en otoño podemos llevarlo, díselo, y Tomás que ya se lo había dicho, que llevaba
2001 Esp.	, añadido con un gesto de asco, no me voy a presentar en el casino, y Tomás papá,	a lo mejor	el casino ya no existe, quizá lo han cerrado, pero el padre negó con la cabeza,
2001 Esp.	hacerse respetar, y usted, señorito, solo desde los quince años, otro en su lugar	a lo mejor	había salido un poco afeminado, o más blando, aunque muchos huérfanos salen duros
2001 Esp.	marido y yo con mi mujer, cada uno en su sitio, ya estará casi tan vieja como yo y	a lo mejor	hasta se arrepiente, se quedó viuda muy joven, treinta y uno o treinta y dos debía
2001 Esp.	Alfredo, lo mismo hasta llegué a ir a la escuela, aunque yo era más niño, claro, pero	a lo mejor	cuando empecé a ir a la escuela ustedes ya se habían ido del pueblo, no lo sé,
2001 Esp.	reparaciones y la baja de la mujer del otro coche, tiene para dos meses más o menos, y	a lo mejor	una indemnización para la prostituta, la chica quiso marcharse, me figuro que no
2001 Esp.	meses, ya me inventaré alguna explicación, con ese dinero, cuando salga el juicio,	a lo mejor	me llega para pagar, estas cosas van lentas y, además, con un recurso se puede
2001 Esp.	más Guzmán que si hubiera sido una Guzmán legítima, con un padre como Dios manda.	A lo mejor	el granuja era de verdad un estudiante de derecho y hoy es ya un abogado famoso
2001 Esp.	los tíos se han ido al pueblo, se fueron ayer y volverán mañana, la próxima vez	a lo mejor	puedes ir tú también, seguro que te gustaría volver al pueblo, abuela, yo no lo
2001 Esp.	rechacé el ofrecimiento, me las tendió un buen rato, lleváte las si quieres, me dijo,	a lo mejor	algún día te sirven para escribir la historia de tu familia, ya que has salido
2001 Esp.	he preguntado muchas veces en estos años, por qué no tuve valor para hacerlo, pero	a lo mejor	tu madre también se echó para atrás en el último momento, vivía sometida por la
2001 Esp.	cambio otra vez, mira, ya llevo ocho, y si le doy con tino las voy a dejar muy bien,	a lo mejor	hasta hago una tacada. Ya está bien, no hay quien te aguante, Tomás, termina el
2001 Esp.	no volvió más, y cuando me enteré de lo que me ocurría él ya había desaparecido,	a lo mejor	solo se había ido a estudiar a Sevilla, estudiaba para abogado en Sevilla, eso

Figura 3. Líneas de concordancia de «a lo mejor» en CORPES XXI

Al examinar varias pantallas como la anterior, los estudiantes obtienen un patrón frecuente, donde observan que esta expresión va seguida de indicativo. Al mismo tiempo aprenden que es de uso frecuente, al menos en España. Además, las líneas muestran un contexto cotidiano, por los temas y por el frecuente uso de la segunda persona del singular. Tras esta búsqueda y primera interpretación, realizan una segunda exploración para ver si se confirma la mayor frecuencia en España y el uso en contextos cotidianos. Llevan a cabo una búsqueda en Estadísticas, otra función del CORPES XXI. Tal y como se aprecia en los gráficos siguientes, «a lo mejor» se utiliza mucho más en España que en América Latina.

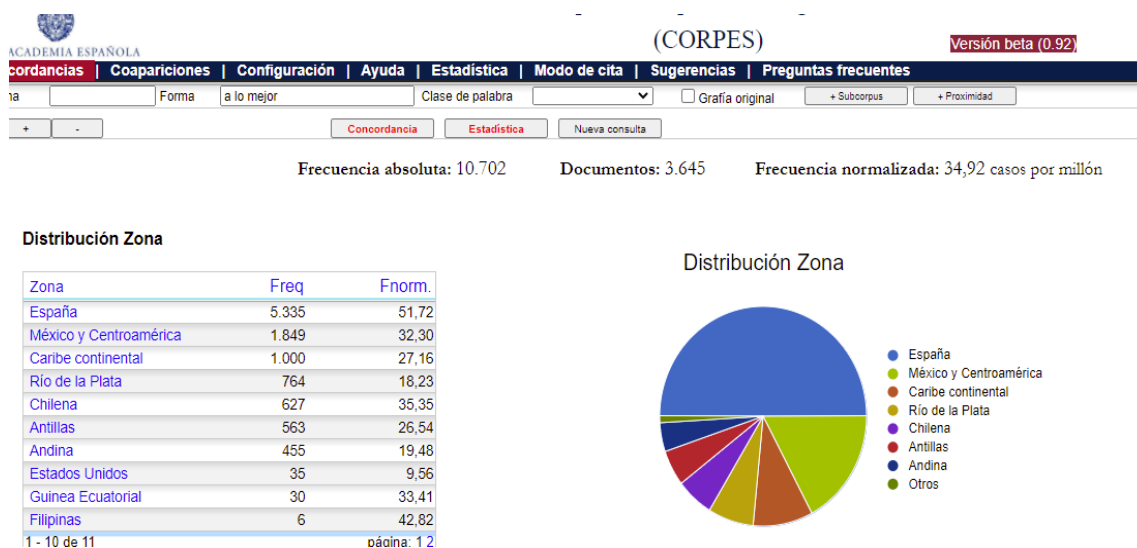


Figura 4. Estadísticas de «A lo mejor», país, CORPES XXI

Otro dato que puede ser muy útil es que esta expresión se usa mucho más en el contexto cotidiano, en conversaciones.

Distribución Tipología

Tipología	Freq	Fnorm
Ficción	7.082	82,08
Entrevista	1.208	229,65
Divulgación	216	8,65
Noticia	214	7,62
Blog	157	31,68
Otros	149	50,37
Académico	93	3,73
Reportaje	84	7,66
Opinión	75	19,99
Biografía memoria	64	32,31

1 - 10 de 24 página: 1 2 3

Distribución Tipología

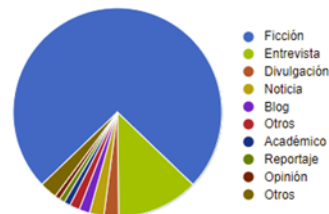


Figura 5. Estadísticas de A lo mejor, género textual, CORPES XXI

Por último, y para comparar la frecuencia entre Indicativo y Subjuntivo, pueden realizar una última búsqueda, esta vez en el Corpus del Español (CdeE).

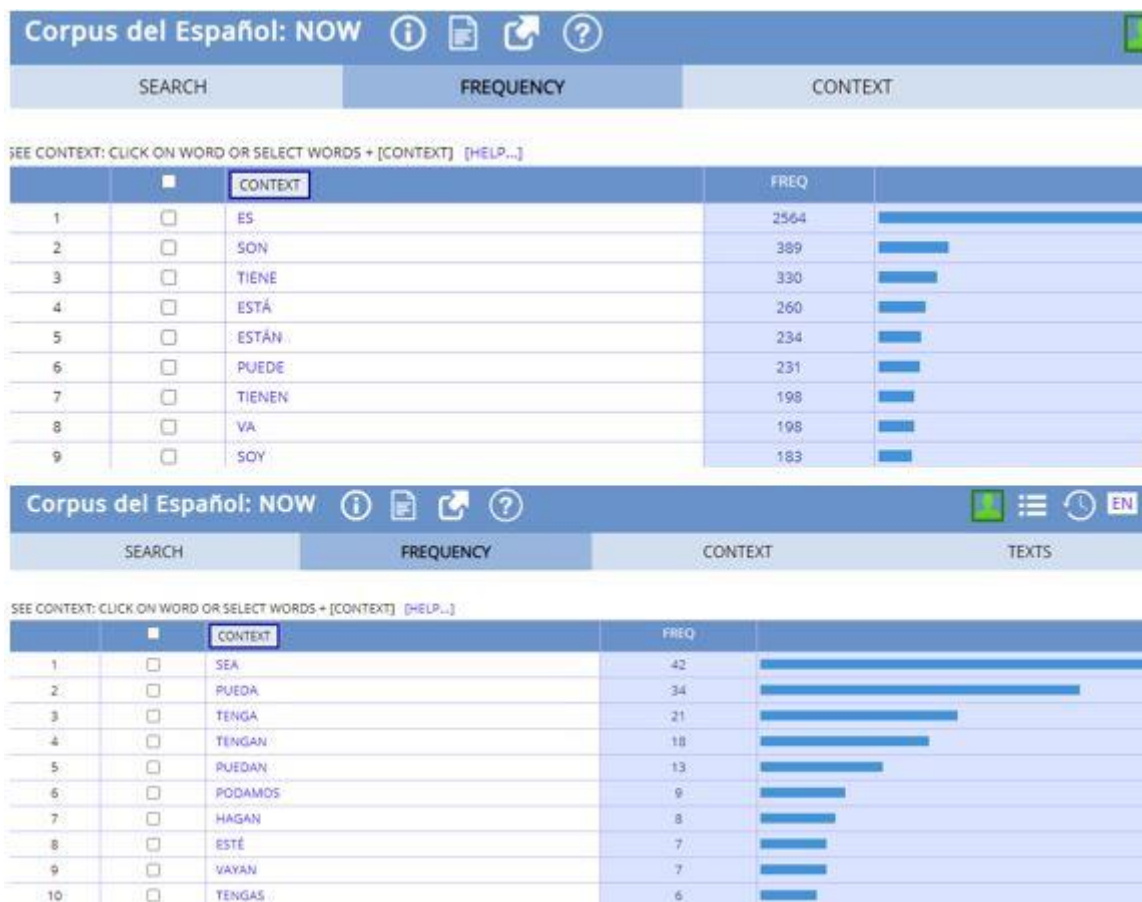


Figura 6. Búsqueda de «a lo mejor», modo verbal. Corpus del Español

Esta búsqueda confirma que la expresión se utiliza mayormente con Indicativo, (2 564 resultados por 42 resultados con Subjuntivo).

De este modo, se confirman las hipótesis de manera que los estudiantes pueden sacar una conclusión y llevarla a la clase. Entre todos, pueden construir la siguiente regla.

«A lo mejor» se usa para expresar hipótesis y conjeturas. Va seguido normalmente de Indicativo y es mucho más frecuente en contextos cotidianos orales. Es mucho más frecuente en España que en América Latina.

3.3. Nuevos recursos en línea para el español

Hemos visto como la investigación empírica en ELE sobre ABD no ha acabado de despegar mientras que las tecnologías siguen cambiando al mismo tiempo que las necesidades y preferencias de los estudiantes. En el apartado anterior se ha expuesto el desafío que entraña para los estudiantes la búsqueda en corpus y su preferencia por el uso de dispositivos móviles. Parece necesario, por tanto, investigar herramientas más accesibles y de fácil interfaz, pero sin descuidar el entrenamiento en habilidades cognitivas.

La cuestión de la consideración de la Web como corpus es motivo de controversia entre los investigadores de lingüística, desde la negación «the World Wide Web is not a corpus» de Sinclair (2005: 21) a la aceptación de otros como Kilgarriff y Grefenstette (2003). Ante las críticas por la falta de anotación de la Web, Boulton (2012) argumenta que tampoco lo están otros corpus y, si bien esta falta de anotación limita algunos tipos de investigación, esto no significa que sea imposible. Por supuesto, es una gran mezcla de reduplicaciones, spam, listas, páginas sin sentido con innumerables tipos de textos, pero esto puede reflejar la naturaleza de la lengua. Bernardini, Baroni y Evert (2006), de hecho, consideran la Web como un «suplente de corpus».

Para extraer la información de la Web se puede usar Google como un programa de concordancias. Es una interfaz muy simple y clara que no requiere un gran conocimiento lingüístico o extralingüístico (Boulton, 2012). Sin embargo, Buyse (2017) encuentra los siguientes inconvenientes en el uso de Google: la falta de selección por expertos y de lematización y anotación y el que no se pueda limitar las búsquedas a temas o a tipos de textos como orales o académicos. No obstante, el área geográfica sí es un criterio de búsqueda.

Una mejor herramienta para analizar la Web es Webcorp. Con este programa se pueden obtener concordancias y también se pueden extraer listas de palabras de textos de la Web. En la Web se pueden encontrar también otros recursos como corpus multilingües que se actualizan constantemente y que, a diferencia de los diccionarios, permiten la traducción en contexto. de modo que se evita la traducción de palabra a palabra como Linguee³, Reverso⁴ o SpanishDict⁵. Estos recursos presentan una interfaz parecida a la de los diccionarios en línea, es decir, mucho más simple que la de los corpus lingüísticos. Además de los equivalentes y sinónimos, ofrecen un listado de ejemplos con las dos lenguas en paralelo, la nativa y la lengua meta. El número de ejemplos es limitado, 30 aproximadamente, lo que facilita la investigación del estudiante aunque reduce la exposición a la forma o expresión y, con ello, a algunos de los

³ <https://www.linguee.com/>

⁴ https://www.reverso.net/text_translation.aspx?lang=ES

⁵ <https://www.spanishdict.com/>

contextos de uso. Buyse recomienda asimismo usar corpus pequeños que se manejan fácilmente como Wortschatz⁶, de la Universidad de Leipzig. Además de concordancias, grados de frecuencia y textos contextualizados, este corpus ofrece una visualización gráfica (un mapa semántico) con las coocurrencias más frecuentes. Buyse (2017) realiza una útil comparativa entre seis recursos, cinco corpus y Google que se muestra en la tabla 1.

Tabla 1. Comparativa entre diferentes recursos en línea (Buyse, 2017)

	Google	WebCorp	Wortschatz	CORPES XXI	Corpus del Español	Linguee
Facilidad de uso	+	+	+	+	+	+
Multilingüe	+	+	+	-	-	+
Histórico vs sincrónico	-	-	+	-	+	--
Selección / actualización	-/+	-/+	+/-	+/-	+/+	+/+
Calidad / cantidad	-/+	-/+	+/-	+/+	+/+	+/+
Lematización	-	-	-	+	+	-
POS tagging	-	-	-	+	+	-
Etiquetado semántico	-	-	-	-	+	-
Mayúsculas / acentuación/ puntuación	- /+/-	+/+/-	+/+/+	+	-/+/+	- /+/-
Comodines y operadores	-/+	-/+	+/-	+	+/-	-
Combinaciones de palabras	+	+	+	++	++	+
Selección temática / geográfica / textual	- /+/-	+/+/-	-/-/-	+	+	-
Comparar subcorpus	-	-	-	+	++	-
Sacar concordancias	-	+	-	+	+	-
Reordenar concordancias	-	+	-	+	+	-
Frecuencias y distribución	+/-	+/-	+/-	+/+	+/+	-

Por último, WordAndPhrase, de Mark Davies, está basado en el CdeE y cuenta con dos funciones muy útiles: el listado de frecuencias y el análisis de textos. La primera función proporciona equivalentes en inglés, sinónimos, colocaciones, información sobre frecuencia por género textual y país y 200 líneas de concordancia. La segunda permite introducir un texto y descubrir el índice de frecuencia de palabra mediante un mapeado de colores.

⁶ <https://corpora.uni-leipzig.de/en>

Además, si se hace clic en la palabra se obtiene toda la información detallada sobre la frecuencia.

Al igual que los corpus, estos recursos se pueden usar como herramientas de consulta, como apoyo a la lectura o escritura, o como herramientas de aprendizaje. En el primero de los casos, el estudiante las usará principalmente de forma autónoma y, en el segundo, serán sobre todo el instrumento de diferentes tareas de investigación, individuales y grupales, propuestas por el profesor. En este caso, estas actividades pueden servir como una iniciación a ABD y, poco a poco, mostrar todas las funciones de los corpus de referencia.

3.4. Un ejemplo de uso de Linguee y WordAndPhrase

Veamos un ejemplo práctico de actividad de investigación de ABD usando dos de los recursos presentados en el apartado anterior. Un estudiante de B1 quiere buscar un equivalente de «creepy» en español y el profesor le anima a consultar el sitio Web Linguee.

The screenshot displays the Linguee interface for the search term 'creepy'. At the top, the language pair is set to English ↔ Spanish. The search results include a dictionary entry for 'creepy' (adjective) with the Spanish translation 'espeluznante' and several other synonyms: 'escalofriante', 'horripilante', 'horroroso', 'estremecedor', 'aterrador', and 'espantoso'. An example sentence is provided: 'creepy-crawly n — homiugo desagradable m'. Below the dictionary entry, there is a grid of search results from various websites, each showing a snippet of text where the word 'creepy' is used in context. For example, 'Spider Brownies are cute, creepy and delicious.' and 'Strings of red or blue Christmas lights provide more dim, creepy lighting.'

Figura 7. Primera búsqueda de «creepy» en Linguee

El estudiante se encuentra con dos adjetivos que no conocía: «escalofriante» y «espeluznante» y decide investigarlos un poco más. Primero los busca en Linguee.

Nos encontramos en un punto muerto escalofriante e inútil y, mientras tanto, la impaciencia y la desconfianza consumen a los Estados Miembros	We are at a chilling and unproductive standstill, while impatience and mistrust are brewing among Member States.
Me sorprendió realmente que habiendo crecido en partes tan diferentes del planeta tengamos opiniones tan similares que a veces es escalofriante .	I was just really surprised that we grew up on different sides of the planet and we have such similar opinions that it is creepy sometimes .
En las páginas 8 y 12 hay otro problema igual de escalofriante . Hemos reforzado los créditos de pago.	On pages 8 and 12, there is an equally terrifying problem: we have increased the payment appropriations.
El número de víctimas infantiles -más de 300 muertos y de 1.500 heridos- escalofriante	The number of child victims - more than 300 killed and in excess of 1,500 injured - is harrowing .
[...] de lo que él dice que hará y cómo va a hacerlo son una descripción escalofriante y exacta de lo que probablemente hará.	[...] perpetrator says he will do and how he will do it are a chilling accurate description of what the perpetrator is likely to do.
En la Evaluación de Ecosistemas del Milenio, publicada recientemente, se llega a la escalofriante conclusión de que, a menos que tomemos medidas inmediatas [...]	The recently released Millennium Ecosystem Assessment chillingly concludes that, unless we take action to mitigate the decline [...]
El vertido a los ríos de cianuro y de metales pesados es un ejemplo escalofriante de hasta qué punto las actitudes negligentes hacia el medio ambiente [...]	The cyanide and heavy metals in the rivers are gruesome examples of how environmental negligence may ruin the waterways for decades to come.
El odio entre los hermanos, el menosprecio, el dolor por la muerte es en Leenane tan escalofriante que hasta hace reír.	Hatred between brothers, contempt and grief is so horrifying in Leenane it even makes you laugh.
[...] desmedido a la riqueza" y escribió con una antelación escalofriante : "En otras tierras se libran peleas de raza y batallas políticas.	[...] about by 'the unbridled cult of wealth' and wrote with chilling prescience . In other countries there are racial conflicts [...]
En un caso especialmente escalofriante , muchos trabajadores somalíes de las Naciones Unidas y sus familiares tuvieron que [...]	In a particularly chilling turn of events, many Somali United Nations staff and their families had to be relocated from [...]
Este hecho espeluznante es ilustrativo del siniestro afán despiadado de Mugabe de aferrarse al poder.	This abhorrent fact marks out the sinister context of Mugabe's ruthless campaign to stay in power.
Esta es un triste, inexcusable y espeluznante ejemplo de la violencia experimentada por los miembros del sindicato en el sector bananero.	This is a sad, inexcusable and horrifying example of the violence experienced by union members in the banana sector.
Debido a la naturaleza espeluznante y un tanto increíble de la noticia, Riegner se abstuvo de pasar esta información hasta [...]	Due to the shocking and somewhat unbelievable nature of the report, Riegner refrained from passing on this information until he investigated its source.
Los autores del informe nos ofrecieron una lectura alarmante y a menudo espeluznante , y pidieron que se tomaran medidas	The authors of the report presented us with disturbing and often shocking reading, and called for action.
El visitante camina a través de esta espeluznante escena como si navegara por una representación bidimensional, una imagen [...]	The gallery visitor walks through this eerie scene as if somehow navigating a two-dimensional representation, a forensic [...]
Las infecciones que se adquieren durante la hospitalización le cuestan a los EE. UU. una cantidad espeluznante en vidas y valor económico de la atención de la salud.	Infections that are acquired during the course of a hospital stay cost the United States a staggering amount in terms of lives lost and health care costs.
Y finalmente, pero no menos importante, la moto tiene un sonido espeluznante con un toque agresivo.	And last but not least, the bike has this great sound with an aggressive touch.
Las atracciones más populares son el espeluznante montaña rusa llamada Jaguar (sólo apto para los más atrevidos) y el teatro.	The most popular attractions are the hair-raising roller-coaster called the Jaguar (only suitable for the very daring) and the theatre shows.

Figura 8. Búsqueda de «espeluznante» y «escalofriante» en Linguee

Las líneas de concordancia muestran ya unos patrones. En los ejemplos de Linguee, «escalofriante» se usa para referirse a información estadística particularmente negativa pero no aparecen ejemplos de escenas de miedo, de terror. Sin embargo, «espeluznante», si bien se usa también para un informe, se presenta más en escenas siniestras o en relación con la violencia.

Para comprobar estas hipótesis, el profesor sugiere al estudiante que utilice otra herramienta, WordAndPhrase. El estudiante lleva a cabo una búsqueda en esta herramienta y obtiene inmediatamente resultados útiles.

The image displays two screenshots of the WordAndPhrase Spanish interface. The top screenshot shows the search results for the word «escalofriante». The interface includes a navigation bar with options like 'BROWSE LIST', 'LIST OF WORDS', 'SEE ENTRY', and 'ACCOUNT'. The search results for «escalofriante» (FREQ 5703) include synonyms such as 'conmover', 'espeluznante', 'emocionante', 'impresionante', 'estremecedor', and 'estremecedor (loud)'. It also lists collocates and topics related to the word.

The bottom screenshot shows the search results for the word «espeluznante». The interface includes a navigation bar with options like 'BROWSE LIST', 'LIST OF WORDS', 'SEE ENTRY', and 'ANALYZE TEXT'. The search results for «espeluznante» (RANK 12343, FREQ 3960) include synonyms such as 'avieso', 'tétrico', 'odioso', 'repugnante', 'inico', 'maligno', 'siniestro', 'conmover', 'escalofriante', 'turbador', 'emocionante', 'impresionante', 'estremecedor', 'espantoso', 'espantoso (frightening)', 'espantoso', 'estremecedor (loud)', 'emocionante', 'impresionante', 'conmover', 'turbador', 'escalofriante', 'pavoroso', 'horrible', 'horroroso', 'apocalíptico', 'espantoso', 'aterrador', 'horrible', 'horrendo', 'terrible', 'espantable', 'espantoso', 'temeroso', 'horrendo', 'horrible', 'horroroso', 'temible', 'trágico', 'conmover', 'desgraciado', 'calamburoso', 'tremendo', 'impresionante', 'emocionante', 'dramático', and 'tremebundo (terrible)'. It also lists collocates and topics related to the word.

Figura 9. Búsqueda de «escalofriante» y «espeluznante» en WordAndPhrase Spanish

Esta simple búsqueda ya muestra que:

- «Escalofriante» es bastante más frecuente que «espeluznante». Se encuentra en la posición 10.278 mientras que «espeluznante» ocupa el rango 12.343.
- «Espeluznante» es más frecuente en español hablado que «escalofriante».
- Por la lista de sinónimos parece que «espeluznante» tiene connotaciones más negativas. También algunas de las colocaciones son más siniestras, por ejemplo, «alarido», «aullido» o «tortura» mientras que el colocativo de «escalofriante» es «grito».

El estudiante puede crear entonces una pequeña ficha (definición, equivalente, sinónimos, colocaciones, frecuencias, registro y varios ejemplos) de cada término y presentarla a la clase.

4. Discusión y conclusiones

Con el presente artículo hemos pretendido acercar al profesorado de ELE la definición de ABD, las investigaciones que existen al respecto y el modo de trabajarlo con los recursos existentes en la actualidad. Hemos presentado los dos tipos de recursos: los grandes corpus de referencia y otros recursos más recientes, en su mayoría híbridos, y hemos mostrado ejemplos de aplicación para los dos tipos de recursos. Los dos grupos de herramientas presentan ventajas e inconvenientes.

La búsqueda con recursos híbridos ofrece varias ventajas con respecto a la búsqueda realizada con corpus «tradicionales». En primer lugar, estos recursos son mucho más simples de usar y su interfaz es muy familiar para el estudiante. En concreto, muchos estudiantes conocen ya Linguee (y/o recursos similares) y la interfaz de WordAndPhrase es muy parecida a la de un diccionario, con todos los datos adicionales presentados de manera compacta en una única página o pantalla. Por añadidura, al ser la búsqueda más simple, es también mucho más rápida. Por último, son menos los ejemplos, por lo que es más fácil leerlos que las líneas de concordancia de un corpus. No obstante, la búsqueda en corpus ofrece muchas más posibilidades para la investigación de patrones como el gran número de ejemplos, los listados de colocaciones, la posibilidad de comparar dos términos o la mayor refinación (al ser corpus anotados) en las búsquedas de estructuras (búsqueda por tipo de palabra, tiempo verbal, etc.).

Es el profesor el que debe decidir qué herramientas pueden ser más adecuadas para sus estudiantes, por nivel y por capacidad tecnológica y habilidades analíticas. Nosotros somos de la opinión de que los recursos híbridos son los más adecuados para los niveles más bajos, A1, A2 y B1, mientras que los corpus se podrían introducir a partir del nivel B2. Es también el papel de profesor diseñar actividades colaborativas en clase para investigar contenidos lingüísticos. Lo más importante es mostrar a los estudiantes ejemplos de investigaciones y de cómo una simple búsqueda no es suficiente. Tras la primera interpretación hay que buscar de nuevo para confirmar o no esa primera hipótesis. La secuenciación propuesta por Sinclair (2003) y Pérez- Paredes et al. (2011) es un buen marco para esta investigación. Ahora bien, el profesor sigue necesitando formación en el uso de corpus y en el diseño de actividades por descubrimiento. El profesor necesita actividades ya creadas y disponibles que, a su vez, sean

buenos modelos para crear las suyas propias. Esta formación se podría llevar a cabo en los programas de formación inicial de profesores, así como en cursos de formación especializada.

Para el alumno, recursos híbridos como Linguee o WordAndPhrase funcionan como una herramienta de consulta de grandes prestaciones y como instrumento de aprendizaje con el que pueden profundizar su conocimiento lingüístico. Estos recursos son un primer paso para iniciarse en el enfoque y poco a poco aprender a usar herramientas más poderosas y sofisticadas. El objetivo final es que el estudiante adquiera destrezas de estudio y destrezas de descubrimiento y análisis que le permitan tomar las riendas de su aprendizaje.

En conclusión, en este artículo mostramos que el enfoque ABD se puede llevar a la práctica con herramientas más sencillas y más conocidas por nuestros estudiantes. Asimismo, en el ámbito de ELE son necesarios más estudios empíricos sobre ABD con diferentes tipos de datos para probar su eficacia y para investigar las percepciones de los estudiantes. Esperamos que este artículo deje entrever a otros educadores las posibilidades de este enfoque. El Aprendizaje basado en datos ha venido para quedarse, los nuevos usos y herramientas así lo confirman.

Referencias bibliográficas

- ASENCIÓN-DELANEY, Y., COLLENTINE, J. G., COLLENTINE, K., COLMENARES, J., y PLONSKY, L. (2015). El potencial de la enseñanza del vocabulario basada en corpus: optimismo con precaución. *Journal of Spanish Language Teaching*, 2(2), 140-151. DOI:10.1080/23247797.2015.1105516
- BALUNDA, S. A. (2009). *Teaching academic vocabulary with corpora: students perceptios of data-driven learning*. Memoria de máster. Indiana University. [<https://scholarworks.iupui.edu/bitstream/handle/1805/2049/Balunda%20MA%20Thesis%20Teaching%20Academic%20Vocabulary%20with%20Corpora.pdf>] [Consultado el 14/11/2021]
- BENAVIDES, C. (2015). Using a Corpus in a 300-Level Spanish Grammar Course. *Foreign Language Annals*, 48(2), 218-235. DOI: 10.1111/flan.12136
- BERNARDINI, S. (2004). Corpora in the classroom: An overview and some reflections on future developments. En J. SINCLAIR (Ed.), *How to Use Corpora in Language Teaching*. Amsterdam: Benjamins. DOI:10.1075/scl.12.05ber
- BERNARDINI, S. (2016). Discovery Learning in the language-for-translation classroom: corpora as learning aids. *Cad. Trad.*, 36 (nº especial 1), 14-35. [http://www.scielo.br/scielo.php?script=sci_arttext&pid=S2175-79682016000500014] [Consultado el 14/11/2021]
- BERNARDINI, S., BARONI, M., y EVERT, S. (2006). A WaCky introduction. En M. BARONI, y S. BERNARDINI (Eds.), *Wacky! Working papers on the Web as corpus* (pp. 9-40). Bologna: Gedit.
- BOULTON, A. (2008). DDL: Reaching the parts other teaching can't reach? En A. Frankenburg-García (Ed.), *Proceedings of the 8th Teaching and Language Corpora Conference* (pp. 38-44). Lisboa, Portugal: Associação de Estudos e de Investigação Científica do ISLA-Lisboa.

- BOULTON, A. (2009). Data-driven learning: reasonable fears and rational reassurance. *Indian Journal of Applied Linguistics*, 35(1), 81-106. [<https://hal.archives-ouvertes.fr/hal-00326990v2>] [Consultado el 14/11/2021]
- BOULTON, A. (2012). What data for data-driven learning? *The EUROCALL REVIEW*, 20(1), 23-27.
- BOULTON, A. (2017). Research Timeline Corpora in language teaching and learning. *Language Teaching*, 50(4), 483-506. DOI:10.1017/S0261444817000167
- BOULTON, A., & COBB, T. (2017). Corpus Use in Language Learning: A Meta-Analysis. *Language learning*, 67(2), 348-393. DOI: <https://doi.org/10.1111/lang.12224>
- BUYSE, K. (2017). Los corpus como herramienta de aprendizaje de léxico. En F. HERRERA (Ed.), *Enseñar léxico en el aula de español. El poder de las palabras* (pp. 121-140). Barcelona: Difusión.
- BUYSE, K., & VERLINDE, S. (2013). Possible Effects of Free on Line Data Driven Lexicographic Instruments on Foreign Language Learning: The Case of Linguee and the Interactive Language Toolbox. *Procedia*, 95, 507-512. DOI:<https://doi.org/10.1016/j.sbspro.2013.10.675>
- CHAMBERS, A. (2010). What is data-driven learning? En A. O'KEEFFE, y M. MCCARTHY (Eds.), *The Routledge Handbook of Corpus Linguistics* (pp. 345-358). Londres y Nueva York: Routledge.
- CONTRERAS IZQUIERDO, N. (2019). Innovación docente en educación superior: recursos lexicográficos en el aprendizaje basado en datos (ABD) para la enseñanza de las variedades del español. En J. GÁZQUEZ LINARES, M. MOLERO JURADO, A. BARRAGÁN MARTÍN, M. SIMÓN MÁRQUEZ, Á. MARTOS MARTÍNEZ, J. SORAINO SÁNCHEZ, y N. F. OROPESA RUIZ (Eds.), *Innovación docente e investigación en Arte y Humanidades* (pp. 521-536). Dykinson.
- FLOWERDEW, L. (2015). Data-driven learning and language learning theories. En A. LENKO-SZYMANSKA, y A. BOULTON (Eds.), *Multiple Affordances of Language Corpora for Data-driven learning* (pp. 15-36). John Benjamins.
- GILQUIN, G., y GRANGER, S. (2010). How can data-driven learning be used in language teaching? En A. O'KEEFFE, y M. MCCARTHY (Eds.), *The Routledge Handbook of Corpus Linguistics* (pp. 359-370). Londres y Nueva York: Routledge.
- JOHNS, T. (1990). From printout to handout: Grammar and vocabulary teaching in the context of datadriven. *CALL Austria*, 10, 14-34.
- JOHNS, T. (1991). From printout to handout: Grammar and vocabulary teaching in the context of data-driven learning. En T. JOHNS, y P. KING (Eds.) *Classroom Concordancing. English Language Research Journal*, 4, 27-45.
- KENNEDY, C., & MICELI, T. (2001). An evaluation of intermediate students' approaches to corpus investigation. *Language Learning and Technology*, 5(3), 77-90.
- KILGARRIFF, A., y GREFENSTETTE, G. (2003). Introduction to the special issue on Web as a corpus. *Corpus Linguistics and Linguistic Theory*, 29(3), 263-275.
- KRASHEN, S. K. (1988). *Second Language Acquisition and Second Language Learning*. Oxford: Pergamon Press.

- LEE, H., WARSCHAUER, M., y LEE, J. H. (2019). The Effects of Corpus Use on Second Language Vocabulary Learning: A Multilevel Meta-analysis. *Applied Linguistics*, 40(5), 721-753. DOI:10.1093/applin/amy012
- MARCOS MIGUEL, N. (2020). Exploring tasks-as-process in Spanish L2 classrooms: Can corpus-based tasks facilitate language exploration, language use, and engagement? *International Journal of Applied Linguistics*(Special issue article). DOI:10.1111/ijal.12314
- MCCARTHY, M., y CARTER, R. (1995). Spoken grammar: what is it and how can we teach it? *ELT Journal Volume*, 49(3), 207-218.
- O'SULLIVAN, I. (2007). Enhancing a process-oriented approach to literacy and language learning: The role of corpus consultation literacy. *ReCALL*, 19(3), 269-286.
- PAZ, Y. (2013). Aplicación pedagógica de la lingüística de corpus a una clase de gramática avanzada de español como segunda lengua. *MarcoELE Revista de Didáctica ELE*, 6. [https://marcoele.com/linguistica-de-corpus/] [Consultado el 14/11/2021]
- PELÁEZ PINO, I. (2019). *Propuesta para la aplicación didáctica de los corpus de nativos en las clases de español de los negocios*. Memoria de máster. Universitat de Barcelona. [http://diposit.ub.edu/dspace/bitstream/2445/141297/1/TFM_Ivan%20Pelaez%20Pino.pdf] [Consultado el 14/11/2021]
- PÉREZ-PAREDES, P. (2010). Corpus Linguistics and Language Education in Perspective: Appropriation and the Possibilities Scenario. En T. HARRIS, y M. MORENO JAÉN, *Corpus Linguistics in Language Teaching* (pp. 53-73). Peter Lang.
- PÉREZ-PAREDES, P. (2019). A systematic review of the uses and spread of corpora and data-driven learning in CALL research during 2011-2015. *Computer Assisted Language Learning*. DOI:https://doi.org/10.1080/09588221.2019.1667832
- PÉREZ-PAREDES, P., y ZAPATA-ROS, M. (2018). Patrones de Pensamiento Computacional y corpus lingüísticos: el aprendizaje de lenguas con datos lingüísticos. *RED. El aprendizaje en la Sociedad del Conocimiento*. [http://eprints.rclis.org/32209/] [Consultado el 14/11/2021]
- PÉREZ-PAREDES, P., ORDOÑANA GUILLAMÓN, C., VAN DE VYVER, J., MEURICE, A., AGUADO JIMÉNEZ, P., CONOLE, G., y SÁNCHEZ HERNÁNDEZ, P. (2019). Mobile data-driven language learning: Affordances and learners' perception. *System*, 84, 145-159. DOI:https://doi.org/10.1016/j.system.2019.06.009
- PÉREZ-PAREDES, P., SÁNCHEZ-TORNEL, M., ALCARAZ CALERO, J., & JIMÉNEZ, P. (2011). Tracking learners' actual uses of corpora: guided vs non-guided corpus consultation. *Computer Assisted Language Learning*, 24(3), 233-253.
- RODRIGUEZ CASTELLANO, S. (2017). *Trabajar el léxico en el aula de ELE mediante el uso directo de un corpus lingüístico*. Memoria de máster. Universitat Jaume I. [http://repositori.uji.es/xmlui/bitstream/handle/10234/172831/TFM_2017_RodriguezCastellano_Sara.pdf?sequence=1] [Consultado el 14/11/2021]
- SCHMIDT, R. (1990). The Role of Consciousness in Second Language Learning. *Applied Linguistics*, 11, 129-158.
- SEPEHRI, M. (2015). *The effects of Data Driven Learning on Iranian EFL Learners' Writing Skills Development*. Tesis doctoral sin publicar. University of Birmingham. [https://etheses.bham.ac.uk/id/eprint/6267/] [Consultado el 14/11/2021]
- SINCLAIR, J. (2003). *Reading concordances: an introduction*. Harlow: Pearson Longman.

- SINCLAIR, J. (2005). Corpus and text: Basic principles. En M. WYNNE (Ed.), *Developing linguistic corpora. A guide to good practice* (pp. 5-24). Oxford: Oxbow Books.
- URIEL DOMÍNGUEZ, M. (2015). Las colocaciones en un corpus de aprendices valón y flamenco. En Y. MORIMOTO, M. V. PAVÓN LUCERO, y R. SANTAMARÍA MARTÍNEZ (Eds.), *La enseñanza de ELE centrada en el alumno. Asociación para la Enseñanza del Español como Lengua Extranjera*, (pp. 977-986). [https://cvc.cervantes.es/ensenanza/biblioteca_ele/asele/pdf/25/25_0977.pdf] [Consultado el 14/11/2021] [Consultado el 14/11/2021]
- Vincze, O. (2015). *Learning multiword expressions from corpora and dictionaries*. [Tesis doctoral. Universade da Coruña]. [https://core.ac.uk/download/pdf/61917869.pdf] [Consultado el 14/11/2021]
- VYATKINA, N., & BOULTON, A. (2017). Corpora in language learning and teaching. *Language Learning & Technology*, 21(3), 1-8. [https://scholarspace.manoa.hawaii.edu/bitstream/10125/44750/1/21_03_commentary.pdf] [Consultado el 14/11/2021]
- VYGOTSKI, L. S. (1978). *Pensamiento y lenguaje*. Buenos Aires: La Pleyade.
- WIDDOWSON, H. (1978). *Teaching Language as Communication*. Oxford: Oxford University Press.
- YAO, G. (2019). Vocabulary learning through data-driven learning in the context of Spanish as a foreign language. *Research in Corpus Linguistics*, 7, 18-46. DOI:10.32714/ricl.07.02