

El Diccionario de Colocaciones del Español: una puesta al día

Margarita Alonso Ramos
Universidade da Coruña

1. INTRODUCCIÓN

El objetivo de este artículo es presentar una puesta al día del *Diccionario de colocaciones del español* (DiCE), disponible en la web desde 2004 (www.dicesp.com). Aunque ya hemos descrito el diccionario en varias ocasiones (entre otras, Alonso Ramos 2004a, 2005, 2006, 2008, 2010, Vincze et al. 2011), aquí intentaré presentar los puntos esenciales, antes de centrarme en cuáles son las nuevas funcionalidades o mejoras introducidas, así como en los próximos retos que nos esperan. Por lo tanto, el artículo se organizará en tres secciones principales que puedan satisfacer a todo potencial lector: 1) lo esencial para los que no conozcan el DiCE; 2) lo nuevo para los que ya lo conocen; y 3) lo próximo para todos aquellos interesados en estar al tanto de los avances del DiCE.

2. PUNTOS ESENCIALES DEL DICE

El diccionario académico define *diccionario* como “Libro en el que se recogen y explican de forma ordenada voces de una o más lenguas, de una ciencia o de una materia determinada”. Como veremos, esta definición no es muy apropiada para el DiCE (y para otros diccionarios tampoco). El DiCE no es un libro, sino un diccionario electrónico, aunque sí se recogen “voces”, creo que ordenadamente. Más en particular, se recogen *colocaciones*. Empezaré, por tanto, presentando cuál es objeto de este diccionario electrónico, las colocaciones, para continuar con una descripción de sus componentes.



2.1. Objeto: Colocaciones en un marco teórico específico

El DiCE está enmarcado en un enfoque teórico, la Lexicología explicativa y combinatoria (Mel'čuk et al. 1995, 2012), con su correspondiente concepción de qué entendemos por *colocación*. En la bibliografía española se ha discutido más sobre el término que sobre el concepto colocación y no quisiera entrar en ese debate (Alonso Ramos, 2010). Para lo que aquí interesa, basta decir que una colocación es una unidad fraseológica formada por dos elementos, la *base*, que es seleccionada por su significado y el *colocativo*, que es seleccionado bajo el control de la base. Así, por ejemplo, los nombres *paseo*, *decisión*, *fotografía* se combinan respectivamente con *dar*, *tomar* y *sacar* para expresar aproximadamente el mismo sentido: 'hacer'. En este caso se trata de tres colocaciones verbales, más en particular, tres construcciones con verbos de apoyo (Alonso Ramos 2004b), en donde el nombre constituye la base y el verbo funciona como colocativo[1]. El hecho de que el DiCE esté concebido dentro de este marco teórico tiene como consecuencia que las colocaciones son descritas semántica y sintácticamente. La razón se debe a que este marco teórico dispone de la herramienta de las *funciones léxicas* (FLL, Mel'čuk 1996), que permiten describir para una unidad léxica dada todos sus *correlatos léxicos*, es decir, las elecciones paradigmáticas y sintagmáticas que dependen de elecciones léxicas hechas previamente. Así, por ejemplo, bajo la entrada de *pena*, incluimos el verbo *dar* (correlato sintagmático *pena*, como en *dar pena* [a alguien]) y el verbo *apenar* (su correlato paradigmático). Si queremos registrar en un diccionario las distintas maneras de expresar 'causar una pena', debemos incluir tanto la colocación verbal como el verbo relacionado paradigmáticamente. Siendo entonces más precisa, debería decir que el DiCE es un diccionario de correlatos léxicos, pero no quiero abundar en la inflación terminológica. Gracias a las FLL subyacentes, el DiCE ofrece una descripción semántica y sintáctica de las colocaciones, a diferencia de otros diccionarios (Benson et al. 1986, Crowter et al. 2002 o Bosque 2006) en donde las colocaciones son simplemente agrupadas semánticamente por la proximidad del significado de los colocativos y clasificadas por la clase de palabras del colocativo. Sin embargo, el beneficio de esa descripción no siempre se aprecia, puesto que el DiCE sigue siendo criticado porque las FLL no son lo suficientemente subyacentes (entre otros, Buendía y Faber 2014)[2].

[1] Para una reflexión crítica sobre la distinción entre combinación libre y colocación, véase Bosque (2011) y Alonso Ramos (en evaluación).

[2] Entraré en cuestiones de usabilidad del DiCE un poco más abajo.

Otra de las críticas que suele recibir el DiCE es que no está “acabado” y que se “limita” al campo semántico de los nombres de sentimiento. Sin embargo, un diccionario, tal y como lo concibo, es necesariamente un proyecto en curso: los “acabados” son procesos más vinculados al mundo editorial y económico que el propiamente científico. El DiCE ha pasado por diferentes versiones desde que se puso en la web en 2004 y sigue siendo objeto de diferentes modificaciones tanto en su contenido como en el acceso a la información. La “limitación” a un campo semántico permite además mostrar las regularidades y las posibilidades de generalización que no serían posibles con una muestra de lemas más aleatoria. Con todo, el DiCE proporciona actualmente más de veinte mil relaciones léxicas (valores de FLL). 2.3. *Diccionario electrónico: arquitectura*

Después de presentar someramente el contenido del DiCE, pasemos ahora al continente. El DiCE ha sido concebido como una base de datos electrónica. A diferencia de otros diccionarios electrónicos que han partido de una versión en papel, el DiCE ya ha sido diseñado con una arquitectura que permite explotar al máximo las posibilidades de acceso a la información que da la versión electrónica. Como señala Schubert (2014), el diseño de un diccionario electrónico debe hacerse antes de construir la base de datos interna: dependiendo de qué información se quiera mostrar y de qué manera, se estructurará la base de datos correspondiente. El DiCE tiene una base de datos y dos interfaces, la del lexicógrafo y la del usuario. Aquí solo me centraré en la segunda, la interfaz de acceso público (pero véase Vincze et al. 2011). Consta de tres componentes: el diccionario propiamente dicho, el componente de *consultas avanzadas* y el componente didáctico, todavía embrionario.

El primer componente es más convencional, mientras que el segundo apunta a consultas puntuales. Con el diccionario, se proporcionan los correlatos de una unidad léxica (UL) dada, mientras que con las consultas avanzadas se intenta dar respuestas a dudas específicas. Más abajo, mostraré algunos ejemplos de las consultas que se pueden realizar. En el componente diccionario ([Acceso al Diccionario](#)), el acceso a la información se hace por lemas y cada lema abarca diferentes UULL. A partir de cada UL, el usuario puede consultar la información semántica o la información combinatoria correspondiente. En cuanto a la información semántica, el artículo lexicográfico de cada UL proporciona: 1) la *etiqueta semántica* que representa el significado genérico; 2) la *forma proposicional*, en la que aparecen los participantes de la situación designada por el nombre; 3) ejemplos de corpus, la mayoría del *Corpus de Referencia del español actual*; y 4) los cuasi-sinónimos y cuasi-antónimos (con el número de la UL correspondiente).

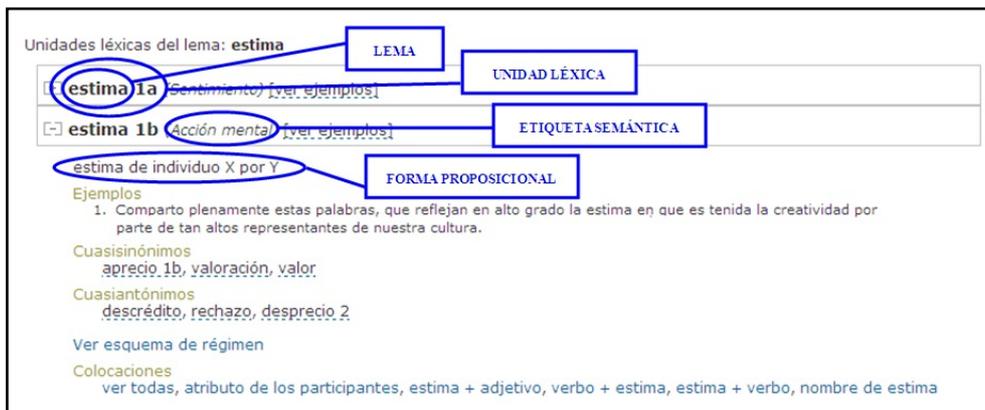


Fig. 1 Zona semántica de un artículo lexicográfico

La información combinatoria ofrecida es de dos tipos: sintáctica, por medio del *esquema de régimen*, y léxica, por medio de las funciones léxicas formateadas en glosas en lengua natural. Con el esquema de régimen se proporciona de forma sucinta la información sintáctica concerniente a la realización lingüística de los actantes. Así, el actante X de *admira* 1 puede realizarse por medio de un sintagma preposicional (*la admiración de Juan*), por un adjetivo posesivo (*su admiración*) o por un adjetivo relacional (*la admiración nacional*). El actante Y se realiza siempre mediante un sintagma preposicional: *admira a/hacia su profesora/ por su trabajo*, etc. Como vemos, el DiCE es también un diccionario de régimen, pero distingue y separa los dos tipos de información combinatoria[3].

» admiración : 1

Esquemas de régimen

Actantes	Realizaciones
1 - X	de N [ver ejemplos] Apos [ver ejemplos] A [ver ejemplos]
2 - Y	a N [ver ejemplos] hacia N [ver ejemplos] por N [ver ejemplos] ante N [ver ejemplos]
3 - Z	por N [ver ejemplos] por Vinf [ver ejemplos]

1. su admiración

Fig. 2. Zona de combinatoria sintáctica

[3] Algunos autores tratan la relación de régimen entre un núcleo léxico y una preposición regida como un tipo de colocación, las *colocaciones gramaticales* (entre otros, Benson et al.1986). Pensamos que este término solo puede dar lugar a confusión.

La combinatoria léxica se muestra en la sección “Colocaciones”. Véase la Fig. 3. del artículo lexicográfico de *admiración*:

ver todas, atributo de los participantes, admiración + adjetivo, verbo + admiración, admiración + verbo, nombre de admiración

Fig. 3. Zona de combinatoria léxica

Como vemos, se pueden hacer cinco tipos de consulta. La primera necesita alguna explicación. Bajo *Atributos de los participantes* agrupamos atributos o nombres que refieren a los participantes de la situación designada; por ejemplo, *digno de admiración* o *admirable* refiere al participante Y, el objeto de la admiración. En el primer caso, hay una relación sintagmática con el nombre *admiración*, mientras que en el segundo, la relación es paradigmática. Obsérvese también que codificamos con diferente rúbrica cuando la base funciona como sujeto (UL+Verbo, *su admiración se dirige a [Y]*) o cuando funciona como objeto o primer complemento (Verbo+UL, *profesar admiración [hacia Y]*). Se trata obviamente de una convención para marcar esta diferencia sintáctica aunque, como sabemos, el orden de palabras en español no tenga un importante papel sintáctico[4]. Véase la siguiente figura con algunas colocaciones verbales de *admiración*:

The screenshot shows the 'admiración 1' entry with a 'Colocaciones' section. The text 'ver todas, atributo de los participantes, admiración + adjetivo, verbo + admiración, admiración + verbo, nombre de admiración' is visible. A blue circle highlights the 'Ordenar alfabéticamente' button. Below, three verb entries are shown: 'tener [~]', 'sentir [~]', and 'rendir [~]', each with a gloss and examples.

Fig. 4. Colocaciones V+*admiración* 1

[4] En cambio, en el *Práctico* (Bosque 2006), no se distingue el papel sintáctico desempeñado por el nombre en una colocación verbal, lo que puede dar lugar a confusión, especialmente a aprendices de español.

Paso ahora a presentar el componente de consultas avanzadas. Desde el inicio del DiCE hemos ido modificándolo con el objetivo de mejorar su usabilidad. Actualmente este componente está organizado en cuatro tipos de consulta: 1) una orientada hacia la comprensión del significado de la colocación; 2) una dirigida hacia la producción de un colocativo correspondiente a un significado dado dependiendo de la base; 3) una que parte del colocativo para proporcionar las distintas bases con que se combina para expresar un sentido o diferentes sentidos; 4) una más dirigida a los especialistas ya que se puede preguntar por una FL en concreto aplicada a una base dada. Podría decirse que las consultas (1) y (3) están más orientadas hacia la comprensión, mientras que la (2) y la (4) están orientadas hacia la producción. Así, por ejemplo, la primera nos sirve para proporcionar la glosa ‘intenso’ que describe el significado de [cerval](#) en relación con *miedo*.

La consulta que pregunta a partir del colocativo es similar a algunas de las entradas del *Práctico* (Bosque 2006). Por ejemplo, desde el colocativo *cerval* se nos proporciona la información de que se combina con *miedo* y también con *pavor*. En donde se saca mayor partido es en las consultas orientadas hacia la producción. Así, por ejemplo, el usuario puede preguntar qué adjetivos expresan ‘intenso’ en relación con el nombre *miedo*. La misma consulta, pero con el metalenguaje de FLL sería: cuál es el valor de la FL Magn(*miedo*).

Por último, presentaré el componente didáctico que, como ya señalé arriba, es embrionario. Se limita a unos pocos ejercicios con el objetivo de ofrecer una muestra de cómo poder integrar actividades didácticas en un recurso lexicográfico. Dicha integración no es siempre fácil. En Alonso Ramos (2006) analicé algunos diccionarios en línea o en CD que ofrecían actividades didácticas con colocaciones y en muchos casos adolecían de problemas de diseño y, lo que es más importante, de una inadecuada concepción de cómo aprender colocaciones. Además de las actividades didácticas, ofrecemos un tutorial orientado a que el usuario adquiera familiaridad con el DiCE y con el concepto de colocación[5]. Véase la Fig.5

[5] En un test de usabilidad del DiCE, del que hablaré más abajo, los participantes que declararon haber realizado este tutorial tuvieron mejores resultados.

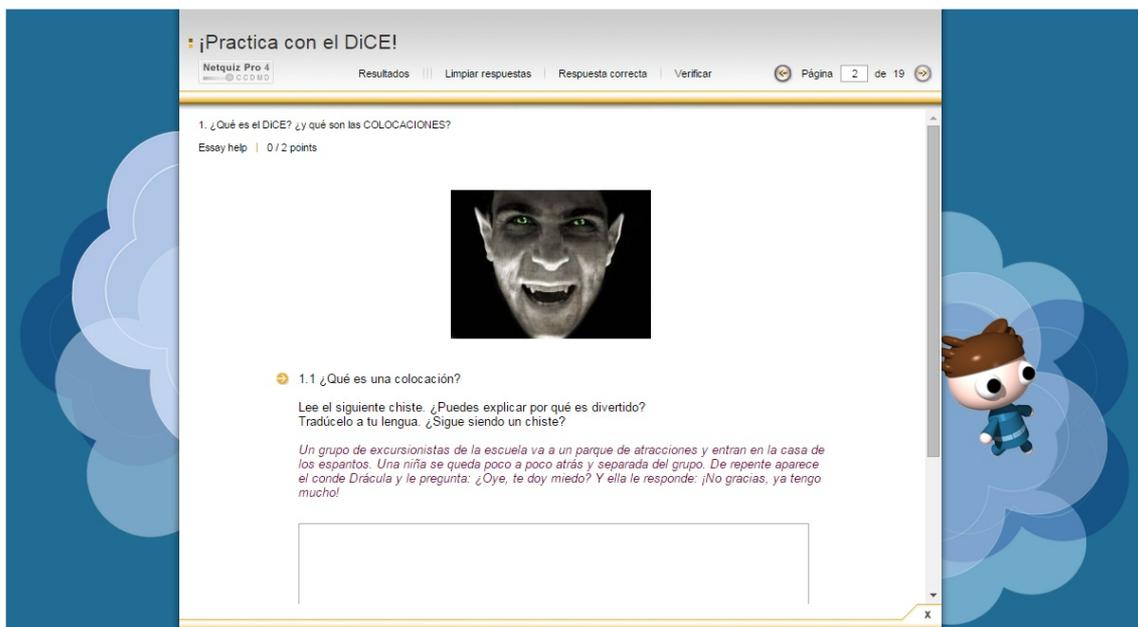


Fig. 5 Muestra de una actividad didáctica del DiCE

A pesar del estado embrionario del componente didáctico, el DiCE ya tiene todo el potencial para integrarse en un entorno de aprendizaje del que hablaré brevemente en la siguiente sección.

3. NUEVAS FUNCIONALIDADES Y MEJORAS EN EL DICE

Como ya avancé al principio, el DiCE está en constante evolución desde su inicio. En esta sección presentaré algunas de los cambios que se han hecho (o se están haciendo) tanto en lo que respecta a la interfaz como al contenido.

Empezando por la interfaz, desde el primer momento, se ha tenido en cuenta mejorar la accesibilidad del DiCE. Con el objetivo de poder evaluar las diferentes opciones de consulta, sometimos al DiCE a un test de usabilidad (Vincze y Alonso Ramos 2013b). Los resultados de dicha prueba mostraron que la calidad global de la interacción entre usuarios y diccionario era buena, aunque también se identificaron áreas de mejora. Nos interesaba especialmente valorar hasta qué punto los usuarios sabían cuándo acceder al módulo diccionario o al módulo de consultas avanzadas que, de hecho, da la respuesta más directa. Los resultados del test señalaron que los participantes, especialmente los de edad más avanzada, se dirigían al componente diccionario, a pesar del rodeo para encontrar la información. Una de las razones para este comportamiento puede ser que esta ruta es la que se da por defecto, además de que es el tipo de acceso más similar a los diccionarios en papel, que es más familiar a

la mayoría de usuarios. Entre las consultas puntuales, la más utilizada es *Qué significa*. De nuevo pensamos que esto se debe a que los usuarios están acostumbrados a consultar un ítem léxico para verificar su significado o su ortografía, pero no lo están para buscar cómo expresar un significado específico. Sin embargo, esta es la principal utilidad de un diccionario de colocaciones. Hay que señalar que entre los distintos grupos de informantes que participaron en la prueba los traductores fueron los que mejores resultados obtuvieron, prueba de que este sector está más habituado que estudiantes o profesores de lengua a utilizar diferentes recursos léxicos. Aunque habría que hacer otro test con informantes que hubieran sido ya familiarizados con el DiCE para ver si los resultados son diferentes, pensamos que en este test ya se trasluce una tendencia general de los usuarios: quieren una herramienta subyacente al proceso de escritura, que no tengan que aprender a utilizar. Volveré sobre esto en la siguiente sección.

El test nos sirvió para detectar diferentes problemas en el diseño actual de la interfaz. Algunas de las mejoras están siendo implementadas en este momento. A continuación paso a presentar las principales. Hemos cambiado los nombres de las consultas avanzadas para ser más transparentes y más ajustados a las expectativas del usuario. La primera *Qué significa* apunta a la comprensión, mientras que *Cómo se dice* apunta a la producción: se parte de una base y después de elegir el tipo sintáctico del colocativo que se busca (N+Adj, V+N, etc.), se despliegan las diferentes glosas correspondientes a cada tipo entre las que el usuario podrá escoger la que más se acerca al sentido que quiere expresar. Vid. la siguiente figura:

The screenshot shows the DiCE website interface. At the top, it says 'Diccionario de colocaciones del Español' and 'DiCE'. There are navigation tabs: 'Bienvenida', 'Acerca del DiCE', 'Acceso al diccionario', 'Consultas avanzadas' (highlighted), and 'Actividades didácticas'. Under 'Consultas avanzadas', there are sub-tabs: '¿Qué significa?', '¿Cómo se dice?' (highlighted), 'Búsqueda por colocativo', and 'Búsqueda por FL'. The main search form has the following fields:

- Base * (unidad léxica opcional): 'admiración' (with a dropdown arrow) and '1' (with a dropdown arrow).
- Tipo de colocación *: '- adjetivo' (with a dropdown arrow).
- Glosa: 'más intensa de lo conveniente' (with a dropdown arrow).

There are 'Buscar' and 'Borrar' buttons. To the right of the form is a box titled '¿Qué puedo buscar aquí?' with the text: '¿Cómo se dice? p.ej. alegría + adjetivo que significa 'que se transmite con facilidad' -> contagiosa'. Below the form, it says 'Encontradas 4 colocaciones' and shows a result for 'más intensa de lo conveniente' (4 valores en total). The result details include the word 'ciega', the glosa 'más intensa de lo conveniente', and two examples: '1. La ciega admiración por el grupo musical Nirvana (R).', '2. Quiero decir que la ciega admiración de nuestro público quizás sea exagerada.'

Fig. 6. Muestra de la consulta *Cómo se dice*

Queremos también mejorar la visibilidad de distintas funcionalidades que ayuden a una navegación más fluida por la web. Así, en el componente diccionario, actualmente cuando se pincha en un lema como [admiración](#), aparece una lista de unidades léxicas que pueden despistar al usuario que está buscando colocaciones de ese nombre. En su lugar, estamos modificando la interfaz para que desde el primer momento se vea el acceso a las colocaciones. Podría ser algo así:



Fig. 7. Visualización inicial de UULL

Otra mejora relacionada con las glosas que estamos implementando se deriva también del test. Comprobamos que algunos usuarios no distinguían correctamente entre glosas y colocativos. En la interfaz actual, al clicar en el enlace [admiración 1+adjetivo](#), por ejemplo, aparece una lista de glosas. Lo que proponemos es cambiar la visualización para que no solo se vea la glosa, sino también algunos de los colocativos. Los colocativos tendrán un hiperenlace para que se pueda acceder a la descripción de esa colocación.

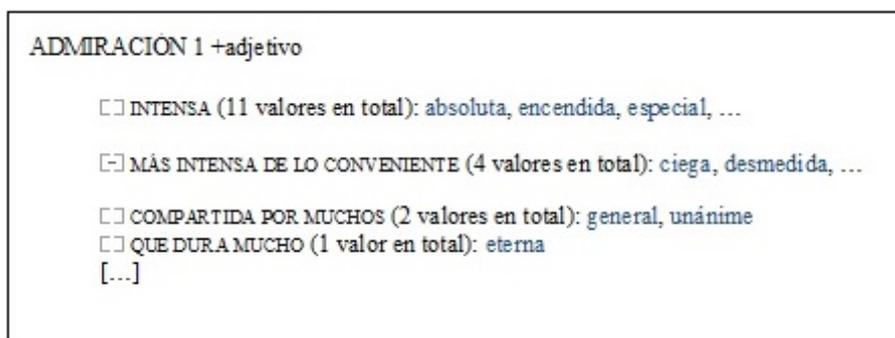


Fig. 8: Visualización inicial de los colocativos de una UL

Los cambios más importantes con respecto al contenido conciernen tres aspectos: 1) la información de frecuencias; 2) el estudio de las marcas de uso de las colocaciones y 3) la información del nivel de ELE. La primera información ya está incorporada en el DiCE, aunque falta darle mayor visibilidad. La obtención de información de frecuencia de colocaciones fue una tarea ardua. Quiero subrayar que hemos asignado frecuencias a sentidos, no a palabras. Hemos optado por asignar un índice de frecuencia a cada UL que constituye la base de la colocación, mientras que asignamos un índice de frecuencia a la colocación entera ponderada por la frecuencia de la base. Cada UL recibe una etiqueta de frecuencia baja, moderada, notable o alta, siguiendo los parámetros de Almela et al. (2005). Los colocativos son presentados en el DiCE bajo la entrada léxica de la base según su tipo sintáctico y su significado, y ordenados por su frecuencia. De esta información, podemos derivar una lista de todas las colocaciones incluidas en el DiCE con información de frecuencia, de importancia vital tanto para propósitos didácticos como evaluativos. La descripción detallada de esta metodología fue presentada en Vincze y Alonso Ramos (2013a). Véase la siguiente figura con la información de frecuencia en cada UL de *admiración*:

[+] Desplegar todo [-] Contraer todo

admiración 1 f. (*Sentimiento*) [ver ejemplos] Frecuencia notable

admiración de individuo X a individuo Y por hecho Z

Ejemplos

1. Juan siente admiración hacia su profesor por su capacidad de trabajo

Cuasisinónimos

adoración 2, aprecio 1a, consideración III.1, devoción 2, entusiasmo 1, respeto I.1a, veneración 2

Cuasiantónimos

aversión 1, desdén 1, desprecio 1

[Ver esquema de régimen](#)

Colocaciones

ver todas, atributo de los participantes, admiración + adjetivo, verbo + admiración, admiración + verbo, nombre de admiración

admiración 2a f. (*Sentimiento*) [ver ejemplos] Frecuencia moderada

admiración 2b f. (*Algo*) [ver ejemplos] Frecuencia baja

Fig. 9. Índice de frecuencia a cada UL

En la siguiente figura se observan los colocativos ordenados por frecuencia, con la posibilidad de ordenarlos alfabéticamente.

admiración 1 (Sentimiento)

Colocaciones
ver todas , atributo de los participantes, **admiración + adjetivo**, verbo + admiración, admiración + verbo, nombre de admiración

Desplegar todo Contraer todo Mostrar funciones léxicas Ordenar alfabéticamente

intensa (11 valores en total)

más intensa de lo conveniente (4 valores en total)

desmedida
Glosa
más intensa de lo conveniente
Ejemplos
1. La desmedida admiración que profesaba el coronel hacia las mujeres (R).
2. En esta desmedida admiración por Leonardo, mago o ingeniero, hay algo más que el reconocimiento de un trabajo conienzudo.

ciega
Glosa
más intensa de lo conveniente
Ejemplos
1. La ciega admiración por el grupo musical Nirvana (R).
2. Quiero decir que la ciega admiración de nuestro público quizás sea exagerada.

exagerada
Glosa
más intensa de lo conveniente
Ejemplos
1. Todo eran frases a medias cuya clave yo únicamente no captaba, y una incesante, exagerada admiración al maestro.
2. El DRAE registra esnobismo como derivado del inglés snob, esclavo de la moda y en consecuencia lo define como exagerada admiración por todo lo que es de moda.

desmesurada
Glosa
más intensa de lo conveniente
Ejemplos
1. Traicionado por mi desmesurada admiración hacia el gran editor que es Jorge Herralde (R).
2. No era el caso discutir con ese maestro por quien sentía una admiración desmesurada.

Fig. 10. Colocativos de *admiración* ordenados por frecuencia

Otra aportación al DiCE aunque todavía no implementada, es la lista de marcas de uso necesarias para las colocaciones. Tras comprobar en los corpus de aprendices que uno de los factores de error en las colocaciones era el uso inadecuado de un registro dado, se emprendió un estudio sobre qué marcas de uso podrían ser interesantes para asignar a las colocaciones en el DiCE. Vázquez Veiga (2014) se ha centrado en cómo asignar las siguientes marcas estilísticas y actitudinales a los datos del DiCE: formal, informal, vulgar, eufemístico, despectivo e irónico. El estudio pone de relieve que el uso de una base o de un colocativo estilísticamente marcado transmite ese carácter a la colocación entera. Queda por implementar una consulta avanzada que permita buscar colocativos con una determinada marca de uso. Así, Vázquez Veiga (2014) propone una búsqueda como la que se muestra en la figura siguiente en donde se pide un colocativo de la base *miedo* con marca “vulgar” y que tenga el sentido ‘morirse’.

Bienvenida | Acerca del DiCE | Acceso al diccionario | **Consultas avanzadas** | Actividades didácticas

Consultas avanzadas : ¿Qué significa? | Ayuda a la redacción | Directas | Inversas | **Marcas de uso**

Base: Colocativo: Marca de uso:

acojonarse VULGAR

Glosa
sentir un ~ intenso

Comentario
expresión fusionada
registro vulgar

Ejemplos

1. Esta vez se acojonaron, les daba miedo que les demandaran (web)
2. Creyó que Emilio Botín estaba empezando a acojonarse y tomó nota para el futuro próximo.

Fig. 11. Búsqueda de colocativos por marcas de uso

En cuanto a la información del nivel de ELE, actualmente estamos volcando la información a la base de datos del DiCE. Asignamos un nivel del *Marco común de referencias europeo* (A1, A2, B1, etc.) a cada UL así como a la colocación. Para ello, nos basamos principalmente en el vaciado de información en el *Plan curricular del Instituto Cervantes* (PCIC). Aunque no todas las UULL ni las colocaciones del DiCE aparecen en el PCIC, hemos elaborado un criterio de derivación basado en frecuencias. Establecemos una correlación entre el nivel asignado en el PCIC y su frecuencia en un corpus de referencia (esTenTen, Kilgarriff y Renau 2013). No ha sido posible, por el momento, encontrar otros criterios didácticos más fundamentados. En los próximos meses presentaremos en una publicación nuestra metodología de nivelación de colocaciones, pero por el momento, se puede mostrar ya cómo podría consultarse en el DiCE la información de nivel. En la figura siguiente, se piden colocativos del nivel C2 seleccionados por la base *alegría*.

Bienvenida | Acerca del DICE | Acceso al diccionario | **Consultas avanzadas** | Actividades didácticas

Consultas avanzadas : [¿Qué significa?](#) | [Ayuda a la redacción](#) | [Directas](#) | [Inversas](#) | [Marcas de uso](#)

Base: Nivel ELE:

alegría 2 f. (Cualidad) [[ver ejemplos](#)] [[ver alegría](#)] (1 valor)

derrochar [~] C2

Glosa
 manifestar una ~ intensa

Ejemplos

1. se supone que es en Navidad donde debemos derrochar alegría, donde debemos derrochar amistad (web)
2. Ser madre es conocer los secretos de la felicidad compartida, sembrar cariño, derrochar alegría (web)

alegría 1a f. (Sentimiento) [[ver ejemplos](#)] [[ver alegría](#)] (1 valor)

desbordante C2

Glosa
 intensa

Ejemplos

1. Opuesto al de la depresión, se caracteriza por la alegría desbordante
2. Observó con sorpresa que todos los detalles nimios de la vida diaria, le producían una alegría desbordante
3. A veces cuando muchos nos ven pasar por los caminos rocieros, con nuestra alegría desbordante, con nuestros cantos y rezos a la virgen

Fig. 12. Búsqueda de colocativos por Nivel ELE

4. PRÓXIMA INTEGRACIÓN EN UNA HERRAMIENTA DE AYUDA A LA REDACCIÓN.

Llegamos a lo próximo. Me detendré ahora en exponer cuáles son los pasos que estamos dando hacia el futuro. Decía arriba que el usuario apuesta por recursos léxicos “invisibles”. En esa línea ya existe alguna *ayuda a la lectura* en donde el usuario clikea en la palabra y se le abre una ventana emergente con la definición de una palabra[6]. A nosotros nos gustaría poder añadir la glosa de la colocación, como presentamos en la siguiente imagen.

[6] Véase el diccionario ClickLex desarrollado por Molino de ideas [<http://www.clicklex.com/>]

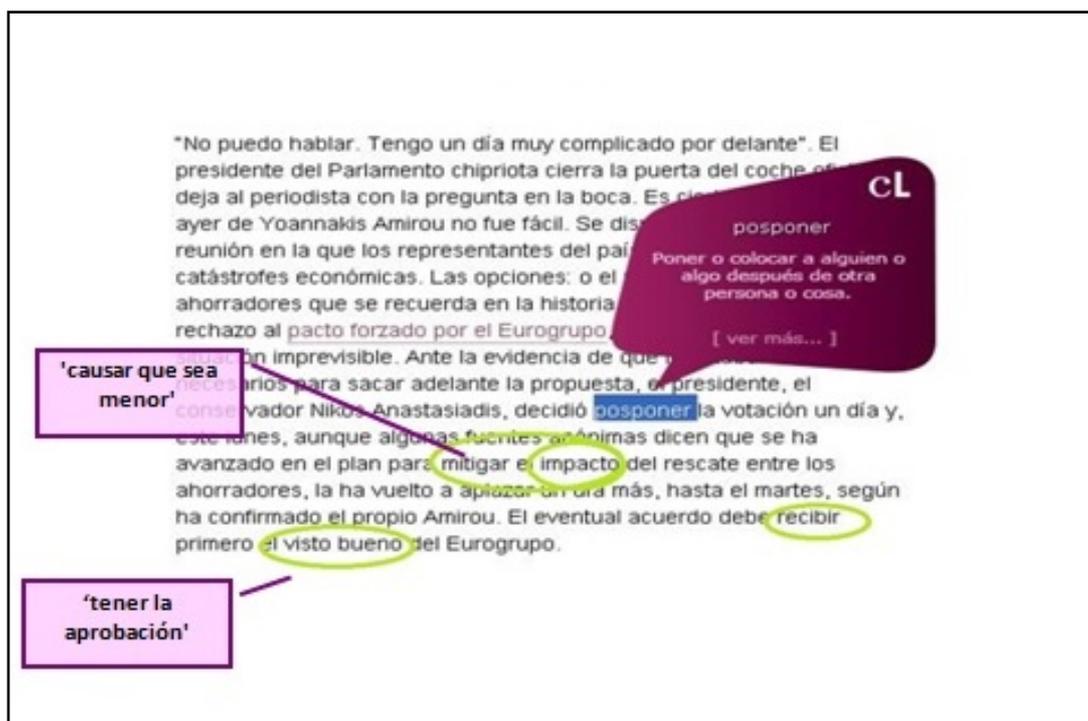


Fig. 13 Ayuda a la lectura en el escritorio

Para llegar a ese reconocimiento de colocaciones, hacen falta técnicas de identificación automática basadas principalmente en medidas estadísticas y en entrenamiento de programas con corpus. El DiCE, dado que contiene ejemplos para cada colocación, es un recurso útil para estas investigaciones puesto que constituye un corpus de colocaciones de casi treinta mil oraciones (Alonso Ramos 2009). Wanner, junto con su equipo, ha utilizado el corpus del DiCE para entrenar sus programas de identificación y clasificación semántica de colocaciones. El corpus del DiCE está ahora etiquetado morfológica y sintácticamente y podemos decir que disponemos de un *treebank* de colocaciones (Moreno et al. 2013).

El concepto de diccionario está cambiando, puesto que sus usuarios lo son principalmente de diccionarios en línea o, siendo sinceros, son usuarios de Google. En todo caso, el usuario espera que el diccionario deje de ser algo estático para convertirse en algo más dinámico e interactivo, que esté vinculado a la web y a otros recursos léxicos (Verlinde et al. 2009, 2012). Con esa idea en mente, estamos construyendo un entorno de ayuda a la escritura centrado en las colocaciones, cuyo módulo didáctico esté directamente vinculado al diccionario y en donde exista conexión directa también con el corpus, sin pasar directamente por el diccionario. Se trata de diseñar una herramienta dinámica que facilite el aprendizaje autónomo y

guiado por los datos (en inglés *data-driven learning* con el término acuñado por Johns 1991), tal y como propugnan en los últimos años los estudiosos sobre aprendizaje de lenguas segundas (entre otros, Geluso y Yamaguchi 2014). Nuestra propuesta es una herramienta de ayuda a la redacción centrada en colocaciones (HARenES[7], García-Salido et al. 2014). Esta herramienta integra el DiCE, pero también un diccionario personalizado del usuario, así como la consulta directa al corpus sin pasar por el diccionario. Aunque todavía no está completamente operativa, su arquitectura está ya diseñada. Véase la siguiente figura con sus principales funcionalidades.

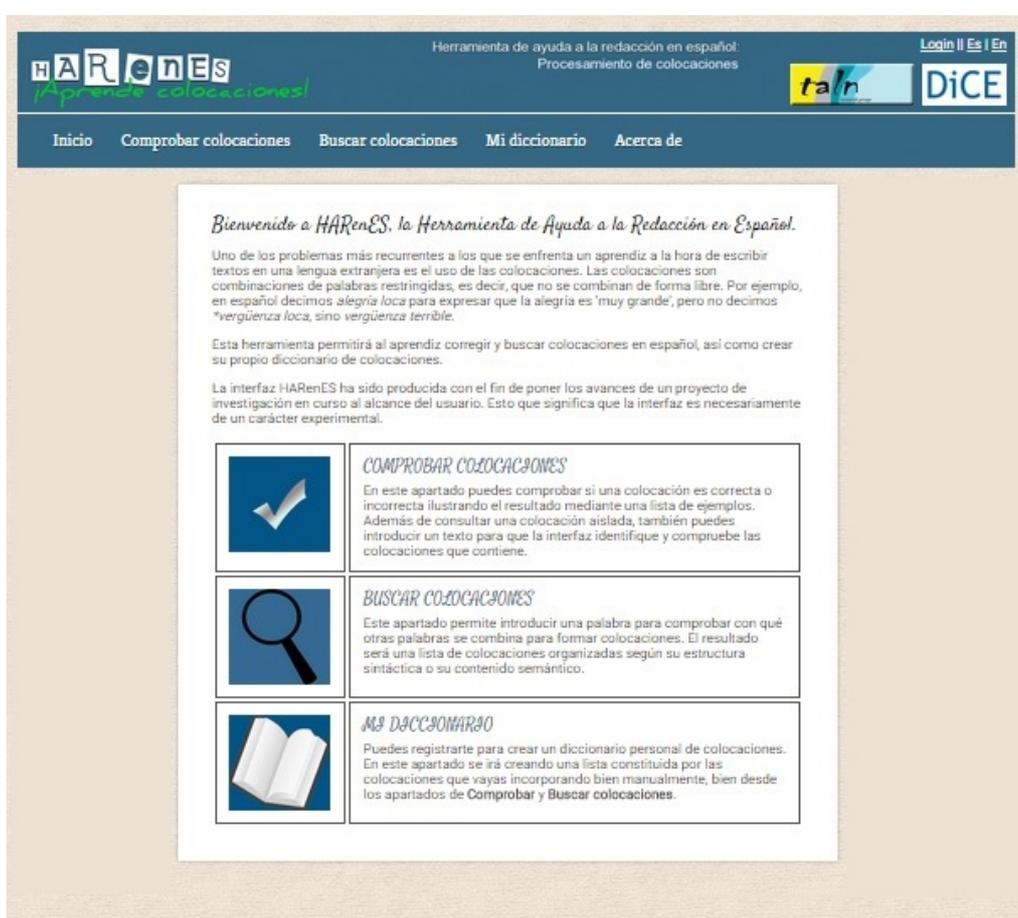


Fig. 4. Pantalla inicial de HARenES

[7] HARenES está siendo desarrollada conjuntamente por el equipo de la Universidad Pompeu Fabra, dirigido por L. Wanner y por mi equipo, en marco del proyecto de investigación del Ministerio FFI2011-30219-C02-01

El entorno consta de tres componentes principales: 1) el verificador de colocaciones, que propone sugerencias en caso de que la colocación consultada sea incorrecta; 2) la búsqueda de colocativos de una base dada, por medio principalmente de medidas estadísticas en el corpus; 3) el diccionario de colocaciones personalizado, en donde el usuario puede guardar las colocaciones que le interesen con sus ejemplos, así como el registro de los errores y sus soluciones.

Como vemos, en algo más de diez años hemos evolucionado de un diccionario de colocaciones en la web hacia una ayuda a la escritura centrada en las colocaciones que integre el DiCE pero que va más allá. Veremos lo que nos depara el futuro.

Agradecimientos

Este trabajo da muestra de las actividades realizadas dentro del proyecto de investigación FFI2011-30219-C02-01, financiado por el MINECO y por fondos FEDER. Agradezco la atenta lectura de Marcos García Salido, así como la ayuda en la edición de figuras a Orsolya Vincze y especialmente a Estela Mosqueira.

REFERENCIAS BIBLIOGRÁFICAS

- ALMELA, R., CANTO SANTOS, P., SÁNCHEZ A., SARMIENTO, R. & ALMELA, M. (2005): *Frecuencias del español: Diccionario y estudios léxicos y morfológicos*, Madrid, Universitas.
- ALONSO RAMOS, M. (2004a): “Elaboración del *Diccionario de colocaciones del español* y sus aplicaciones”, en BATTANER, P., DECESARIS, J. (eds.): *De Lexicographia. Actes del I Symposium Internacional de Lexicografia*. Barcelona, IULA y Edicions Petició, 149-162. Alonso Ramos, M. (2004b): *Las construcciones con verbo de apoyo*, Madrid, Visor Libros.
- ALONSO RAMOS, M. (2005): “Semantic Description of Collocations in a Lexical Database”, en KIEFER, F. et al. (eds.). *Papers in Computational Lexicography COMPLEX 2005*, Budapest, Linguistics Institute and Hungarian Academy of Sciences, 17-27.
- ALONSO RAMOS, M. (2006): “Towards a Dynamic Way to Learn Collocations in a Second Language”, en CORINO, E., C. MRELLO & C. ONESTI (eds.) *Proceedings of the Twelfth EURALEX International Congress*, Torino, Accademia della Crusca, Università di Torino, Edizioni dell’Orso Alessandria, 909-923.
- ALONSO RAMOS, M. (2008): “Papel de los diccionarios de colocaciones en la enseñanza de español como L2”, en BERNAL, E. & J. DECESARIS (eds.) *Proceedings of the XIII EURALEX International Congress*, Barcelona, IULA, Documenta Universitaria, 1215-1230.
- ALONSO RAMOS, M. (2009): “Hacia un nuevo recurso léxico: ¿fusión entre corpus y diccionario?”, en CANTOS GÓMEZ, P. & A. SÁNCHEZ PÉREZ (eds.). *A Survey of Corpus-based Research. Panorama de investigaciones basadas en corpus*, Murcia, AELINCO, 1191-1207.
- ALONSO RAMOS, M. (2010): “No importa si la llamas o no colocación, descríbela”, en Mellado, C. et al. (eds.), *Nuevas perspectivas de la fraseología del siglo XXI*, Berlin, Frank & Timme, 55-80.

- ALONSO RAMOS, M. (en evaluación): “Can collocations be deduced? A lexically-driven analysis from the perspective of language production”, en TORNER, S.; BERNAL, E. (eds.). *Collocations and other lexical combinations in Spanish. Theoretical and applied approaches*. Ohio, Ohio State University Press.
- BENSON, M., BENSON, E., ILSON, R. (1986): *The BBI Combinatory Dictionary of English. A Guide to Word Combinations*. Amsterdam/Philadelphia, John Benjamins.
- BOSQUE, I. (2011): “Deducing Collocations”, plenary talk at the *5th International Conference on the Meaning-Text Theory (MTT-2011)*, Barcelona, Universitat Pompeu Fabra, in L. Boguslavsky & Leo Wanner (eds.), *Proceedings of the 5th International Conference on Meaning-Text Theory*, Barcelona, 2011, p. vi-xxiii. Disponible en línea: <http://olst.ling.umontreal.ca/pdf/proceedingsMTT2011.pdf>
- BOSQUE, I. (dir.) (2006): *Diccionario combinatorio práctico del español contemporáneo*. Madrid, SM.
- BUENDÍA, M., FABER, P. (2014): “Collocation dictionaries: a comparative analysis”, *MonTi: Monografías de Traducción e Interpretación* 6, 203-235.
- Consejo de Europa (2001). *Marco común europeo de referencia para las lenguas: aprendizaje, enseñanza, evaluación*. Madrid, MEC y Anaya. [http://cvc.cervantes.es/ensenanza/biblioteca_ele/marco/].
- CROWTHER, J., DIGNEN, S., LEA, D. (eds.) (2002): *Oxford Collocations Dictionary for Students of English*. Oxford: Oxford University Press.
- GELUSO, J., YAMAGUCHI, A. (2014): “Discovering formulaic language through data-driven learning: Student attitudes and efficacy”, *ReCALL* 26(2), 225-242.
- INSTITUTO CERVANTES (2006): *Plan Curricular del Instituto Cervantes. Niveles de referencia para el español*. Madrid, Biblioteca nueva [http://cvc.cervantes.es/ensenanza/biblioteca_ele/plan_curricular/].
- JOHNS, T. (1991): “Should you be persuaded: Two examples of data-driven learning”, en JOHNS, T., KING, P. (eds.), *Classroom concordancing*. English Language Research Journal, 4:1-16.

- KILGARRIFF, A., RENAU, I. (2013): “esTenTen, a Vast Web Corpus of Peninsular and American Spanish”, *Procedia-Social and Behavioural Sciences*, 95(25), 12–19.
- MEL'ČUK, I. (1996): “Lexical Functions: A Tool for the Description of Lexical Relations in the Lexicon”, en. L. Wanner (ed.), *Lexical Functions in Lexicography and Natural Language Processing*, Amsterdam/Philadelphia: John Benjamins, 37-102.
- MEL'ČUK, I. (2012): “Phraseology in the Language, in the Dictionary, and in the Computer”, *Yearbook of Phraseology*. 3(1), 31–56.
- MEL'ČUK I., A. CLAS & A. POLGUÈRE (1995): *Introduction à la lexicologie explicative et combinatoire*, Louvain-la-Neuve, Duculot.
- MORENO, P., FERRARO, G., WANNER, L. (2013): “Can we determine the semantics of collocations without using semantics?” en Kosem, I., et al. (eds.) *Electronic lexicography in the 21st century: thinking outside the paper. Proceedings of the eLex 2013 conference*, Ljubljana/Tallinn, Trojina, Institute for Applied Slovene Studies/Eesti Keele Instituut, 106-121.
- SCHUBERT, K. (2014): *The design of prototypes of a dictionary of English collocations*, Projekt im Studiengang Internationales Informationsmanagement, Universidad de Hildesheim.
- VÁZQUEZ VEIGA, N. (2014), “Marcas de uso en el *Diccionario de colocaciones del español*”, *Zeitschrift für romanische Philologie*, 130(3), 698-724.
- VERLINDE, S., P. LEROYER & J. BINON (2009): “Search and you will Find. From Stand-alone Lexicographic Tools to User Driven Task and Problem-Oriented Multifunctional Leximats”, *International Journal of Lexicography*, 23(1):1–17.
- VERLINDE, S., PEETERS, G. (2012): “Data access revisited: the Interactive Language Toolbox”, en S. GRANGER, M. PAQUOT (eds). *Electronic lexicography*. Oxford, Oxford University Press, 147-162.
- VINCZE, O., ALONSO RAMOS, M. (2013a): “Incorporating frequency information in a collocation dictionary: Establishing a methodology”, *Procedia - Social and Behavioral Sciences*, 96, 241-248.

- VINCZE, O., ALONSO RAMOS, M. (2013b): “Testing an electronic collocation dictionary interface: Diccionario de Colocaciones del Español”, en Kosem, I., Kallas, J., Gantar, P., Krek, S., Langemets, M., Tuulik, M. (eds.) *Electronic lexicography in the 21st century: thinking outside the paper. Proceedings of the eLex 2013 conference*, Ljubljana/Tallinn, Trojina, Institute for Applied Slovene Studies/Eesti Keele Instituut, 328-337.
- VINCZE, O., MOSQUEIRA, E., ALONSO RAMOS, M. (2011): “An online collocation dictionary of Spanish”, en Boguslavsky, I. y L. Wanner, eds. *Proceedings of the 5th International Conference on Meaning-Text Theory*, Barcelona, Universitat Pompeu Fabra, 275-286. [disponible en <http://meaningtext.net/mtt2011/proceedings/>]