

El potencial de la enseñanza del vocabulario a través del Aprendizaje Basado en Datos (ABD) en el aula de ELE

The potential of teaching vocabulary through Data-Driven Learning (DDL) in the Spanish classroom

Yan He
Universidad de Estudios Extranjeros de Guangdong, China
heyang@oamail.gdufs.edu.cn

Jie Fang
Universidad de Guangzhou, China
jessie_fang@163.com

Yan He es doctora en Estudios Hispánicos. Lengua, Literatura, Historia y Pensamiento por la Universidad Autónoma de Madrid, máster en Lengua Española: Investigación y Prácticas Profesionales por esta misma universidad y licenciada en Filología Hispánica por la Universidad de Estudios Extranjeros de Guangdong, China. Actualmente es profesora titular de español en el Departamento de Lengua Española de la Universidad de Estudios Extranjeros de Guangdong, China. Sus principales campos de investigación son la semántica, la enseñanza del español como segunda lengua en el aula de China y la traducción jurídica español-chino.

Jie Fang es estudiante de licenciatura en Matemáticas y Matemáticas Aplicadas de la Universidad de Guangzhou, China. Ha participado y ganado varios concursos nacionales de modelización matemática.

Resumen

En este trabajo se ha llevado a cabo un estudio cuantitativo para evaluar los efectos del aprendizaje basado en datos (ABD) en el aula de ELE de nivel intermedio. Concretamente, se ha realizado una comparación entre el método de acceso al corpus con mediación del profesor y el método tradicional basado en diccionarios en un mismo grupo de personas en una clase de vocabulario. Los resultados de la prueba han mostrado que el ABD puede promover eficazmente la adquisición de conocimientos léxicos a corto plazo. Sin embargo, los resultados de la prueba final han revelado que el ABD no ayuda significativamente a la memoria a largo plazo. Los resultados de la encuesta han mostrado

que los estudiantes toman una actitud positiva frente al ABD, pero el problema del tiempo es el mayor obstáculo para su futura implementación en el aula.

Abstract

In this paper, a quantitative research has been carried out to evaluate the effects of Data-Driven Learning (DDL) in the intermediate level ELE classroom. Specifically, a comparison was made between the teacher-mediated corpus access method and the traditional dictionary-based method in the same group of people in a vocabulary class. The test results show that DDL can effectively promote the acquisition of lexical knowledge in the short term. However, delayed test results reveal that DDL does not significantly help long-term memory. Furthermore, the survey results show that students take a positive attitude towards DDL, but the problem of lack of time is the biggest obstacle to its future implementation in the classroom.

Palabras clave

Aprendizaje basado en datos, corpus, enseñanza de vocabulario, ELE.

Keywords

Data-driven learning, corpus, vocabulary teaching, ELE.

Introducción

En los últimos años, el desarrollo de la lingüística de corpus no solo ha cambiado los métodos utilizados en los distintos niveles de análisis de la lengua, como fonético, fonológico, morfológico, sintáctico, semántico y pragmático, sino que también se han aplicado en áreas como la enseñanza de lenguas, el análisis del discurso, la lexicografía, etc.

Desde un punto de vista didáctico, tal como se indica en Tsui (2004), los corpus son una herramienta útil para los profesores, sobre todo, para los profesores no nativos, porque les ayudan a solucionar cuestiones que no aparecen en los libros de gramática y corroboran sus creencias. Además, los corpus pueden ser útiles para la preparación de actividades de producción, por ejemplo, en las actividades de expresión o interacción escritas, sirven para comprobar cuestiones sintácticas como régimen de palabras u orden de palabras y para consultar dudas sobre vocabulario como la variedad sociolingüística o colocaciones habituales.

En cuanto al contexto de enseñanza de China, los progresos informáticos han estimulado a los profesores de lengua extranjera a traer los corpus en el aula. Esto se ha evidenciado en muchos trabajos al respecto (Tsui, 2004; Zhu, 2009; Zhang y Zhou, 2012, etc.). Sin embargo, estos trabajos se han centrado en la enseñanza del inglés. En la enseñanza del español se ha recibido menos atención. En la búsqueda de referencias, se han encontrado tres trabajos relativamente relevantes en este tema: Chao (2015) hizo una presentación detallada sobre la herramienta Sketch Engine y propuso una actividad didáctica en la distinción de *querer* y *necesitar* basada en el corpus EsTenTen 11; Lu, Hung y Lu (2016) evaluaron la efectividad de aplicar el Corpus Escrito de Aprendices Taiwanesees de Español y comprobaron que la aplicación del corpus puede facilitar la

extensión de la longitud de los textos; y Wei (2020) propuso algunas actividades didácticas basadas en los corpus paralelos chino-español.

Actualmente los procedimientos de enseñanza y aprendizaje del español son relativamente simples y anticuados en China. Por ejemplo, en la clase de enseñanza del vocabulario, los profesores suelen extraer ejemplos de los diccionarios tradicionales monolingües o bilingües o, en algunos casos, inventar o compilar esos ejemplos por conveniencia. Tales oraciones, carentes de autenticidad o de interferencia contextual, no ayudan a los estudiantes a dominar completamente el uso de las palabras. A pesar de que cada día más profesores tratan de innovar los métodos didácticos en el aula de ELE¹, los estudiantes siguen basándose en gran medida en la explicación de los ejemplos sacados de diccionarios por parte del profesor en clase y la práctica de pura memorización fuera², lo que dificulta el aprendizaje de un español “natural”, ya que no han tenido contacto con el uso más auténtico de la lengua.

Una posible explicación a la falta de interés por parte de los profesores en la implementación del corpus en el aula de ELE puede ser que necesiten una formación especializada y un periodo de entrenamiento para poder almacenar mucha de la información que puede ofrecer un corpus (Vyatkina y Boulton, 2017, citado en Abad Castelló, 2019). Más allá de este aspecto, creemos que la razón principal es la falta de investigaciones empíricas que demuestren la eficacia de este método de enseñanza del español. Por ello, el presente trabajo tiene como objetivo evaluar la utilidad de este modelo mediante una investigación empírica en el aula de ELE.

Aprendizaje basado en datos³

Los progresos computacionales y tecnológicos están estrechamente relacionados con el desarrollo de los corpus. A medida que surgen recursos de corpus disponibles de forma gratuita y de fácil acceso, el aprendizaje de idiomas asistido por corpus se ha hecho realidad. En 1990 Johns propuso el término “aprendizaje basado en datos” (ABD) para definir un enfoque pedagógico en el que los estudiantes pueden conocer inductivamente las características y regularidades lingüísticas mediante la exploración de datos lingüísticos reales y auténticos, es decir, de los corpus. Durante este proceso, el estudiante es el investigador y el profesor es un mero facilitador. Es el alumno el que examina datos provenientes de los corpus lingüísticos y el que extrae sus conclusiones de los mismos (Johns, 1991). A partir de la formulación del término, muchos lingüistas han destacado que dicho modelo de aprendizaje permite a los estudiantes observar ejemplos de uso evidenciados, les ayuda a desarrollar destrezas analíticas y les capacita para trabajar de manera autónoma (Asención-Delaney, Collentine, Collentine, Colmenares y Plonsky, 2015; Bernardini, 2016; Vyatkina y Boulton, 2017, etc.).

Existen dos maneras para trabajar con el corpus en el aula: el acceso directo por parte de los estudiantes o con mediación del profesor. Al comienzo de la formulación del ABD, el método de acceso directo (hands on DDL) fue elegido para su implementación

¹ En Zhen, Jiménez Cruz y Montero Lorenzo (2020) se ha hecho una revisión de la enseñanza del español como lengua extranjera en China y el desarrollo del profesorado. El trabajo indica que coexisten una serie de teorías didácticas para facilitar el aprendizaje de esa lengua extranjera, como por ejemplo, el método estructural, el método audio-lingual, el análisis contrastivo, etc.

² Según Blanco (2011), el pensamiento de Confucio forma parte de la tradición educativa china y se erige la norma de que el alumno no refute o contradiga la enseñanza impartida por su profesor. Influida por esta herencia cultural, durante un largo periodo en la enseñanza prevalecía la concepción lineal del aprendizaje que partía de la memorización del contenido impartido por el profesor.

³ Data-driven learning (DDL), en inglés.

en el aula. Los estudiantes entraban en contacto directo con el ordenador y aprendían directamente los datos recuperados del corpus. Un gran número de investigaciones empíricas han estudiado y analizado el efecto de este método didáctico. Los resultados obtenidos han demostrado que el método hace que los alumnos sean más autónomos (Daskalovska, 2015) pero también puede abrumar a los estudiantes porque los datos pueden ser “irrelevantes”, “incomprensibles” y “extremadamente caóticos” (Boulton 2010: 6). Además, los datos del corpus normalmente son textos producidos por hablantes nativos y, por lo tanto, la información podría estar fuera de la competencia lingüística de un alumno. Por último, la consulta directa del corpus requiere acceso a un ordenador, lo que significa que los alumnos deben tener conocimientos básicos de informática y los centros educativos deben estar equipados con salas multimedia. Pero esta condición es difícil de alcanzar para algunos estudiantes y centros educativos (Lenko-Szymanska, 2014).

En cambio, el acceso con mediación del profesor (hands off DDL) introduce recursos del corpus previamente seleccionados por el profesor. En este caso es el profesor quien consulta los corpus directamente, luego selecciona los datos lingüísticos apropiados según el nivel del alumno y finalmente prepara ejercicios basados en las concordancias. Las ventajas de esta forma son bastante obvias: por un lado, los estudiantes pueden beneficiarse del acceso directo a datos de idioma auténticos (en lugar de la consulta directa del corpus) y por el otro, no se requieren conocimientos y competencias informáticas. Thompson (2006) también señala que los materiales previa y adecuadamente seleccionados pueden ayudar a los alumnos a centrar su atención en los elementos clave, reducir la confusión y limitar el rango de posibles respuestas.

En una clase de ELE, en la que los estudiantes normalmente no disponen de un nivel avanzado del idioma, optamos por implementar el segundo del ABD, es decir, el acceso al corpus con mediación del profesor, en una clase de la enseñanza y aprendizaje del vocabulario para comprobar su efecto didáctico.

Metodología

Este experimento estaba destinado a comparar la eficacia del aprendizaje del vocabulario en español a través del método ABD y el método basado en el diccionario. Para resaltar las diferencias con mayor claridad, se seleccionó un mismo grupo de estudiantes como sujeto del experimento, es decir, se repitieron los dos distintos modelos de aprendizaje del vocabulario en este mismo grupo.

1. Participantes

Los sujetos de este estudio eran 34 estudiantes universitarios (11 hombres y 23 mujeres) con especialización en Lengua Española y unos dos o tres años de formación en este idioma. En general, su nivel de dominio del español era intermedio-alto⁴. Antes de realizar el experimento, ninguno de ellos había tenido contacto con el corpus, ni experiencia con el aprendizaje basado en datos.

2. Materiales

⁴ Elegimos como participantes a estudiantes de nivel intermedio-alto, ya que creemos que tienen suficiente dominio del idioma para leer materiales de aprendizaje de alto nivel (como corpus sobre concordancia y diccionarios monolingües) sin necesidad de consultar otros libros de referencia. Agradecemos el apoyo y la participación de los estudiantes en este experimento.

Con el fin de analizar si el aprendizaje de vocabulario a través de ABD es más efectivo que a través de la consulta del diccionario, seleccionamos un total de 30 palabras en español como candidatas para los ítems objetivo. Las 30 palabras fueron extraídas del Programa de enseñanza para cursos superiores de las especialidades de lengua española en escuelas superiores chinas y se consideró el vocabulario de nivel avanzado, supuestamente desconocido para los participantes. Luego se realizó una prueba previa sobre estas 30 palabras y se eliminaron 20 ítems ya conocidos por ellos. Las 10 palabras seleccionadas eran: *constancia, contingencia, envergadura, englobar, fosilizar, inducción, prospección, rumiar, soldar, subyacer*. Además, se seleccionaron 8 palabras agrupadas en pares para examinar si el ABD es más efectivo que el método tradicional en el caso de la distinción de los sinónimos. Las palabras en cuestión eran *alegre/agradable, andar/caminar, demostrar/mostrar, saludable/sano*. La selección se basó en la observación de los errores más comunes por los alumnos en clase. Aunque estas palabras se consideran como fáciles, los estudiantes suelen confundirlas en la formación de oraciones o en la traducción. Agrupamos las palabras en orden alfabético y luego por la paridad de los números de serie, las dividimos en dos grupos, que se utilizaron como grupo experimental y grupo de control, tal y como se muestra en la siguiente tabla:

grupo experimental	envergadura, englobar, fosilizar, inducción, prospección, demostrar/mostrar, alegre/agradable
grupo de control	constancia, contingencia, rumiar, soldar, subyacer, andar/caminar, saludable/sano

Los materiales de aprendizaje para el grupo experimental (en adelante, GE) se basaron en oraciones extraídas directamente del Corpus de Referencia del Español Actual (CREA) y el corpus esTenTen18 de Sketch Engine, presentados en formato KWIC (keyword in context). En palabras de Cruz Piñol (2012: 43), “presentar la palabra clave en su contexto proporciona mucha información sobre el funcionamiento de ese término en la lengua real y en el nivel textual”. Las oraciones fueron cuidadosamente seleccionadas de acuerdo con los criterios de Gilquin y Granger (2010: 362): “legibilidad”, “frecuencia” y “utilidad”. Por lo tanto, se tuvo en cuenta la frecuencia de uso de los distintos significados de la palabra clave, eliminando significados que pudieran desviar la comprensión. Además, para garantizar la familiaridad y la comprensibilidad, se evitaron colocar en el material frases cortadas o demasiado largas. A continuación, se muestra un fragmento del material que preparamos para el GE⁵:

soldar tr.

1. Los generadores nuevos, cada uno de los cuales pesa 343 toneladas, ya han sido introducidos en el interior del edificio de contención y está previsto que esta semana los operarios acaben de soldar las tuberías que los conectan a los circuitos primario y secundario.
2. Recordar siempre que una buena unión entre estaño y metal no se produce si el metal no está a la temperatura de fusión del estaño (340° aprox.); por ello, antes de aplicar el estaño sobre el metal, se debe calentar previamente la pieza a soldar, apoyando sobre ella la punta del soldador.

⁵ En el material se han colocado 5 frases para cada palabra clave. Por limitación de espacio, aquí presentamos solo 2 frases para dar a conocer el modelo.

saludable/ sano adj.

1. Las organizaciones asistentes hicieron un llamamiento acerca de la necesidad de incrementar los fondos para la promoción de una alimentación saludable.
2. Cuando compras los postres algunos no son muy saludables, y terminan ocasionando mala digestión.
3. El objetivo del cuidado es conseguir una madre y un niño sanos, con el menor nivel posible de intervención de la manera más segura.
4. Conscientes de ello, el municipio prepara un paquete de medidas para la promoción de estilos de vida saludables que fomentarán la alimentación sana y equilibrada, el no consumo de tabaco y sustancias tóxicas, el control del estrés y la práctica.

Como hemos mencionado arriba, para el grupo de control (en adelante, GC) se empleó el método tradicional basado en diccionarios. En el material se incluyeron las definiciones de la palabra clave y los ejemplos citados en los diccionarios consultados. Se utilizaron principalmente dos diccionarios online: *Diccionario de la Lengua Española de la Real Academia Española* (diccionario monolingüe)⁶ y *Esdict* (diccionario bilingüe online)⁷. Abajo se ofrece una muestra del material de aprendizaje basado en diccionarios.

prospección

DLE

1. f. Exploración del subsuelo basada en el examen de los caracteres del terreno y encaminada a descubrir yacimientos minerales, petrolíferos, aguas subterráneas, etc.
2. Exploración de posibilidades futuras en indicios presentes. Prospección de mercado, de tendencias de opinión.

Esdict

1. 勘探, 勘察
2. 【引】寻找销路

Ejemplo: La prospección biológica es una cuestión emergente.

alegre/agradable

DLE

alegre

1. adj. Poseído o lleno de alegría. Juan está alegre.
2. adj. Que siente o manifiesta de ordinario alegría. Ser hombre alegre.
3. adj. Que denota alegría o viveza. Cara, juego alegre.
4. adj. Que ocasiona alegría. Noticia alegre.
5. adj. Pasado o hecho con alegría. Día, vida, plática, cena alegre.
6. adj. De aspecto o circunstancias capaces de infundir alegría. Cielo, prado, casa alegre.
7. adj. Dicho de un color: Vivo o intenso.
8. adj. De color o colores alegres.

agradable

1. adj. Que produce complacencia o agrado.
2. adj. Dicho de una persona: Afable en el trato.
3. adj. desus. Dicho de una persona: Que tiene complacencia o gusto en hacer algo.

Esdict

⁶ La página web: <https://dle.rae.es/>

⁷ La página web: Esdict.cn

alegre adj.

1. «estar, sentirse» 高兴的, 快乐的, 欢喜的:

Todo el mundo se siente ~ en este hermoso día de primavera. 在这个春光明媚的日子里大家都感到愉快.

Hoy trae cara ~ . 他今天满面春风.

Estamos ~s con la noticia. 我们都为这个消息高兴.

Estamos ~s de tenerlos aquí. 我们很高兴看到你们.

2. «ser» 快活的:

Es un hombre ~ . 他是一个开朗的人.

3. 使人高兴的, 令人愉快的:

noticia ~ 喜讯.

día ~ 高兴的日子.

paisaje ~ 秀丽的景色.

4. 明亮的, 豁亮的 (房间) .

agradable adj.

«a, para; de» 令人愉快的, 使人感到惬意的, 令人喜欢的.

3. Instrumentos

3.1. Prueba inmediata

La prueba constaba de dos partes. La primera parte estaba destinada a medir la profundidad del conocimiento de las palabras aprendidas en dos formas diferentes. La prueba siguió *el Formato de Asociaciones de Palabras* propuesto por Read (1993)⁸, pidiendo a los participantes que seleccionaran el sinónimo de la palabra objetivo en la casilla de la izquierda, luego una palabra que pudiera entrar en combinación con la misma en la de la derecha. Los participantes podían encontrar más de una opción correcta. Las respuestas constaban de tres posibilidades: un sinónimo y una colocación, dos sinónimos y dos colocaciones o tres sinónimos y tres colocaciones. Esta combinación aleatoria de respuestas podía evitar que los participantes adivinaran el uso de las palabras. En la segunda parte, los participantes tenían que elegir una de las palabras sinónimas dadas completando el hueco de cada frase. El modelo del test se muestra abajo:

Parte I.

Prospección

proceso sondeo compra explotación	petrolera medicina indirecta amigable
-----------------------------------	---------------------------------------

Parte II.

saludable/sano

1. Nadie defendería en su _____ juicio una situación en que los trabajadores perderían periódicamente todos sus derechos.

2. El 75% de los consumidores (un 2% menos que en 2016) considera que sus hábitos son _____ y que la información nutricional presente en el etiquetado es idónea, destacando como la más valorada la información sobre el contenido en grasas, seguido de los azúcares y de las proteínas.

⁸ El test en formato de asociación léxica tiene una alta fiabilidad para examinar la profundidad del conocimiento del vocabulario (Read, 1993; Qian, 1999).

Las respuestas correctas para cada uno puntuaron 1 punto y las respuestas incorrectas, a su vez, 0 punto. Así, la puntuación máxima de cada prueba inmediata fue de 24. El tiempo para la realización del test fue de 20 minutos.

3.2. Prueba final

En cuanto a la prueba final, su implementación estuvo dirigida principalmente a medir el efecto de la retención del conocimiento del vocabulario durante un período prolongado. La prueba fue adaptada a la escala de calificación del conocimiento del vocabulario (*Vocabulary Knowledge Scale* en inglés, Wesche y Paribakht, 1996). Los participantes realizaron una autoevaluación sobre las diez palabras aprendidas. Específicamente, había cuatro escalas y cada escala representaba puntuaciones diferentes, tal y como se ilustra a continuación:

- A. No recuerdo haber aprendido la palabra (0 punto)
- B. Recuerdo haber aprendido la palabra, pero no sé lo que significa (1 punto)
- C. Conozco la palabra y sé su significado, pero no estoy seguro, el significado es _____ (2 puntos)
- D. Estoy familiarizado con la palabra y sé su significado, y su significado es _____ (3 puntos)

Si los participantes elegían la opción C y acertaban en su significado, recibirían 3 puntos; si elegían la opción D, pero no escribían correctamente el significado de la palabra, otorgarían 1 punto.

3.3. Encuesta

Al terminar el test inmediato de la segunda parte de la clase, se solicitó a los participantes que llenaran un cuestionario acerca de su evaluación y reflexión sobre el método de ABD. La encuesta constaba de dos partes. En la primera parte se les presentaban 15 preguntas con una escala de tipo Likert de 5 puntos, según la cual se calificó su acuerdo o desacuerdo sobre el contenido de la pregunta. Por ejemplo, totalmente de acuerdo/muy satisfecho = 5; de acuerdo/satisfecho = 4; ni de acuerdo ni en desacuerdo = 3; en desacuerdo/insatisfecho = 2; muy en desacuerdo/muy insatisfecho = 1. Las preguntas se podían agrupar en 3 categorías: a. Uso del modelo de autodescubrimiento en el aprendizaje (por ejemplo: Cuando encuentras una palabra desconocida, ¿inferes el significado según el contexto en que aparece?); b. evaluación sobre el aprendizaje de palabras en contextos (por ejemplo: ¿Crees que la recurrencia de la palabra en múltiples contextos favorece la memorización?); c. evaluación sobre la validez de la aplicación del método ABD (por ejemplo: ¿Crees que la duración de la clase experimental fue demasiado corta y no aprendiste todas las palabras?).

La segunda parte de la encuesta estaba compuesta por cuatro preguntas abiertas, en las que nuestros encuestados tenían la oportunidad de comentar sobre las ventajas y desventajas del método ABD. Las preguntas eran: 1. ¿Qué más aprendiste de la observación de las concordancias en distintos contextos? 2. En comparación con el método tradicional basado en diccionarios, ¿cuáles son las ventajas y desventajas del método basado en corpus? 3. ¿Quieres aplicar este método de aprendizaje en futuros estudios? Justifica la respuesta. 4. Si se aplica el método en el aula, ¿cuáles son los aspectos que quedan por mejorarse?

Procedimiento

El estudio intentó comparar el efecto de la enseñanza y el aprendizaje del vocabulario mediante dos métodos distintos: uno basado en corpus y el otro, basado en el diccionario. Para poner de manifiesto las diferencias con mayor claridad y evitar las influencias de los sujetos, los tratamientos experimentales se realizaron en el mismo grupo.

En la primera parte de la clase, los sujetos aprendieron el vocabulario siguiendo el método tradicional, según el cual el profesor mostraba primero la definición de la palabra clave, luego explicaba sus reglas de uso con precisión (cómo se conjuga, con qué preposiciones se combina, en qué locuciones suele aparecer, etc.) y daba ejemplos sacados del diccionario. Finalmente, permitía a los estudiantes hacer oración por sí mismos para favorecer la memorización. Todo el proceso duró 30 minutos. Al final del aprendizaje, los estudiantes le devolvieron el material al profesor e inmediatamente participaron en una prueba.

Tras una pausa de 10 minutos, en la segunda parte de la clase se aplicó el ABD en el aprendizaje de las palabras del grupo experimental. Tras una breve introducción, se distribuyó a los estudiantes el material preparado de antemano. Se pidió a los estudiantes que averiguaran las propiedades léxicas, los significados y las principales colocaciones de las palabras clave y que las utilizaran para hacer frases. Durante el proceso, que duró 30 minutos, el profesor no dio ninguna explicación, dejando a los estudiantes llevar a cabo el aprendizaje de forma independiente. Al terminar el estudio, los estudiantes hicieron otra prueba.

Una semana después de la experimentación, los estudiantes realizaron una prueba con todas las palabras que habían aprendido tanto en clase tradicional como en clase ABD.

Análisis de resultados

Como se mencionó anteriormente, para asegurar la validez de los resultados, en vez de aplicar el método ABD en el grupo experimental y el método tradicional en el grupo de control, decidimos aplicar los dos métodos de aprendizaje en un mismo grupo de estudiantes, dividiendo las palabras pendientes de aprenderse en el grupo experimental y el grupo de control. Habíamos pensado clasificar nuestros participantes en dos grupos (grupo experimental, grupo de control) según la nota que obtuvieron en el examen final del semestre pasado. Sin embargo, la nota como el único criterio de selección no parecía suficientemente convincente porque aparte del nivel de idioma existían otros factores que pudieran repercutir en el resultado, como por ejemplo la motivación de los estudiantes, su actitud hacia el aprendizaje del español, etc. Ahora veamos los resultados.

1. Resultados de la prueba inmediata

En la prueba inmediata tanto en la clase tradicional como en la clase experimental, la puntuación media del grupo experimental y del grupo de control fue de 17.26 y 15.00 respectivamente⁹, siendo el grupo experimental superior al grupo de control. Basado en los datos recolectados en la prueba inmediata, como se muestra en la siguiente tabla, se puede notar claramente que el GE ($M = 17.25$, $SD = 3.612$) logró mayor desempeño que el GC ($M = 15.00$, $SD = 3.181$)¹⁰.

⁹ Recuérdese que la puntuación máxima de la prueba inmediata fue de 24 puntos.

¹⁰ La letra M se refiere a la nota media y la sigla SD, a la desviación estándar. La SD es un concepto estadístico que representa la variabilidad de las observaciones individuales. Cuanto mayor sea la desviación

Sin embargo, esto no mostró realmente si la diferencia entre los datos era estadísticamente significativa, por lo que parecía apropiada una prueba t^{11} de muestras emparejadas basada en la puntuación media de los dos grupos. Pero antes de probar esta hipótesis, era necesario verificar si los datos recopilados satisfacían los supuestos de la prueba t , es decir, el supuesto de normalidad. Realizamos la prueba de normalidad de Shapiro-Wilk¹² en la diferencia de nota entre el GE y el GC, cuyo resultado demostró que los datos se ajustaban a una distribución normal ($p = .760$), tal y como se muestra abajo¹³:

Pruebas de normalidad

	Estadístico	Shapiro-Wilk	
		GL	Sig.
GC-GE	.980	34	.760

El siguiente paso fue realizar la prueba t , cuyo resultado indicó que hubo una diferencia estadísticamente significativa entre el grupo ABD y el grupo basado en el diccionario en la prueba inmediata ($p = 0.003 < 0.05$). Es decir, el método ABD ayudó a los estudiantes a profundizar a corto plazo en el aprendizaje del vocabulario. Véase el resultado en la siguiente tabla¹⁴:

Prueba de muestras emparejadas

Diferencias emparejadas

	Media	SD	Desv. Error promedio	95% de intervalo de confianza de la diferencia		t	GL	Sig. (bilateral)
				Inferior	Superior			
GC - GE	-2.265	4.158	.731	-3.716	-.814	-3.176	33	.003

estándar, mayor será la dispersión de los datos. Así, en comparación con el GC, la SD de las notas obtenidas en el GE es algo superior, lo que indica mayor variación en las puntuaciones individuales.

¹¹ En la estadística, la prueba t puede usarse para determinar si un único grupo difiere de un valor conocido, si dos grupos difieren entre sí, o si hay una diferencia significativa en medidas pareadas.

¹² La selección del tipo de pruebas de normalidad depende del tamaño de la muestra. Si el tamaño de la muestra es grande, se utiliza la prueba de Kolmogorov-Smirnov y, en caso contrario, se utiliza la prueba de Shapiro-Wilk. Aunque no existe un estándar preciso sobre el tamaño de la muestra, el convencionalmente considerado como pequeño es menor de 50 muestras. Para más detalles, consulte la página web en la que se presenta el uso del instrumento SPSS.

<https://statistics.laerd.com/spss-tutorials/testing-for-normality-using-spss-statistics.php>

¹³ Como el presente no es un artículo estadístico, sino que utiliza conocimientos de estadística para el análisis de los datos, para comodidad del lector, solo explicaremos los conceptos estadísticos pertinentes al tema de este estudio. En la siguiente tabla, la sigla GL se refiere a los grados de libertad, que se determinan según el número de observaciones en la muestra. La sigla Sig. se refiere a la significación estadística (a veces representada con la letra p). Por lo general, si el valor es menor o equivalente a 0.05, la diferencia de dos grupos es estadísticamente significativa; al contrario, no existe una diferencia real entre los datos.

¹⁴ Igual que en la tabla de arriba, en esta el lector puede fijarse solo en el dato de la Sig.

2. Resultados de la prueba final

Una semana después del experimento se realizó la prueba final. La puntuación media del grupo experimental y del grupo de control fue de 15.88 y 16.82 respectivamente¹⁵, siendo el grupo de control ligeramente superior al grupo experimental.

Estadísticas de muestras emparejadas

	Media	N	SD	Desv. Error promedio
GC	16.82	34	2.139	.367
GE	15.88	34	2.086	.358

Igual que en el test inmediato, realizamos primero la prueba de normalidad de Shapiro-Wilk, lo que mostró que los datos recopilados de los dos grupos también tenían una distribución normal ($p=0.317$).

Pruebas de normalidad

Shapiro-Wilk

	Estadístico	GL	Sig.
GC-GE	.964	34	.317

El siguiente paso fue realizar la prueba t. Como podemos ver en la siguiente tabla, el resultado indicó que hubo una diferencia estadísticamente significativa entre el grupo ABD y el grupo basado en diccionario ($p = 0.024 < 0.05$). Es decir, el método ABD no mostró ninguna ventaja en términos de memoria a largo plazo. Al contrario, las palabras aprendidas a través del método tradicional guardaron más espacio y duraron más tiempo en el lexicón mental.

Prueba de muestras emparejadas

Diferencias emparejadas

	Media	SD	Desv. Error promedio	95% de intervalo de confianza de la diferencia		t	GL	Sig. (bilateral)
				Inferior	Superior			
GC - GE	.941	2.322	.398	.131	1.751	2.364	33	.024

3. Resultados de la encuesta

¹⁵ Recuérdese que la puntuación máxima de la prueba final fue de 30 puntos.

En primer lugar, se analizaron las actitudes de los participantes sobre el auto-descubrimiento de conocimientos durante el proceso del aprendizaje. Los resultados mostraron que los participantes estaban de acuerdo en que el significado de las palabras podía obtenerse mediante la contextualización y que a menudo la utilizaban para inferir el significado del vocabulario. En cuanto a la clase experimental, esta mostró interés en buscar pistas útiles en los distintos contextos para la descodificación de la palabra clave. Solían acudir a las palabras con que se combinaba la palabra clave para inferir su significado. Además, los participantes pensaban que eran capaces de adquirir los significados de las palabras a través de su propio análisis, y se creían capaces de descubrir conocimientos por sí mismos. Los resultados de la primera parte de la encuesta se presentan abajo:

Preguntas	M	SD
1. ¿El significado se puede obtener mediante el contexto en que aparece?	3.68	0.843
2. ¿Inferes el significado de la palabra desconocida según el contexto cuando lees?	3.68	0.806
3. En la clase experimental, para inferir el significado, ¿has acudido a otras palabras o locuciones a partir del contexto?	3.76	0.741
4. En la clase experimental, ¿has analizado el contexto en que aparecía la palabra clave para inferir el significado?	3.97	0.797

En segundo lugar, veamos los resultados de las actitudes sobre el aprendizaje de palabras en multi-contextos. Los resultados mostraron que los participantes reconocían que las múltiples apariciones de las palabras clave contribuían a la memorización de la palabra. Pero en la clase experimental no aprovecharon debidamente todos los contextos, sino que se limitaron a leer los párrafos cortos o fáciles de entender para inferir el significado. Al captar el significado dejaron de leer otros sin verificar si lo habían comprendido realmente o no. Además, creían que la presencia de múltiples entradas contextuales afectara a su motivación y les hizo sentirse frustrados. Evidentemente, los participantes no estaban del todo acostumbrados a este nuevo método de aprendizaje y carecían de motivación suficiente para el autoaprendizaje.

Preguntas	M	SD
5. ¿Es útil aprender palabras en múltiples contextos?	4.15	0.774
6. ¿Las repetidas apariciones de una palabra en distintos contextos ayudan a la memorización?	3.85	0.958
7. En la clase experimental, ¿usaste todos los contextos para inferir la palabra?	3.59	0.783
8. ¿Utilizaste los contextos para comprobar si la conjetura era acertada?	3.29	0.871

9. La aparición de la palabra en múltiples contextos no garantizó completamente la precisión en la inferencia de significado.	3.65	0.849
10. El método ABD, con párrafos largos y normalmente difíciles, causó estrés a los estudiantes y afectó la motivación.	3.26	0.931

Por último, analizamos la evaluación sobre la validez de la aplicación del método ABD. En cuanto a los materiales de aprendizaje, los participantes consideraban que aparte de la palabra clave, la existencia de otras palabras desconocidas influía negativamente en el resultado del aprendizaje. Para facilitar la lectura, optamos por la forma KWIC para presentar las palabras y evitamos párrafos demasiados cortos o largos. El cuestionario puso de manifiesto uno de los problemas más pertinentes de la aplicación del ABD en el aula: la pérdida de tiempo. La mitad de los participantes no finalizaron el aprendizaje de las palabras del grupo experimental, porque no tuvieron suficiente tiempo.

Preguntas	M	SD
11. La existencia de otras palabras desconocidas en los contextos influyó en el aprendizaje de la palabra clave.	3.28	0.871
12. A veces las frases eran demasiado difíciles de leer, y no pude ni quise inferir el significado mediante el contexto.	3.00	0.816
13. A veces las frases eran demasiado cortas, y no pudieron proporcionar información útil.	3.21	0.808
14. ¿Utilizaste los contextos para comprobar si la conjetura era acertada?	2.91	1.026
15. El tiempo asignado para la clase experimental no era suficiente y no aprendí todas las palabras.	3.74	0.864

Ahora veamos las respuestas de los participantes acerca de su reflexión sobre el método ABD. Para la primera pregunta (¿Qué más aprendiste de la observación de las concordancias en distintos contextos?), la mayoría contestó que además de ayudar a la comprensión del significado de la palabra clave, este método le empujaba a pensar y analizar, voluntaria o forzosamente, más elementos aparecidos en el contexto. A través de la lectura de una serie de párrafos en un tiempo relativamente corto, se concentraban más en el estudio. Sin embargo, 3 participantes confesaron que perdieron la paciencia al encontrarse con párrafos de más de tres líneas y como consecuencia no quisieron seguir leyéndolos. Para la segunda pregunta (¿En comparación con el método tradicional basado en diccionarios, ¿cuáles son las ventajas y desventajas del método basado en corpus?), todos los participantes creían que entendían mejor el significado de las palabras con la ayuda de los contextos si estos eran fáciles y comprensibles. Ocho de ellos afirmaron tener una comprensión más profunda de esas palabras debido a sus repetidas apariciones en el contexto. La mitad opinó que este método ayudaba mucho en la discriminación de los sinónimos.

En cuanto a las desventajas, muchos se quejaron del tiempo asignado para la clase experimental. 18 participantes decían que no habían podido leer todas las frases

formuladas en el material. Esta sensación también fue evidenciada por el docente: en la clase tradicional, casi todos los participantes terminaron el aprendizaje de las palabras con anticipación mientras que en la clase experimental todos se quedaron trabajando hasta el último minuto. Un comentario interesante fue que “con el método ABD, sí pude entender el significado de la palabra. Pero como no tenía un diccionario, no pude asegurarme de su traducción al chino y por lo tanto, no pude confirmar si había entendido bien. El método tradicional me dio más seguridad.”

Para la tercera pregunta (¿Quieres aplicar este método de aprendizaje en futuros estudios? Justifica la respuesta.), la opinión fue divergente: 20 participantes afirmaron que querían aprovechar el método ABD en el futuro porque era favorable para la memorización de las palabras, para fortalecer su paciencia, o parecía más “especializado” en el aprendizaje; otros 14 contestaron que no querían seguir el método porque les costaba mucho tiempo acostumbrarse y leer tantas frases largas sin poder entenderlas les hacía sentir frustrados. Para la última pregunta, los dos aspectos más mencionados por los participantes fueron: 1. el tiempo debería planificarse mejor; 2. el material debería simplificarse.

Observaciones de los datos

Este estudio ha explorado el uso del ABD con mediación del profesor en la enseñanza del vocabulario en el aula de ELE.

Según los análisis cuantitativos de las dos pruebas posteriores y una evaluación del cuestionario, se ha descubierto que el método ABD es más efectivo en la memorización de las palabras a corto plazo. En cuanto a la memorización a largo plazo, en nuestro estudio el método ABD no tiene ninguna ventaja en comparación con el método tradicional basado en diccionario. Creemos que esto se debe principalmente a la incoherencia en la forma de aprender y de evaluar. En la clase experimental, los alumnos solo disponían del material en español, y sin la guía del profesor ni un diccionario para consultar, se limitaban a inferir el significado a través del contexto en que aparecen, o de forma intuitiva, o reflexiva, como consecuencia no estaban completamente seguros del significado de las palabras. En la prueba final, examinamos el grado de familiaridad del vocabulario por parte de los alumnos, pidiéndoles que escribieran su traducción al chino si pensaban conocer la palabra. Si en el momento del aprendizaje no sabían exactamente qué significaba la palabra, resultaba más difícil acertar su traducción en el test. Al contrario, el grupo de control aprendió las palabras mediante diccionarios tanto monolingües como bilingües, una vez dominadas, los alumnos siempre pudieron escribir correctamente su traducción al chino.

A pesar de que en la prueba final no se lograron resultados satisfactorios, los participantes tuvieron una actitud positiva frente a este nuevo método. La mayoría estaba de acuerdo en su utilidad en la comprensión de las palabras clave y en el fortalecimiento de la memorización (aunque a corto plazo). También mostraron interés en su futura aplicación en el aula, siempre y cuando los materiales pudieran ser más simplificados. Cabe destacar que este estudio se llevó a cabo con estudiantes universitarios de nivel intermedio de español, que no tenían experiencia previa en el aprendizaje con corpus, por lo que encontraron cierta dificultad al leer las líneas de concordancia con palabras que iban más allá de su nivel de competencia. Aunque en el proceso de elaboración de los materiales intentamos elegir frases relativamente fáciles, no pudimos evitar palabras consideradas de nivel avanzado. Y esto es común en la comunicación cotidiana, en que no siempre entendemos cada frase y cada palabra que escuchamos, y ahí es donde

tenemos que movilizar nuestras habilidades para poder interpretar el significado de las palabras. Creemos que con el método ABD se puede desarrollar esta competencia.

Para la futura aplicación del método ABD, hay que superar los siguientes obstáculos. El primero y más importante es el tiempo. El enfoque ABD requiere más tiempo que el enfoque tradicional porque los alumnos necesitan tiempo para observar y analizar un gran número de concordancias para sacar conclusiones de uso. El estudio de Boulton (2010: 547) descubrió que no había diferencias significativas en términos de tiempo empleado entre estos métodos, pero esa conclusión se sacó de una clase sin el control estricto del tiempo. En nuestro caso, la clase experimental se desarrolló estrictamente en 40 minutos, por eso la mayoría de los sujetos no completaron el estudio por falta de tiempo. Además, la preparación del material también exigía más dedicación por parte del profesor. Johns (1991a) tardó ocho horas en preparar una sola copia del material de estudio. En nuestro caso, también nos encontramos con el mismo problema. El segundo obstáculo se refiere al rol del docente. El ABD pone más énfasis en el papel central del alumno durante el proceso de aprendizaje, estando más “centrado en el alumno” (Mukherjee, 2006: 12). Por eso, en nuestra experimentación no hicimos ninguna intervención durante el proceso de aprendizaje por parte del alumno. Los datos obtenidos en el test final parecían poner de manifiesto que sin la adecuada explicación, corrección y retroalimentación del profesor, los conocimientos no se quedaban en la memoria. En consecuencia, cómo equilibrar la guía del profesor con la autonomía del alumno queda por discutir. Por último, muchos profesores de lenguas extranjeras no han recibido una formación relevante y probablemente se muestran reacios a adoptar un nuevo método como el ABD (Pérez-Paredes, 2005). Por lo tanto, cómo hacer que los maestros se sientan cómodos con la mecánica del análisis del corpus es un problema.

Conclusión

El estudio ha proporcionado evidencia empírica de la efectividad del método ABD en el aula de ELE, lo cual es significativo para su popularización y aceptación. Además, el estudio ha seguido los procedimientos de la investigación empírica basándose en estudios pioneros anteriores (por ejemplo, Cobb 1997 y Boulton 2010). Por lo tanto, el diseño de investigación es metodológicamente sólido.

El estudio ha constatado que el método ABD es eficaz para facilitar el aprendizaje del vocabulario de la lengua extranjera a los alumnos de nivel intermedio, lo que confirma la viabilidad del uso en el aula.

A pesar de que el presente estudio tiene como objetivo examinar la efectividad y eficiencia del ABD, es importante señalar que el objetivo no radica en demostrar que los métodos tradicionales como las actividades basadas en diccionarios no son efectivos en el aprendizaje. En realidad, como una de las herramientas de estudio más tradicionales y clásicas, el diccionario juega un papel fundamental en el aprendizaje y la enseñanza de idiomas. Nuestros resultados en la prueba final también revelaron su importancia. Lo que intentamos es comprobar la viabilidad del método ABD e integrar mejor los dos en la enseñanza y el aprendizaje.

Basados en las experiencias obtenidas en esta investigación, en el futuro vamos a extender la longitud de la experimentación para ver si las actividades ABD son más beneficiosas a medida que los alumnos se familiaricen con la lectura de párrafos largos y con la guía del profesor.

Bibliografía

ABAD CASTELLÓ, M. 2019. Uso de corpus lingüísticos por y para profesores de español como lengua extranjera. *redELE*, número 31. Obtenido de: <https://www.educacionyfp.gob.es/dam/jcr:39778ef0-dddf-422f-b3c2-a8f08730225b/07-uso-de-corpus-linguisticos.pdf>

ALLAN, R. 2006. *Data-driven learning and vocabulary: Investigating the use of concordances with advanced learners of English*. Trinity College Dublin.

ASENCIÓN-DELANEY, Y., COLLENTINE, J.G., COLLENTINE, K., COLMENARES, J., & 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.

BERNARDINI, S. 2016. Discovery Learning in the language-for-translation classroom: corpora as learning aids. *Cad. Trad*, 36(1), 14-35.

BLANCO, J.M. 2011. Principios metodológicos de la enseñanza de E/EL en contextos sino hablantes. *SinoELE*, 5, 60-81.

BOULTON, A. 2009. Data-driven learning: reasonable fears and rational reassurance. *Indian Journal of Applied Linguistics*, 35(1), 81-106.

BOULTON, A. 2010. Data-driven learning: taking the computer out of the equation. *Language Learning*, 60(3), 534-572.

CHAO, W. 2015. A Corpus-based Study of Spanish Synonymous Verbs. *Journal of Yangzhou College of Education*, 33, 53-56.

COBB, T. 1997. *From concord to lexicon: Development and test of a corpus-based lexical tutor*. [Unpublished doctor's thesis]. Concordia University, Montreal, Canada.

CRUZ PIÑOL, M. 2012. *Lingüística de corpus y enseñanza del español como segunda Lengua*. Madrid: Arco Libros.

DASKALOVSKA, N. 2015. Corpus-based versus traditional learning of collocations. *Computer Assisted Language Learning*, 28(2), 130-144.

GILQUIN, G. & GRANGER, S. 2010. How can data-driven learning be used in language teaching?. In A. O'Keeffe & M. McCarthy (eds) *The Routledge Handbook of Corpus Linguistics*. London: Routledge, 359-370.

JOHNS, T. 1990. From printout to handout: Grammar and vocabulary teaching in the context of datadriven. *CALL Austria*, 10, 14-34.

JOHNS, T. 1991a. Should you be persuaded: Two examples of data-driven learning. *English Language Research Journal*, 4, 1-16.

JOHNS, T. 1991b. From printout to handout: Grammar and vocabulary teaching in the context of data-driven learning. *English Language Research Journal*, 4, 27-45.

LENKO SZYMANSKA, A. 2014. Is this enough? A qualitative evaluation of the effectiveness of a teacher-training course on the use of corpora in language education. *ReCALL*, 26(2), 260-278.

LU, H.C., HUNG, S.Y. y LU, L.H. 2016. La aplicación de un corpus de aprendices en la autocorrección de composiciones escritas. *Porta Linguarum*, 26, 149-160.

MUKHERJEE, J. 2006. Corpus linguistics and language pedagogy: The state of the art—and beyond. In S. Braun, K. Kohn & J. Mukherjee (Eds.), *Corpus technology and language pedagogy: New resources, new tools, new methods*. Peter Lang Publishing, 5-24.

PÉREZ GIMÉNEZ, M. 2015. Propuestas didácticas para estudiar la sintaxis coloquial en Bachillerato: las “construcciones suspendidas”. *Tejuelo*, 22, 215-234.

PÉREZ PAREDES, P. & 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*. Obtenido de: http://eprints.rclis.org/32209/1/patrones_pensamiento_computacional_preprint.pdf

QIAN, D. 1999. Assessing the roles of depth and breadth of vocabulary knowledge in reading comprehension. *Canadian Modern Language Review*, 6(2), 283-307.

READ, J. 1993. The development of a new measure of L2 vocabulary knowledge. *Language Testing*, 10(3), 355-371.

THOMPSON, P. 2006. Assessing the contribution of corpora to EAP practice. In Z. Kantaridou, I. Papadopoulou & I. Mahili (Eds.), *Motivation in learning language for specific and academic purposes*. Macedonia: University of Macedonia.

TSUI, A. 2004. What Teachers have always wanted to know - and how corpora can help. In J. Sinclair (Ed.), *How to use Corpora in Language Teaching*. John Benjamins Publishing Company, 39-61.

VINCZE, O. 2015. *Learning multiword expressions from corpora and dictionaries*. University of A Coruña, Galicia, Spain.

VYATKINA, N. & BOULTON, A. 2017. Corpora in language learning and teaching. *Language Learning & Technology*, 21(3), 1-8.

WEI, W.Q. 2020. *Propuesta de aplicación didáctica de los corpus paralelos en el aula de ELE para los estudiantes sinohablantes*. TFM, Universidad de Barcelona.

WESCHE, M. & PARIBAKHT, S.T. 1996. Assessing Second Language Vocabulary Knowledge: Depth Versus Breadth. *The Canadian Modern Language Review*, 53(1), 13-40.

ZHANG, B.Z. & ZHOU, J.L. 2012. Study on the classroom implementation pattern of data-driven learning (DDL). *Foreign Languages and Their Teaching*, 3, 41-44.

ZHEN, Y., JIMENEZ CRUZ, I. & MONTEJO LORENZO, M.N. 2020. La enseñanza del español como lengua extranjera en China y el desarrollo del profesorado. *Transformación*, vol.16,152-167.

ZHU, H.M. 2011. Data-driven Learning: A New Trend of College English Vocabulary Learning and Teaching. *CAFLE*, (01), 46-50+59.