

De las Humanidades Digitales a la enseñanza de ELE

Aportes y aplicaciones

Humanidades Digitales (HD) y enseñanza de ELE

La extraordinaria expansión de dispositivos móviles con acceso a Internet durante la última década ha propiciado y acelerado la comunicación digital. Las Humanidades, al igual que otras disciplinas académicas, también han sido permeables a este proceso de digitalización constitutivo del denominado cuarto paradigma científico. Aunque su definición (e incluso su definibilidad) sea un territorio de desacuerdo, no parece posible negar el impacto que las tecnologías y la cultura digital tienen y tendrán sobre la práctica humanística.

En ese sentido, este taller introduce herramientas sencillas y susceptibles de integrarse al aula, a la vez que invita a profundizar en la reflexión sobre las posibilidades que la perspectiva digital ofrece a las humanidades.

Corpus

Un concepto relevante es el de *corpus*. Como su etimología lo indica, un corpus es un cuerpo o un conjunto de datos textuales, lo más extenso y ordenado posible, que puede servir como base para una investigación (DRAE, *corpus*²). Generalmente contiene cientos o miles de textos, y millones de palabras.

Considérese, por ejemplo, el *Corpus Thomisticum*, un proyecto iniciado en 1940 por el sacerdote Roberto Busa y que se identifica como el puntapié inicial de las HD: contiene 179 obras e indiza más de 10 millones de palabras (¡originalmente utilizando tarjetas perforadas!). CREA y CORPES XXI, los *corpus de referencia* de la RAE, son colecciones de cientos de miles de documentos (aproximadamente 500.000 entre ambos), con cientos de millones de formas.

¿Cómo producir conocimiento a partir de un *corpus* de esas dimensiones? Si una persona dedicara ocho horas diarias, tardaría más de ocho años en finalizar la lectura del CORPES XXI. Se hace necesario recurrir a formas de "lectura distante". Una forma sencilla, pero potente, de hacerlo es *calcular la frecuencia de las palabras* y representarla visualmente mediante **nubes de palabras**. En este tipo de visualización, la frecuencia de un término se refleja en su tamaño: las palabras más usadas son más grandes. Y no se necesitan millones de palabras para utilizar nubes de palabras: son también una forma creativa de comentar y mejorar las redacciones en ELE.

Crear nubes de palabras es extremadamente sencillo. Existen herramientas de escritorio, como [Wordle](#), u online como [nubesdepalabras.es](#).

Considerando esta nube de palabras, creada con la redacción de una estudiante (nivel A2) sobre un viaje, ¿qué aspectos del texto pueden observarse con claridad? ¿Qué tipo de sinónimos y otros recursos lingüísticos necesitaría la estudiante para repetir menos palabras?



María
Martínez Casas
(DSV Bayern)

Nigel
Manchini
(ANEP, Uruguay)

*XXII Jornadas para el
profesorado de ELE.*
Instituto Cervantes
de Múnich.
15 de junio de 2024

Acceso al
material digital



[www.
bit.ly/
de-HD-a-ELE](http://www.bit.ly/de-HD-a-ELE)

Textos introductorios

Galina y Priani (2020)
[¿Qué son las
humanidades digitales?](#)
del Río y González (2015)
[Introducción a las
Humanidades Digitales.](#)
Boyd (2017)
[Psychological text analysis
in the digital humanities.](#)

Las emociones en palabras: emoFinder y la valencia

La herramienta *emoFinder* permite buscar propiedades subjetivas de palabras utilizando bases de datos baremadas. Dispone de más de 30.000 palabras, anotadas por múltiples evaluadores, en variables subjetivas como Valencia, *Arousal*, Concreción, Edad de Adquisición, Familiaridad y Asociación con emociones discretas (alegría, ira, tristeza, etc.).

Centrémonos en la *valencia* o tono hedónico, una de las dimensiones básicas del afecto, vinculada al placer o displacer asociado a una experiencia o (en este caso) una palabra. Por ejemplo, una palabra como “felicidad” tiene una valencia muy positiva, ya que está asociada a experiencias agradables, mientras que “tristeza” u “horror” tienen una valencia negativa.

Dentro del aula, *emoFinder* puede ayudar a potenciar la dimensión emocional al trabajar con vocabulario. Una forma sencilla de hacerlo es explorar la afectividad dentro de distintos campos semánticos: ¿qué comidas, qué miembros de la familia, qué actividades, qué animales *tienen connotaciones más positivas*? Para el ejemplo, averiguaremos la valencia de algunos términos relacionados con ELE: “lengua”, “curso”, “español”, “enseñanza”, “universidad”, “aprendizaje”, “examen”. Para tener una referencia, también veremos las palabras “felicidad” y “tristeza”.

Lo primero que tenemos que hacer es seleccionar la **Base de Palabras** que vamos a utilizar (en *Acerca de* se encuentra la referencia a cada una de ellas). En este caso, marcaremos la de Stadthagen-González et al. (2017), una base de datos con más de 14.000 palabras.

Posteriormente, debemos seleccionar en qué **Variables** obtener valores; en este caso: *valencia_m*. Como vemos, el valor va desde 1 a 9.

Al buscar así, obtendremos los valores para *todas* las palabras de la base de datos seleccionada. Pero, como nos interesan palabras específicas, podemos escribirlas en el recuadro **Consultar palabras**. También podríamos, si lo necesitáramos, completar otros Parámetros. Una vez realizado esto, podemos comenzar la búsqueda.

The screenshot shows the emoFinder interface. On the left, under 'Bases de Palabras', the 'stadthagen_gonzalez_2017' base is selected. On the right, under 'Variables', 'valencia_m (1 - 9)' is selected. Below these, there is a 'Consultar palabras' field containing the words 'lengua, curso, español, enseñanza, universidad, aprendizaje, examen, tristeza, felicidad'. To the right of this field are 'Parámetros' for 'Número de letras' and 'Patrón de palabra', both currently empty.

Estudiar una lengua tiene aspectos muy positivos, pero también implica aspectos con connotaciones negativas: como vemos, aunque tener que rendir *examen* se parece a la tristeza, el *aprendizaje* en sí se asemeja a la felicidad.

Para familiarizarse con *emoFinder*, les invitamos a:

- Elegir algunos campos semánticos (actividades, familia, animales, lenguas) y explorar su valencia.
- Buscar palabras de 8 letras, finalizadas con “ar” y con valencia muy positiva.

9 / 9 resultados

Palabra	stadthagen_gonzalez_2017 valencia_m x
tristeza	1.80
examen	3.05
lengua	5.70
curso	5.80
español	6.70
enseñanza	6.85
universidad	6.90
aprendizaje	7.05
felicidad	8.40

Fraga, I., Guasch, M., Haro, J. Padrón, I., Ferré, P. (2018). *EmoFinder: The meeting point for Spanish emotional words. Behavior Research Methods* 50, 84–93.

Corpus de referencia de la Real Academia Española

El Corpus de Referencia del Español Actual (CREA) y el Corpus de Referencia del Español del Siglo XXI (CORPES XXI) son *corpus* desarrollados por la Real Academia Española. Cuentan con una interfaz sencilla que permite su utilización sin necesidad de conocimientos especializados. CREA 3.2 incluye más de 135.000 documentos producidos entre 1975 y 2000. CORPES XXI 1.1, por su parte, incluye más de 380.000 documentos y más de 410 millones de formas ortográficas, procedentes de textos escritos y transcripciones orales de este siglo. Ambas herramientas nos permiten obtener estadísticas, concordancias, co-ocurrencias y distribuciones de palabras.



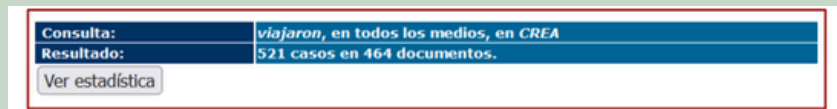
Para familiarizarnos con CREA, volvamos a nuestro primer ejemplo: los viajes. Nos interesa conocer la perspectiva de pasado de la forma de indefinido “viajaron”.

Comenzamos escribiendo la palabra en el recuadro **Consulta**. Podemos refinar la búsqueda, pero una palabra es suficiente para comenzar.

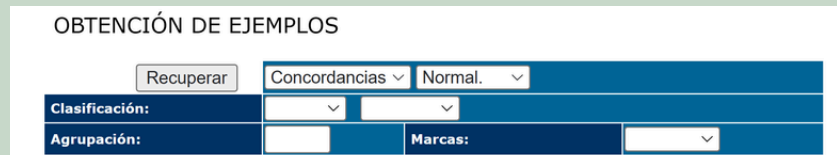
Al presionar “Buscar” obtendremos el primer resultado: el número de **casos** de esa palabra en CREA, y el número de **documentos** en que aparece. Este dato nos permite tener una aproximación a la frecuencia del término.

Podemos presionar “**Ver estadística**” para obtener más detalles acerca del año, el país y los temas en que esa palabra aparece.

Si nos interesa conocer el contexto o desambiguar, podemos **obtener ejemplos**. CREA permite recuperar documentos, párrafos o **líneas de concordancia** (el contexto inmediato de una palabra, con algunas palabras antes y algunas después).



Estadísticas									
Año	%	Casos	País	%	Casos	Tema	%	Casos	
1997	25.77	124	ESPAÑA	41.63	199	5.- Ocio, vida cotidiana.	34.16	178	
1996	11.85	57	ARGENTINA	10.66	51	3.- Política, economía, comercio y finanzas.	33.39	174	
1995	8.10	39	MÉXICO	7.74	37	4.- Artes.	9.78	51	
2001	7.90	38	CUBA	5.43	26	7.- Ficción.	7.67	40	
2000	6.86	33	EE. UU.	5.02	24	2.- Ciencias sociales, creencias y pensamiento.	6.90	36	
2004	6.65	32	CHILE	4.81	23	1.- Ciencia y Tecnología.	4.03	21	
2002	5.61	27	COLOMBIA	4.81	23	6.- Salud.	2.49	13	
1994	4.57	22	VENEZUELA	4.39	21	9.- Oral.	1.53	8	



CONCORDANCIA
 y después de tanta lluvia quizás hasta los que no viajaron les hubiese venido bien estar aquí
 umnos de sexto curso del colegio público “Flavia” viajaron a Barcelona para concursar en el p
 piadas y diversas pruebas de habilidad. Los niños viajaron hasta Barcelona en avión y regres
 or con respecto a los que el pasado fin de semana viajaron a Gijón. El Real Mallorca por su p
 agradecimiento a los ochenta miembros del coro que viajaron a Berlín, así como a todos aquéllos

Tenemos conocimiento suficiente para abordar empíricamente una pregunta interesante: *¿es más frecuente la forma del indefinido o del imperfecto del verbo viajar?*

Podemos, simplemente, comparar la cantidad de casos y documentos en que aparece cada una de esas conjugaciones. En la tabla completamos los verbos “hacer”, “tener” y “decir”, en ambos pasados en la tercera persona plural. Les invitamos a buscar el verbo “pensar” y tres verbos más.

- ¿Qué pasado es más frecuente en cada caso?
- ¿Las frecuencias coinciden con tu intuición inicial?
- En caso de que algún caso les resulte extraño, pueden explorar las concordancias.

hacer	hicieron	hacían
Casos	14797	7916
Documentos	7815	2899
tener	tuvieron	tenían
Casos	9646	17733
Documentos	5741	6550
decir	dijeron	decían
Casos	7357	4459
Documentos	4417	1896
pensar	pensaron	pensaban
Casos		
Documentos		
Casos		
Documentos		
Casos		
Documentos		
Casos		
Documentos		

CORPES XXI: diferencias léxicas en español peninsular y rioplatense

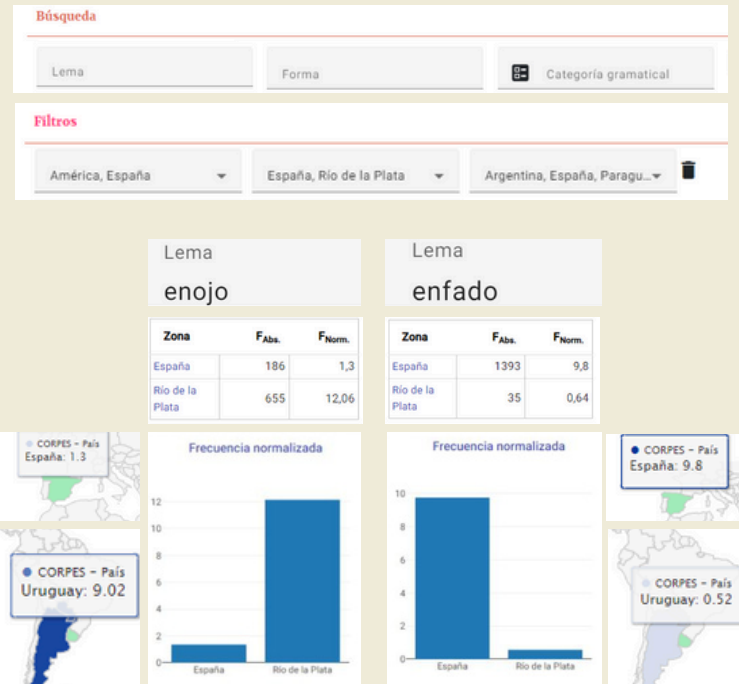
¿Qué diferencias existen entre las distintas variantes del español? ¿Qué términos son distintivos? Abordar este tipo de problema enriquece las aulas desde una visión que reconoce el pluricentrismo del español. CORPES XXI es una herramienta adecuada para hacerlo de manera simple e interactiva.

En nuestro segundo ejemplo ya trabajamos con algunas emociones. Sigamos explorándolas. Para familiarizarnos con CORPES XXI, utilizaremos otras: “enfado”, “enojo” y “amor”. También exploraremos toda una categoría gramatical asociada a las emociones: las *interjecciones*.

Comencemos por la búsqueda: CORPES XXI permite buscar por forma (ej. “enojada”), pero también por *lema* (según el DRAE, *palabra que encabeza un artículo de diccionario*, ej. “enojo”) o por categoría gramatical.

CORPES XXI permite filtrar considerando distintos criterios, como la zona lingüística, el año o el área temática. En este caso, vamos a filtrar las variantes del español de quienes escriben este material: peninsular y rioplatense.

Al presionar “Buscar”, el primer resultado que obtenemos es la frecuencia. Obtenemos la **frecuencia absoluta (F Abs.)**, el número de veces que el lema o la forma aparece en el *corpus*, y la **frecuencia normalizada (F Norm.)**, el número de veces que aparece por millón de palabras en esa región. Esta última es la cifra que conviene utilizar para comparar países. También podemos ver esta información en gráficos de barras y en mapas interactivos; en todos los casos, los resultados sugieren que en España se *enfadan*, mientras que en el Río de la Plata se *enojan*.



CORPES XXI ofrece otra información relevante al respecto de las palabras: en qué medio se utilizan generalmente, género y la edad del autor/a, etc. Además, si queremos profundizar, también podemos obtener concordancias e índices estadísticos más refinados, como el MI (*Mutual Information*), que estima la tendencia de dos palabras a aparecer juntas en un mismo texto del *corpus*.

Volvamos al primer gráfico que aprendimos a realizar: la nube de palabras. En la izquierda representamos los verbos, adverbios y adjetivos más asociados a “amar” en España y el Río de la Plata; en la derecha, las interjecciones más comunes en cada variante: ¿Qué interjecciones y asociaciones son comunes y cuáles propias de cada variante?

Verbos, adverbios y adjetivos asociados a “amar” en España y el Río de la Plata



Interjecciones más comunes en España y el Río de la Plata



Para adquirir familiaridad con CORPES XXI, les invitamos a explorar la frecuencia normalizada de algunas palabras que sean frecuentes en un área, pero no en otra, por ejemplo, “heladera vs nevera”, “móvil vs celular” o “valija vs maleta”. Cada una de estas variantes encierra una historia; investigarlas no solo es una forma de adquirir vocabulario reconociendo el carácter pluricéntrico de la lengua española, sino también de acercarse a su dimensión histórica y sociocultural.