

# CÓRPORA Y ENSEÑANZA DE LENGUAS: SE BUSCAN

## COLOCACIONES<sup>1</sup>

Sabela Prieto González  
Estela Mosqueira Suárez  
Nancy Vázquez Veiga  
*Universidade da Coruña*

### **Abstract**

El objetivo de nuestro trabajo es explotar el corpus CEDEL2 para realizar un estudio comparativo del uso de colocaciones entre hablantes nativos del español y estudiantes de español como L2. Este corpus resulta idóneo para dicha tarea, pues se compone de un número similar de redacciones producidas por hablantes nativos y no nativos de español, lo que facilita la descripción comparativa del uso de las colocaciones en ambos grupos, pues nos permite medir su frecuencia en torno a un determinado tópico y hacer un análisis cualitativo de los resultados obtenidos.

Este trabajo nos servirá para analizar tanto los errores que con más frecuencia cometen los estudiantes de español L2 como las carencias detectadas en sus textos, con el propósito de elaborar propuestas didácticas específicas, potencien el uso adecuado de las colocaciones por parte de los aprendices de ELE.

**Palabras clave:** colocaciones, lingüística de corpus, ELE

## **I. INTRODUCCIÓN**

En este trabajo presentaremos los resultados de la primera fase del proyecto de investigación *Colocaciones y Tecnologías Lingüísticas (COLOCATE)*, cuyo objetivo final es construir un entorno multimedia para la enseñanza y aprendizaje de las colocaciones en español. Aunque inicialmente ha sido concebido para mejorar las destrezas de los estudiantes de español L2, consideramos que también puede constituir una herramienta útil en el aula de español L1.

Esta primera fase, que podemos llamar de diagnóstico, consiste en observar los problemas que el uso de las colocaciones plantea a los aprendices de español. Hemos tomado como punto de partida el *Corpus Escrito del Español L2 (CEDEL2)* (Lozano, *en prensa*), compuesto por textos producidos por hablantes nativos y no nativos de español, a partir de una temática previamente fijada. Esto facilita la descripción comparativa del uso de las colocaciones, pues nos permite medir la frecuencia de colocaciones en torno a un determinado tópico en ambos tipos de textos y hacer un análisis cualitativo de los resultados obtenidos.

---

<sup>1</sup> Parcialmente financiado por el Ministerio de Ciencia e Innovación (FFI2008-06479-C02-01) y el Programa de Becas FPU (MEC)

El trabajo que a continuación presentamos se estructura en tres partes principales. En primer lugar, describiremos el corpus CEDEL2, del cual hemos tomado los datos para nuestro análisis. En segundo lugar, nos referiremos a la importancia de la enseñanza de las colocaciones a los estudiantes de una lengua extranjera. Finalmente, mostraremos los resultados de la comparación de los textos producidos por hablantes nativos y no nativos, y analizaremos tanto los errores que con más frecuencia cometen los hablantes de español L2 como las carencias detectadas en sus textos. Así, partiendo de este trabajo previo, podremos elaborar propuestas didácticas específicas, que se centren en los errores más frecuentes y que potencien el uso adecuado de las colocaciones por parte de los estudiantes de ELE.

## II. EL CORPUS CEDEL2

El corpus CEDEL2<sup>2</sup> pertenece al proyecto de investigación *Word Order in second Language acquisition corpora* (WOSLAC), dirigido por Amaya Mendikoetxea, y realizado por la Universidad Autónoma de Madrid y la Universidad de Granada. Este corpus incluye, por un lado, textos redactados por aprendices de español como L2, cuya lengua materna es el inglés; y, por otro, redacciones escritas por nativos de español. De cada usuario se ofrecen datos personales que pueden resultar relevantes a la hora de diseñar un estudio comparativo, como su grado de dominio de la L2, su lengua materna, o su conocimiento de otras lenguas.

Tomando la información de la propia web del proyecto, sabemos que los datos incluidos en CEDEL son recogidos de forma *online*, mediante tres formularios electrónicos que el usuario debe rellenar. Uno de ellos es una redacción en la que se valora la espontaneidad, con una extensión de aproximadamente 500 palabras, y sobre un tema escogido entre las doce materias propuestas por los investigadores. Gracias a todas estas intervenciones voluntarias, actualmente el corpus consta de unas 400.000 palabras, aunque su objetivo es alcanzar la cifra de un millón.

Para nuestro estudio, hemos revisado una muestra de 252 textos tomados de CEDEL2, una mitad escrita por nativos de español y la otra, por estudiantes de ELE de un nivel medio-avanzado. Los textos redactados por extranjeros fueron organizados de mayor a menor grado de conocimiento de su L2, según el nivel que se le había asignado dentro del corpus CEDEL2. A partir de aquí realizamos un análisis de cada uno de los documentos, basado en la detección y señalización de las colocaciones empleadas. En el caso de los estudiantes de ELE, establecemos una clasificación preliminar entre colocaciones correctas, incorrectas o mejorables, en la que seguimos trabajando para elaborar una tipología de errores que nos servirá para un futuro etiquetado de corpus. A continuación, llevamos a cabo un análisis cuantitativo de las colocaciones encontradas, para hallar el promedio de aparición tanto en el caso de los nativos como en el de los no

---

<sup>2</sup> Proyecto de investigación en marcha, que se puede consultar en la página web <http://www.uam.es/proyectosinv/woslac/cedel2.htm>

nativos, y de este modo concluimos que los datos son lo suficientemente rentables como para llevar a cabo nuestro estudio.

### III. COLOCACIONES Y ENSEÑANZA DE LENGUAS

La importancia de la enseñanza de colocaciones en la adquisición de una segunda lengua es cada vez más reconocida (Alonso Ramos, 2006; Howarth, 1998; Lewis, 2000; Nesselhauf, 2003 y 2005) pero, para que su aprendizaje sea significativo, debemos descubrir los errores cometidos con mayor frecuencia por los estudiantes, de modo que podamos identificarlos, clasificarlos y crear un material didáctico específico orientado a su enseñanza.

Con *colocación* nos referimos a una combinación en la que una palabra exige la presencia de otra para expresar un sentido dado, como *vencer el dolor* o *matar el aburrimiento* ('causar que el ~ desaparezca'). Siguiendo la terminología de Hausmann (1979: 191) llamamos *base* al término de la colocación que es autónomo semánticamente y que selecciona, para un significado concreto, una UL específica que será denominada *colocativo*. Como ya adelantamos, el sustantivo *dolor* selecciona el verbo *vencer* para expresar el significado 'causar que el ~ desaparezca', mientras que el sustantivo *aburrimiento* selecciona el verbo *matar* para expresar el mismo significado.

El fenómeno colocacional aparece tanto en el lenguaje cotidiano (*café solo*, como 'café sin leche') como en ámbitos más específicos (*oficiar la misa*, *clonar un gen*); por esto, su enseñanza debe afrontarse desde un principio en el proceso de aprendizaje, ya que las colocaciones permiten designar actividades habituales como *dar un paseo* o *hacer una pregunta*. Su estudio es tan necesario como complejo pues, por ejemplo, muchos idiomas disponen de un vocablo que significa 'pregunta', pero en cada caso se combina con un verbo diferente: en español *hacemos una pregunta*, en inglés la "pedimos" (*to ask a question*) y en francés la "colocamos" (*poser une question*) (Sanromán, 2006: 145). En la actualidad, varios autores afirman que para conocer una lengua extranjera, el aprendiz debe dominar sus colocaciones (Lewis, 2000) pues, si las palabras se aprenden en su contexto, la recuperación de bloques léxicos será más sencilla al tiempo que mejorará la fluidez y la corrección del alumno (Higueras García, 2004: 15).

La gran presencia de colocaciones en las lenguas y la importancia que tienen en la adquisición del idioma han hecho que, en los últimos años, hayan florecido abundantes estudios sobre este tema para el caso español (Bosque, 2001; Corpas Pastor, 2001; Koike, 2001; Koike, 2002; García Platero, 2002).

Pero, para contar con datos reales que nos permitan realizar un análisis significativo, la lingüística de corpus resulta de gran ayuda, pues nos permite acceder a textos escritos por hablantes nativos y extranjeros de la lengua que pretendemos estudiar.

#### IV. ESTUDIO COMPARATIVO DEL EMPLEO DE COLOCACIONES EN HABLANTES NATIVOS Y NO NATIVOS

La lingüística de corpus ha abierto nuevas vías de investigación en el terreno de la enseñanza de lenguas, pues nos permite acceder a muchos textos producidos por aprendices de L2, para detectar los problemas que pueden tener y ponerles solución. En nuestro trabajo, el hecho de contar con dos corpus de similares características, uno de redacciones escritas por nativos de español y otro escrito por estudiantes de ELE, nos permite examinar el uso de las colocaciones en ambos casos y tomar a los nativos como referencia.

Nuestra hipótesis inicial suponía que los aprendices evitarían el empleo de colocaciones, debido a la dificultad que entraña su aprendizaje y el uso idóneo del colocativo apropiado en cada caso. Para corroborar nuestra teoría, tomamos 126 textos escritos por nativos (los que estaban recogidos en la muestra del CEDEL2 que nos fue facilitado para este estudio) y el mismo número de redacciones escritas por extranjeros, para proceder a señalar todas las colocaciones empleadas en ellos. Los resultados de esta búsqueda inicial resultaron sorprendentes, pues el número de colocaciones empleadas por estudiantes de español superaba al utilizado por hispanohablantes. Ante esta observación, decidimos realizar un análisis más exhaustivo de algunos textos en concreto, para llevar a cabo un estudio descriptivo de las colocaciones encontradas en cada grupo de textos y poder comparar los resultados obtenidos.

Para nuestro análisis seleccionamos las redacciones de autores que habían elegido el tema “Describe un viaje que has hecho recientemente”, puesto que en ambos corpus existían aproximadamente veinte textos con esa temática y, pudimos comprobar en nuestra aproximación inicial que esta materia destacaba por el elevado número de colocaciones empleadas.

En un primer recuento, comprobamos que los extranjeros utilizaron un total de 278 colocaciones, frente a las 190 de los nativos; pero este dato, que podría llamar nuestra atención inicialmente, no resulta tan llamativo si tenemos en cuenta que las redacciones de los extranjeros son más extensas que las de los nativos, con lo cual, el uso de colocaciones sería aproximadamente el mismo. De todos modos, la igualdad en el número de colocaciones existentes en ambos corpus tampoco era lo que esperábamos en un principio, pues creíamos que los extranjeros evitarían dichas estructuras y que sería necesario hacer hincapié en este aspecto durante las clases de ELE. Por tanto, la alta presencia de colocaciones en las redacciones escritas por aprendices nos lleva a realizar un análisis cualitativo de las mismas, que nos permita llevar a cabo una comparación del uso de colocaciones en nativos y no nativos, partiendo de una misma temática. Para ello, organizamos las colocaciones más recurrentes en torno a determinadas palabras que funcionan como base de la colocación.

En primer lugar, analizamos las colocaciones que emplean la palabra *vacaciones* como base de la colocación. Los nativos utilizan los colocativos *irse* o *pasar* (*irse de vacaciones*; *pasar las vacaciones*) como verbos que indican la realización de las

*vacaciones*; mientras que, en el caso de los extranjeros, encontramos la colocación correcta *ir de vacaciones* junto al uso erróneo de la expresión *estar en vacación*, resultante del calco de la expresión inglesa *to be on holiday*, en donde se traslada al español la estructura morfosintáctica del inglés, incluyendo el singular de la palabra *vacación*, y dando como resultado una forma inexistente en nuestra lengua.

En segundo lugar, nos centramos en la palabra *viaje*, que en los textos redactados por nativos aparece como base de las siguientes colocaciones: *realizar un viaje*, *hacer un viaje*, *disfrutar de un viaje*, *organizar un viaje* o *planear un viaje*. De las expresiones mencionadas, *hacer un viaje* es la más recurrente, probablemente por su carácter coloquial. Entre los textos de los no nativos sólo encontramos la colocación *hacer un viaje*, formada con una “palabra comodín”, lo que nos lleva a pensar que los aprendices no tienen una percepción clara de la colocación correcta y presentan importantes carencias en el registro de colocativos adecuados. Se trataría, por tanto, de un ejemplo ilustrativo de aquellas estructuras que es necesario potenciar en el ámbito didáctico.

A continuación, estudiamos las colocaciones que tienen como base la palabra *tiempo*, tanto en su acepción climatológica como en su acepción temporal, por los que nos estamos refiriendo a dos unidades léxicas diferentes.

En primer lugar, observamos que los textos escritos por nativos ofrecen gran variedad de colocaciones en las que aparece alguna expresión temporal, como en *pasar el tiempo*, *hace tiempo*, *tener tiempo*, *dar tiempo (a)*, *permanecer tiempo*, *perder el tiempo*. Las correspondientes en el caso de los extranjeros son *tardamos mucho tiempo*, *aprovechar el tiempo*, *tener tiempo (suficiente)*. Se trata, por tanto, de un tipo de expresiones que los participantes dominan y que dan lugar a una amplia variedad de registro.

En segundo lugar, para referirnos a las condiciones climatológicas, reunimos en un mismo grupo colocaciones que emplean como base las palabras *tiempo*, *calor* o *frío*, en combinación con el verbo *hacer*. Los nativos utilizan construcciones habituales como *hacer calor* o *hacer frío*, mientras que los extranjeros dominan este espectro de colocaciones, pues encontramos ejemplos como *hizo buen tiempo*, *tener un tiempo (fantástico)*, *hizo calor*, *hacía mucho calor*, *hacía mucho frío*. A pesar de tratarse de estructuras que nada tienen que ver con su correspondencia en inglés apenas cometen errores. Como excepción, recogemos una construcción incorrecta, *el tiempo hacía calor*, en la que se asigna al *tiempo* la función de sujeto para evitar la oración impersonal, inexistente en inglés.

Otro grupo nutrido de colocaciones lo conforman aquéllas cuya base hace referencia a las comidas del día. En las redacciones de nativos, encontramos tres colocaciones que estarían dentro de este bloque temático: *tomar el desayuno*, *preparar la comida* y *servir la comida*. No ocurre lo mismo en el caso de los extranjeros donde encontramos las siguientes colocaciones: *comiendo la cena*, *tomar el desayuno*, *comiendo comida*, *comida típica*, *probar comida*, *comida rica*, *comida auténtica*. Cuantitativamente, estos términos resultaron mucho más productivos entre los

extranjeros a la hora de emplear colocaciones. Mientras que un nativo español tiende a utilizar las formas sintéticas *desayunar*, *comer* o *cenar* (y sus variantes), los extranjeros crean colocaciones probablemente porque en la estructura inglesa equivalente siempre se emplea un verbo de apoyo, en este caso *to have*. Es importante evitar la tendencia a la hipercorrección cuando analizamos las colocaciones de los no nativos. Nuestra intuición nos puede jugar malas pasadas a la hora de evaluar las estructuras empleadas por un estudiante de ELE. A priori, desde nuestro punto de vista, si tenemos que escoger entre la opción sintética *desayunar* y la analítica *tomar el desayuno*, nos quedaríamos con la primera y consideraríamos que la recurrencia en las redacciones de extranjeros de la colocación *tomar el desayuno* se debe, en parte, al calco que hacen de su correspondiente en inglés. Sin embargo, y en contra de este argumento, vemos que también aparece la expresión *tomo el desayuno* en una de las redacciones escritas por nativos, con lo que confirmamos que ambas estructuras son igualmente válidas.

A continuación, recogemos colocaciones que toman como base sustantivos que designan medios de transporte. Por una parte, en los textos de nativos encontramos expresiones como *tomar el ómnibus*, *tomo una motocar*, *tomar un bote*, *cogí un taxi*, *coge el tren*, *coger el avión*. Se puede observar que alternan los verbos de apoyo *tomar* y *coger*, pues en el corpus intervienen igualmente españoles y latinoamericanos. Por otra parte, los textos de aprendices presentan las siguientes expresiones: *cogíamos un taxi*, *fuimos en taxi*, *habíamos cogido un tren*, *fuimos en tren*, *tomamos el tren*, *coger el avión*. En algún caso, el mismo hablante emplea las dos variantes: *tomar el tren* y *coger el tren*. Al igual que ocurre entre el inglés británico y el americano, hay una gran variedad de construcciones que varían entre el español hablado en España y el de los países de Sudamérica. Esta información es vital y deberá tenerse en cuenta a la hora de elaborar material didáctico, con el fin de que el estudiante de español como L2 sea consciente de las connotaciones a la hora de utilizar una u otra expresión.

En último lugar, mostramos las colocaciones que se refieren a expresiones temporales. En las redacciones de los nativos tenemos *el año pasado*, *el mes pasado*, *mañana siguiente*. Una vez más, las colocaciones en los textos de aprendices son más numerosas, aunque en algunos casos cometen fallos a la hora de escoger el colocativo adecuado. Entre las expresiones correctas encontramos *año pasado*, *el mes de marzo pasado*, *el fin de semana pasado*, *el semestre pasado*, *el año pasado*, *el verano pasado*, *los próximos tres días*. Sin embargo, también detectamos colocaciones incorrectas como *los cuatro meses pasados* ('los últimos cuatro meses') o *el próximo día* ('al día siguiente'). Del análisis realizado en torno a este grupo de colocaciones se concluye que hay un uso extendido de las expresiones *el próximo día* y *la mañana próxima* en lugar de *al día siguiente* y *a la mañana siguiente*, respectivamente. Se trata, por tanto, de una construcción errónea en la que habría que incidir para alcanzar resultados satisfactorios en el aprendizaje del español como L2.

## V. REFLEXIÓN FINAL

En España, el estudio de las colocaciones ha llegado a la lingüística aplicada con un importante retraso, que se traduce en una frecuente invisibilidad de este tipo de estructuras en el aula de lengua extranjera. En la actualidad, el número de investigadores que reclaman una mayor atención para el ámbito de la fraseología va en aumento. Entre los trabajos que se centran en el tema de la didáctica destacan autoras como Ruiz Gurillo (2000), Higuera (1997) y Tortajada (2000), quienes critican la falta de obras específicas que aborden la enseñanza de las unidades fraseológicas y que presenten nuevas metodologías dirigidas al docente a la hora de abordar la enseñanza de las colocaciones.

También se ha llamado la atención sobre la necesidad de primar la comprensión y producción de las colocaciones por parte del estudiante, en lugar de dar preferencia a los aspectos formales. Por ejemplo, Fernández Prieto (2005) propone profundizar en la investigación de las colocaciones, expresiones muy abandonadas hasta el momento y que dan lugar a multitud de errores cuando se aprende una L2. Nosotros tomamos ese testigo y pretendemos contribuir a la labor del docente, ofreciéndole un buen referente para elaborar material didáctico adecuado a las necesidades reales del estudiante.

Este análisis comparativo constituye una pequeña muestra que nos permite evidenciar el peso que tienen las colocaciones en cualquier lengua, pues son expresiones de elevada frecuencia en el lenguaje cotidiano. Por consiguiente, el correcto aprendizaje de las mismas ha de considerarse como un objetivo primordial en el proceso de adquisición de una segunda lengua. Como hemos mostrado en este trabajo, una detenida observación de los datos nos permite disponer de información no sólo de los errores colocacionales de los extranjeros, sino también de su destreza a la hora de producir colocaciones y de las carencias detectadas en su uso. En esta fase exploratoria, la utilización del corpus CEDEL2 ha resultado de suma utilidad, pues nos ha proporcionado un valioso primer diagnóstico sobre el que basaremos nuestras investigaciones posteriores.

## Referencias bibliográficas

- Alonso Ramos, M. (2006). Towards a dynamic way of learning collocations in a second language». In *Proceedings of the Twelfth EURALEX International Congress*, 909-923.
- Bosque, I. (2001). Sobre el concepto de *colocación* y sus límites. *Lingüística Española Actual*, 23/1, 9-40.
- Corpus Escrito Del Español L2 (CEDEL2) <<http://www.uam.es/woslac/cedel2.htm>> [Fecha de la consulta: 09-02-2009]
- Corpas Pastor, G. (2001). Apuntes para el estudio de la colocación. *Lingüística Española Actual*, 23/1, 41-56.
- Fernández Prieto, M. J. (2005). La enseñanza de la fraseología. Evaluación de recursos y propuestas didácticas. In *Las gramáticas y los diccionarios en la enseñanza del español como segunda lengua, deseo y realidad Actas del XV Congreso Internacional de ASELE*, 349-356
- García Platero, J. M. (2002). Aspectos semánticos de las colocaciones. *Lingüística Española Actual*, 24/1, 25-34.
- Hausmann, F. J. (1979). Un dictionnaire des collocations est-il possible?. *Travaux de littérature et de linguistique de l'Université de Strasbourg*, 17, 1, 187-195.
- Higueras García, M. (1997). La importancia del componente idiomático en la enseñanza del léxico a extranjeros. *Frecuencia L*. Madrid: Edinumen, pp. 15-19.
- Higueras García, M. (2004). *La enseñanza aprendizaje de las colocaciones en el desarrollo de la competencia léxica en el español como lengua extranjera*, Tesis doctoral, Facultad de Educación, Universidad Complutense de Madrid.
- Howarth, P. (1998). The phraseology of learners' academic writing'. In A. P. Cowie (Ed.), *Phraseology. Theory, analysis, and applications*. Oxford: Clarendon Press, pp.161-186.
- Koike, K. (2001). *Colocaciones léxicas en el español actual. Estudio formal y léxico-semántico*, Alcalá de Henares: Universidad de Alcalá/Takushoku University.
- Koike, K. (2002). Comportamientos semánticos en las colocaciones léxicas. *Lingüística Española Actual*, 24/1, 5-24.
- Lewis, M. (ed.) (2000): *Teaching Collocation. Further Developments in the Lexical Approach*, LTP, London.
- Lozano, C. (en prensa). CEDEL2: Corpus Escrito del Español L2. In *Actas del XXVI Congreso de AESLA*. Universidad de Almería.
- Nesselhauf, N. (2003). The use of collocations by advanced learners of English and some implications for teaching. *Applied Linguistics*, 24/2, 223-242.
- Nesselhauf, N. (2005). *Collocations in a learner corpus*. Amsterdam and Philadelphia: John Benjamins.
- Ruiz Gurillo, L. (2000). Un enfoque didáctico de la fraseología española para extranjeros. *Espéculo*. Madrid: Universidad Complutense de Madrid. <http://www.ucm.es/info/especulo/ele/fraseolo.html>
- Sanromán Vilas, B. (2006). Observaciones sobre el uso de colocaciones en aprendices finlandeses de español: hacia una aplicación didáctica. In M. Alonso Ramos (Ed.), *Diccionarios y fraseología, Anexos de Revista de Lexicografía*, 3, Universidade da Coruña, 145-160.
- Tortajada Millán, M.A. (2000). Las unidades fraseológicas: una propuesta didáctica mediante situaciones comunicativas ilustradas en imágenes. *Frecuencia L*, 15. Madrid: Edinumen, pp. 27-39.