LOS CÓRPORA ELECTRÓNICOS EN LA ENSEÑANZA DEL VOCABULARIO ESPAÑOL LE: LA UTILIDAD PEDAGÓGICA DE LAS CONCORDANCIAS

Mª Carmen Izquierdo Gil
Universidad de Avignon

1. INTRODUCCIÓN

La necesidad de controlar, limitar y graduar el aprendizaje del vocabulario de las lenguas extranjeras motivó en el primer tercio del siglo XX la constitución de corpus lingüísticos[1]. El recuento estadístico de las unidades léxicas aparecidas en dichos corpus daría lugar a los diccionarios o listados de frecuencias léxicas, conocidos también como vocabularios básicos. Mediante dichos recuentos se pretendía dar cuenta del vocabulario más empleado por los hablantes de una lengua y observar la repartición de las unidades léxicas desde el punto de vista de la frecuencia en el léxico general. Al margen de ciertas limitaciones pedagógicas observadas en los listados de frecuencias léxico-estadísticos, diversos investigadores han considerado que "la frecuencia" constituye uno de los criterios útiles de selección del vocabulario en la enseñanza de lenguas extranjeras (Benítez y Zebrowski, 1993: 229; Carter, 1996: 140, 181; Nation, 1990: 16-20, 2001: 9; McCarthy, 1990: 66-67; Sinclair y Renouf, 1996: 148; Carter y McCarthy, 1996: 210). Fundamentalmente en la planificación léxica del nivel elemental debemos dedicar un tiempo considerable a la adquisición de unidades léxicas de alto índice de uso[2] (Gómez Molina, 2001; Nation, 2000: 16; 1990: 209; Carter, 1996: 190). Picoche ha apuntado justamente en diversas ocasiones la necesidad de establecer una enseñanza sistemática del vocabulario basada en el criterio de frecuencia[3].

---

[1] Amplias colecciones de bases de datos extraídos de textos auténticos de cualquier tipo (orales, escritos; literarios, periodísticos, científicos, coloquiales, etc.).

[2] El índice de uso nos ofrece un perfil más preciso sobre la frecuencia de una unidad léxica en un corpus dado, ya que pondera la dispersión (o reparto) de una unidad léxica en los tipos discursivos y la frecuencia de aparición de dicho elemento.

[3] A partir de numerosos vocabularios básicos se ha constatado que las palabras vacías, gramaticales o funcionales (determinantes -artículos, demostrativos, posesivos, numerales-, pronombres personales, preposiciones, conjunciones, verbos auxiliares, ciertos adverbios) ocupan los primeros puestos de las listas. Al margen de estos elementos funcionales hallamos asimismo la presencia de conceptos nominales de significado muy general (*cosa, casa, tiempo, lugar, persona, sitio),* una serie de sustantivos referidos al tiempo *(año, día, mes, momento, semana, noche, hora, tarde, mañana, rato, minuto, mediodía),* y a entidades humanas *(niño, amigo, hijo, mujer, persona, hombre, gente).* Encontramos igualmente una serie de adjetivos con gran extensión semántica (*bueno, malo, pequeño, grande, nuevo, bonito, viejo, largo, difícil, importante, interesante, fuerte, fácil, moderno, antiguo, diferente...*), así como una serie de verbos polisémicos (diversos significados) y polivalentes (diversas construcciones) tales como *ser, haber, tener, estar, ir, hacer, decir, ver, saber, venir, dar, poder* entre otros.

ASELE. Actas XII (2001). M.ª CARMEN IZQUIERDO GIL. Los córpora electrónicos en l...
Centro Virtual Cervantes

> Un enseignement systématique du vocabulaire réside dans l'étude des mots très peu spécifiques, très peu connotés, à la fois hyperonymiques et très polysémiques[4] qui apparaissent en tête des diverses listes de fréquences et que tout le monde croit connaître. (Picoche, 1993: 57)

> Un enseignement systématique du lexique doit logiquement commencer par assurer un maniement correct et aisé des mots les plus fréquents, et remonter, par le jeu des synonymies, des antonymies et des dérivations vers des mots moins fréquents. (Picoche, 1992: 53)

2. LOS CÓRPORA ELECTRÓNICOS

2.1. Las concordancias.

Uno de los principales inconvenientes observados en los listados de frecuencias léxico-estadísticos consiste en presentar las unidades léxicas de forma descontextualizada. Señalaremos que un diccionario de frecuencias carece de microestructura (definiciones, ejemplos...) a diferencia de un diccionario de uso. Es decir, los vocabularios básicos se presentan como listas de palabras acompañadas de índices. Frente a estas constataciones, algunos investigadores han avanzado la necesidad de analizar los patrones léxicos y sintácticos más recurrentes ofrecidos por los córpora electrónicos como un complemento indiscutible de los diccionarios de frecuencias (Mallikamas, 1999 ; Schmitt, 2000 : 69; López Morales, 1996; Conrad, 1999). Así pues, respecto a los listados de frecuencias clásicos impresos, los actuales córpora electrónicos, consultables vía Internet, nos proporcionan las voces contextualizadas en las pantallas de concordancias. Entendemos por concordancia el programa informático que presenta un conjunto o lista de las ocurrencias de una palabra dada en un corpus. En dichas pantallas, la palabra clave aparece en el centro de cada línea con sus propios entornos textuales o el contexto lingüístico inmediato, es decir, las palabras que aparecen antes y después de la forma lingüística objeto de estudio (Schmitt y McCarthy, 1997: 328; Sinclair, 1991: 32; Cantos 1995: 55; Sánchez Pérez, 2000: 15-29; Brodine, 1998: 172).

Concordancias (RAE)

Consulta: *sé*, en todos los medios, en *CREA*, en *Oral*, en *ESPAÑA*

Resultados: 4619 casos en 547 documentos.
Filtro sobre documentos. Ratio: 1/5. 965 Casos en 110 Documentos.
Pantalla 8 de 39.

Nº CONCORDANCIA (Pantalla 8)

176 a seguir con los libros, por si te apetece y no **sé.** Si hay algo más, ya te llamaré. Hasta ma
177 hacer unas cosas a la Complutense. Así es que, no **sé,** si necesitas mandarlo mañana,

[4] Entendemos por hiperonímicos aquellos términos más abarcadores (*persona, cosa, lugar, sitio, ciudad*...) y por polisémicos, aquellas unidades que poseen dos o más significados (*hacer, tener, dar, tomar, coger).*

ASELE. Actas XII (2001). M.ª CARMEN IZQUIERDO GIL. Los córpora electrónicos en l...
Centro Virtual Cervantes

178 ez de vuelta, en eurotra, ¿vale? Y si no, pues no **sé**, llámame allí. ¿De acuerdo? Hasta luego.
179 tu casa ahora cuando he llegado, viernes, hoy, no **sé**, a las ocho y cuarto o por ahí y no estabas
180 Vamos a ir al cine. ¿A qué cine? Pues no **sé**, no sabemos qué cine, a ver si nos espabila
181 tío, y nos decidimos. ¿Tú qué haces? Pues no lo **sé**. Que no sé, depende. Pues cuando hayam
182 cidimos. ¿Tú qué haces? Pues no lo sé. Que no **sé**, depende. Pues cuando hayamos decidido
183 pero tú me dio ese teléfono. Es incorrecto, no **sé**. Creo que se ha cambiado el número de pri
184 Sé perfectamente que ya te has ido a París o no **sé** qué. Te has ido al extranjero para mucho t
185 rico, soy David. Que ya he vuelto de mi viaje. No **sé** si estás en Madrid o si no estás en Madri
186 se ha cortado. Que cuando puedas me llames. No **sé** si vendrás a comer. Si vienes a comer no
187 dábamos para lo de las fotocopias. Que, bueno, no **sé**, el jueves si vas a ir o yo voy a ir. De tod
188 por la hora y todo. Llámame cuando puedas. Y no **sé** le has comentado a Ritchie y a toda esta
189 í, ese que no se le entiende. Es que es que no **sé** quién tendréis. ¿Cuál tenéis ahora de Sist
190 tío así, calvo, con gafas. ¿Se llama Ruz? No sé no **sé**. No, o Tirado Tirado. Hay uno que se lla
191 sí, calvo, con gafas. ¿Se llama Ruz? No sé no **sé**. No, o Tirado Tirado. Hay uno que se llam
192 ioneta. Exactamente. ¡Hombre, pero eso no no no **sé** si lo veo, inusitado. A ver, señor Casamaj
193 asando por Burgos, o pasando por Zaragoza". , ya **sé** cómo. Como en Eloísa está debajo de un
194 e obra de don Enrique Jardiel Poncela, pero yo no **sé** si realmente vais por aquí lo que estamos
195 de películas famosas, como por ejemplo, pues no **sé**, Los Hermanos Marx en el Oeste o La Di
196 z. O sea desinencia. ¿La raíz? Pero si yo no **sé** cómo se pone. Claro, la raíz. La primera f
197 las que se derivan de la rama céltica?" yo qué **sé**, "¿en cuántas épocas se puede dividir el g
198 al, se considera que el germánic sí, yo sí que lo **sé**. Entonces se puede considerar que el germ
199 nte tuyo. No, a mí no te no te pienses que me lo **sé** muy bien, pero yo con las que me voy a po
200 tes se lo saben todo, que te alucinan. Pero yo no **sé**, o sea, sacar chuleta siempre es más arries

Gracias a la contextualización de las unidades léxicas, las concordancias nos ofrecen informaciones valiosas. Así, por ejemplo, las concordancias de la forma verbal *sé* en una muestra del corpus CREA[5] (tema Oral) indican una asociación recurrente de dicho verbo con el adverbio de negación "no sé"[6]. Un análisis de la expresión coloquial *y pico* en 144 casos (86 documentos) aparecidos en el tema Oral del mismo corpus CREA nos revela su asociación con números para designar:

- precios (*doce millones y pico de pesetas)*
- longitud (*dos metros y pico)*
- hora (*a las diez y pico, un minuto y pic)* o tiempo (*un mes y pico)*
- edad (*setenta y pico años)*
- cantidad (*seis mil y pico millones de habitantes)*

Las pantallas de concordancias nos facilitan, por lo tanto, la identificación de las combinaciones, patrones o esquemas léxicos y sintácticos más recurrentes, más típicos y,

[5] El corpus electrónico CREA (Corpus de Referencia del Español Actual) elaborado por la R.A.E. analiza 180 millones de palabras recopiladas en documentos entre 1975-2000. Acerca de la distribución temporal y temática de los textos recopilados, consúltese CREA en http://www.rae.es

[6] Se observa en una consulta al corpus CREA (todos los medios, tema Oral, España) que la combinación *no sé* aparece en un 75,33% de ocurrencias de la muestra (a partir de un filtrado de 1/5 de 4619 casos en 547 documentos igual a 965 casos, 110 Documentos). Otras expresiones encontradas son: *yo qué sé, no lo sé, no sé a lo mejor...*

ASELE. Actas XII (2001). M.ª CARMEN IZQUIERDO GIL. Los córpora electrónicos en l...
Centro Virtual Cervantes

en consecuencia, estadísticamente más significativos de las palabras-clave objeto de nuestra búsqueda. Puesto que dichos programas de concordancias pueden ofrecernos cientos o miles de ejemplos de un término determinado con sus propios contextos en cuestión de segundos, puede resultar útil seleccionar cierto número de ejemplos mediante la operación de filtrado[7] u ordenar las ocurrencias alfabéticamente por la primera, segunda o tercera palabra a la izquierda o a la derecha de la palabra clave con el fin de visualizar las combinaciones más destacables. Por ejemplo, al ordenar la forma verbal *sé* por la primera palabra a su derecha constatamos, en dos pantallas escogidas al azar, la presencia de esta forma verbal en interrogaciones indirectas (*no sé cuándo, cuántos, por qué...).*

2.2. Las colocaciones léxicas.

Las concordancias suelen revelar la presencia de colocaciones, es decir, combinaciones frecuentes de dos ó más palabras o unidades que tienen una tendencia muy fuerte a ser utilizadas juntas. Así pues, algunas consultas en el corpus CREA sobre unidades léxicas con un elevado índice de uso en los léxicos básicos de Avila (1999), Morales (1986) y Juilland (1965) tales como *cada, menos, mundo, ejemplo, lado, mejor, momento* nos muestran colocaciones recurrentes en producciones lingüísticas orales. Dicho corpus electrónico puede evidenciar las combinaciones más destacables no sólo a partir de las concordancias, sino también a partir de la operación de "agrupación" en 2 ó 3 palabras. Consideramos que dichas combinaciones léxicas o sintácticas deberían tenerse muy en cuenta en la elaboración de un programa léxico elemental.

- ***cada***. Sinclair y Renouf (1996: 154) habían observado algunos esquemas colocacionales en palabras gramaticales muy comunes. Así pues, los autores indican que *each* ("cada") aparece de forma significativa en la lengua junto a unidades de tiempo *hour, day, week* ("hora, día, semana"). Las concordancias sobre el determinante *cada* en el corpus CREA (todos los medios, tema Oral)[8] nos brindan un tipo de información muy similar:

cada vez 20%
cada día 10%
cada año 2%
cada momento 2%
cada semana 2%

[7] La misión del filtro consiste en reducir el número de ejemplos cuando la abundancia de los mismos obstaculiza la labor de consulta. De ese modo, un filtro de tipo 1/2 seleccionará los documentos 1, 3, 5, 7, 9... del conjunto inicial. Si un documento contiene 10 ejemplos de la consulta realizada, la aplicación de este filtro en la proporción 1/2 reduce a 5 los casos seleccionando alternativamente un caso de cada dos.

[8] Los porcentajes ofrecidos en los siguientes casos están extraídos mediante la selección de todos los medios, tema Oral, en España del corpus CREA.

ASELE. Actas XII (2001). M.ª CARMEN IZQUIERDO GIL. Los córpora electrónicos en l...
Centro Virtual Cervantes

- ***menos.*** El filtrado de ocurrencias en todos los medios, tema Oral indica que la forma léxica *menos* aparece en un 40% de ocasiones en dos locuciones: un 20% de ocurrencias se deben a la locución *por lo menos* y casi otro 20% a la locución *más o menos*. Esta última locución muestra, sin embargo, porcentajes muy superiores a la precedente en otras diversas búsquedas.
- ***mundo***. Esta forma léxica aparece en un 37% de casos en la expresión *todo el mundo*.
- ***ejemplo.*** Esta forma léxica aparece en un 92% de ocurrencias en la locución *por ejemplo*.
- ***lado.*** Esta forma léxica aparece en un 31% de casos en la locución *al lado de/del*, en un 19% de ejemplos en la locución *por un lado* y en un 24% de ocurrencias en la locución *por otro lado*.
- ***mejor***. Al margen de su uso como adjetivo comparativo, observamos que esta forma léxica alcanza una alta frecuencia en la locución *a lo mejor* (46% de ocurrencias).
- ***momento.*** Casi un 30% de las ocurrencias de la forma léxica *momento* se deben a su presencia en dos locuciones: 11% *de momento* y 17% *en este momento*.

Consulta: *ejemplo*, en todos los medios, en CREA, en ORAL, en ESPAÑA

2535 Casos en 478 Doc. Resultado (Filtrado 1/3): 901 casos en 160 Documentos

Agrupaciones

De 2 palabras	%	Casos
por ejemplo	77,24	696
Por ejemplo	14,53	131
un ejemplo	2,88	26
el ejemplo	1,66	15
Como ejemplo	0,77	7
El ejemplo	0,44	4
ese ejemplo	0,33	3
propio ejemplo	0,11	1
darían ejemplo	0,11	1
Otros	1,88	17

Consulta: *mejor*, en todos los medios, en CREA, en ORAL, en ESPAÑA

3005 Casos en 540 Doc. Resultado (Filtrado 1/4): 793 casos en 135 Documentos

De 3 palabras	%	Casos
a lo mejor	40,35	320
A lo mejor	6,05	48
es el mejor	1,13	9
es lo mejor	0,75	6
es la mejor	0,50	4
en el mejor	0,50	4
que es mejor	0,50	4
siempre es mejor	0,37	3
que lo mejor	0,37	3
Otros	49,43	392

ASELE. Actas XII (2001). M.ª CARMEN IZQUIERDO GIL. Los córpora electrónicos en l...
Centro Virtual Cervantes

Algunos listados de frecuencias del español tales como el de Juilland y Chang-Rodríguez (1965) y el de Morales (1986) revelan una incorrecta definición del concepto de unidad léxica. En ellos verificamos la presencia únicamente de unidades monoverbales. Esta constatación nos indica que los autores partieron de una definición ortográfica y no semántica de la palabra. Procediendo de tal manera, estos listados nos ocultan la presencia de combinaciones fraseológicas en la lengua e incluso nos impiden localizar las más frecuentes de éstas[9]. Acabamos de ver, sin embargo, que las concordancias y agrupaciones en el corpus electrónico CREA ponen en evidencia la importancia de ciertas locuciones en la lengua (*más o menos, a lo mejor, por ejemplo...*). Sólo el reciente recuento de frecuencias léxicas de Avila (1999) ha superado la identificación tradicional de la unidad léxica al incluir, a diferencia de los anteriores listados, unidades fraseológicas (locuciones, fórmulas rutinarias...).

Los modernos trabajos de léxico-estadística muestran, por consiguiente, que ciertos enunciados fraseológicos tienen un peso considerable en el vocabulario de una lengua. Como apunta McCarthy, los más frecuentes de dichos enunciados deben hallarse presentes en un vocabulario básico.

> The raw frequency list hides the frequency of phrasal combinations, and extra research is needed to ensure that the most frequent phrasal items are not lost from the basic vocabulary. (McCarthy, 1999: 247)

También McCarthy y otros autores han verificado, al observar las concordancias, que algunas formas lingüísticas alcanzan una alta frecuencia debido a su presencia en la lengua principalmente bajo la forma de unidades fraseológicas. Así pues, McCarthy (1999: 237) indica que la palabra 14 en el corpus CANCODE (*know)* y la 37 (*think)* alcanzan dicha frecuencia esencialmente por combinación sistemática con los pronombres *you* y *I* respectivamente, en expresiones formulaicas del tipo *you know* y *I think*. Sinclair y Renouf (1996: 148) evocan asimismo el ejemplo de la palabra inglesa *see*. Al margen de su significado prototípico como verbo transitivo ("ver" algo o a alguien), se ha determinado mediante estudios de léxico-estadística que el primer y segundo usos de *see* se encuentra en las expresiones familiares *you see* y *I see*. En un análisis de 9 manuales de inglés como lengua extranjera se observó que ambas expresiones sólo representaban un 10 % de todas las ocurrencias de *see*, mientras que en el Birmingham Corpus alcanzaban un 53 % de realizaciones.

> Hence usages which are discoursal or pragmatic which carry out functions to do with the larger patterns in texts, are often missed. (Sinclair y Renouf, 1996: 152)

[9] Algunos investigadores tales como Schmitt (2000 : 82) o Sinclair y Renouf (1996 : 147) han puesto asimismo en tela de juicio la ausencia de unidades fraseológicas muy comunes (*never mind, sort of, of course, all right*) en ciertos listados de frecuencias clásicos anglófonos.

ASELE. Actas XII (2001). M.ª CARMEN IZQUIERDO GIL. Los córpora electrónicos en l...
Centro Virtual Cervantes

Esto no significaría, sin embargo, que el significado más común en el corpus hablado *you see* ("¿sabes ?", "¿comprendes ?") tenga que desplazar a un segundo plano al significado prototípico de *see* ("ver"). Las concordancias tan sólo revelan la existencia de unidades fraseológicas de elevada frecuencia que no podemos pasar por alto en el aprendizaje de una lengua extranjera. Actualmente se insiste en que un conocimiento de dichas unidades es esencial, crucial, en la enseñanza del léxico L2 para desarrollar la competencia comunicativa, especialmente la habilidad léxicodiscursiva y socio-comunicativa (McCarthy, 1990: 67; Carter, 1996: 176 ; Sinclair, 1991; Willis, 1999).

2.3. Las colocaciones gramaticales.

Bogaards (1994 : 122) advirtió que una palabra puede alcanzar una alta frecuencia gracias a otras con las que se asocia a menudo en la lengua y que facilitan su aparición. Así pues, en el *Français Fondamental* el verbo *aller* alcanzaba un alto índice de frecuencia gracias a su combinación con verbos en infinitivo, con nombres mediante preposición (*aller à, aller chez, aller dans)* y construcciones sin preposición (*aller rue..., aller ailleurs, y aller).* También algunas consultas en el corpus CREA sobre verbos muy frecuentes en los vocabularios básicos del español nos revelan colocaciones sintácticas importantes que deben reflejarse en la planificación léxico-gramatical de un nivel elemental. En primer lugar, la consulta de agrupaciones de concordancias de la forma verbal *tiene* nos permite constatar mediante filtrado que la expresión de obligación *tiene que* alcanza un 18 % de ocurrencias, lo cual coincide con las consultas de Sánchez Pérez (1995: 77) en el corpus *Cumbre.* El resto de éstas hacen referencia al significado prototípico de este verbo ("tener = poseer") y a construcciones deslexicalizadas con *tener* como verbo soporte. Se observará asimismo que ciertas formas verbales tales como *creo, parece* obtienen un elevado índice de uso en diversos listados de frecuencias (Ávila, 1999: 189; 277) debido a su recurrencia en patrones léxico-sintácticos del tipo *creo que, (me) parece que*, lo cual se puede explicar por su frecuente función atenuadora en la vida cotidiana. En las agrupaciones del CREA, las expresiones *creo que*[10] y *parece que* alcanzan respectivamente un 88% y un 46% de ocurrencias en un filtrado.

La forma verbal *acaba* alcanza un 62% de ocurrencias en la colocación sintáctica *acaba de + infinitivo*. La forma verbal *deja* obtiene un 33% de ocurrencias en la colocación *deja de + infinitivo*. La forma verbal *volver* presenta un 55% de ocurrencias en la colocación *volver a + infinitivo*. La forma verbal *sé* (verbo saber) aparece en un 75% de ocurrencias con el adverbio de negación "no sé". La forma verbal *siento* se encuentra en un 45% de ocurrencias en la fórmula rutinaria "lo siento". La consulta de la forma verbal *voy* nos revela el predominio del esquema colocacional *voy a* en un 77% de ocurrencias. La

[10] En el corpus CREA (todos los medios y temas) la forma verbal *creo* aparece en 15.522 casos (3.781 documentos), de los cuales 12.027 casos (en 3.273 documentos) pertenecen a la expresión *creo que*.

ASELE. Actas XII (2001). M.ª CARMEN IZQUIERDO GIL. Los córpora electrónicos en l...
Centro Virtual Cervantes

mayoría de los ejemplos presentan la combinación *voy a + infinitivo idem (ven ir a + infinitivo).*

2.4. Las construcciones con verbo-soporte.

Algunos investigadores (Fox, 1992: 46; McCarthy, 1999: 239; O'Dell, 1997: 271) han observado que los verbos fuertemente polisémicos con un elevado índice de uso y frecuencia (*tener, dar, hacer, tomar...),* actúan frecuentemente en los córpora electrónicos como verbos soporte o de apoyo. Dichas unidades se denominan asimismo verbos deslexicalizados a causa de su bajo contenido semántico. La carga semántica de dichas construcciones queda asumida por el componente nominal. La función principal de estos verbos consiste, por lo tanto, en servir de soporte sintáctico y morfológico (marcas de tiempo, modo y persona) a un sustantivo que expresa un predicado semántico.

Señalaremos que los denominados verbos soporte conocen un empleo estándar en el cual mantienen su significado léxico lleno prototípico. Así pues, el significado central o básico del verbo *dar* contiene la idea de transmisión (¿Me *das* un caramelo?), el significado básico de *tener* implica la idea de posesión (*Tengo* dos libros de aventuras), el significado básico del verbo *hacer* implica la idea de elaboración o producción (*hacer* un bocadillo), etc. Sin embargo, a menudo cuando se enseñan dichos verbos polisémicos en el aula, se insiste más en el significado prototípico (como verbos llenos de contenido semántico) que en los usos deslexicalizados de estos verbos. Los verbos soporte representan, por consiguiente, un fenómeno insuficientemente explorado en los manuales de enseñanza de lenguas extranjeras (Sinclair y Renouf, 1996: 148; Fox, 1992) a pesar de ser comunes en el lenguaje informal oral (Carter y McCarthy, 1997: 13). Los mencionados autores y muchos otros (O'Dell, 1997; McCarthy, 1999) han subrayado la necesidad de prestar más atención a tales usos en la enseñanza del vocabulario. Nos parece, por lo tanto, fundamental que los conceptores de manuales y los profesores mismos indaguen acerca de los términos más frecuentes y útiles que suelen combinarse con dichos verbos[11].

Con el fin de delimitar algunas construcciones con verbo soporte comunes en español hemos revisado[12] las concordancias de cuatro verbos (*tener, dar, hacer, tomar)* en el corpus CREA:

[11] No debemos olvidar que numerosos ejemplos presentan divergencias en la elección del verbo soporte entre diversas lenguas extranjeras, así por ejemplo, entre el español y el francés: *tener cuidado/faire attention; hacer una pregunta/poser une question; dar una vuelta/faire un tour; dar un beso/faire la bise.*

[12] También hemos verificado las concordancias de dichos verbos en los artículos de los siguientes diccionarios bilingües: *Grand Dictionnaire Larousse Español-Francés/Français-Espagnol* (1998), de García Pelayo y Testas, Paris, Larousse-Bordas. *Harrap's Hispano. Dictionnaire Espagnol-Français/Français-Espagnol* (1997) de Jean-Paul Vidal, Paris, Larousse-Bordas. *Diccionario Actif para estudiantes de francés. Francés-Español/Español-Francés* (1999) de Nicholas Rollin (dir.), Madrid, Santillana. *Diccionario Didáctico de Francés* (1999), de Concepción Maldonado, Madrid, S.M.
TENER: *suerte, cuidado, tiempo, miedo, (un) sueño, cuidado, prisa, paciencia, ganas de..., hambre, frío, sed..., dolor de, ... años, fiebre, razón.*

ASELE. Actas XII (2001). M.ª CARMEN IZQUIERDO GIL. Los córpora electrónicos en l...
Centro Virtual Cervantes

- *TIENE una idea, miedo, sentido, interés, prisa, suerte, tiempo, importancia, ganas, razón, ... años, la intención, la impresión, la sensación, derecho, la costumbre de..., la culpa...*
  [Tema Oral, 6877 casos en 666 Doc, filtrado 1/7, 915 casos en 96 Doc]

- *DAR/DA un paseo, una vuelta, lástima, pena, tiempo, suerte, un beso, un consejo, un golpe, clase de, una –buena/mala– noticia, una opinión, las gracias, permiso, importancia, una sorpresa, asco, vergüenza, la gana, miedo...*
  [DAR, Tema ORAL (1306 casos en 436 doc, filtrado ½, 679 casos en 218 doc), Tema OCIO (2043 casos en 1151 doc, filtrado 1/3, 669 casos en 384 doc) ; DA, tema ORAL (1619 en 459 doc, filtrado ½, 829 casos en 230 doc)]

- *HACE calor, daño, falta, fresco, frío, tiempo, sol, –un– rato, ruido,... años –días, meses, un momento–, fotos, regalos, cas...*
  [Tema ORAL, 8277 en 1130 doc, filtrado 1/8, 927 casos en 143 doc]

- *TOMAR agua, algo, chocolate, copas, el aperitivo, el té, vino, una ensalada, un plato (caliente/frío), la medicina, leche, un vaso de..., el almuerzo, un café, un refresco, un zumo, un cerveza, una aspirina, una sopa, una taza de.., las vacaciones, una foto, una decisión, el aire, el avión, el ascensor, el sol, el tren, un taxi, un camino, una carretera, la calle...*
  [Tema ORAL (369 casos en 215 doc) y en el tema OCIO (869 casos en 476 doc)]

Tras los datos expuestos queda claro que la enseñanza del vocabulario de una lengua extranjera debe considerar tanto las unidades léxicas monoverbales con mayor índice de uso como los patrones de coocurrencia sintácticos o léxicos más significativos localizados en los programas de concordancias. Al fin y al cabo, las unidades léxicas más relevantes en los listados de frecuencias clásicos constituyen únicamente la base o el punto de partida para el desarrollo de la competencia comunicativa, "no more than *basis* for a usable competence. Learners will need a large number of other words if they are to communicate effectively in the language." (Willis, 1999: 119)

DAR: *una vuelta, un paseo, las gracias, vergüenza, un beso, un golpe, un abrazo, un salto, miedo, clases, una (buena/mala) noticia, lástima, pena, un consejo, la hora, una sorpresa.*
HACER: *una pregunta, daño, deporte, una foto, un favor, frío, calor, sol, aire, buen/mal tiempo, una película, ruido, comida, la cama.*
HACER cola, deporte (gimnasia), un viaje (una excursión), ruido, daño, la cama, una pregunta, caso, mal a alguien, falta, una foto, una visita, una llamada, un buen día, frío, calor, un regalo, una película (Construcciones aparecidas en el corpus de Calderón Campos, 1994: 112-124).
TOMAR: *el desayuno, un café, postre, una medicina, una aspirina, una decisión, el sol, el autobús, el tren, un taxi, un camino, un baño, una foto, las vacaciones.*

ASELE. Actas XII (2001). M.ª CARMEN IZQUIERDO GIL. Los córpora electrónicos en l...
Centro Virtual Cervantes

Nos unimos, así pues, a McCarthy (1999: 238) cuando apunta que los datos brutos que nos ofrecen los diccionarios de frecuencias o vocabularios básicos deben imperativamente ser afinados mediante un trabajo analítico a partir de las concordancias y agrupaciones ofrecidas en los corpus electrónicos. En este sentido, las concordancias de los córpora constituyen un complemento indiscutible de los listados de frecuencias léxico-estadísticos.

3. REFERENCIAS BIBLIOGRÁFICAS

Ávila Muñoz, Antonio Manuel (1999): *Léxico de frecuencia del español hablado en la ciudad de Málaga*, Málaga, Universidad de Málaga.

Benítez Pérez, Pedro y Zebrowski, Jerzy. (1993): "El léxico español en los manuales polacos", en Salvador Montesa Peydró y Antonio Garrido Moraga (Eds.): *Actas III Congreso Nacional de ASELE, El español como lengua extranjera: De la teoría al aula,* Málaga, 223-230.

Bogaards, Paul (1994): *Le vocabulaire dans l'apprentissage des langues étrangères*, Paris, Hatier-Didier, Collection LAL.

Brodine, Ruey (1998): "Concordanze e consapevolezza lessicale. L'uso dei *corpora* per l'esplorazione del léxico", en María Teresa Prat Zagrebelsky (Ed.): *Lessico e apprendimento linguistico. Nuove tendenze della ricerca e pratiche didattiche,* Firenze, la Nuova Italia, Quaderni del Lend 1, 171-191.

Calderón Campos, Miguel (1994): *Sobre la elaboración de diccionarios monolingües de producción,* Granada, Universidad de Granada.

Cantos, Pacual (1995): "Tratamiento informático y obtención de resultados", en Aquilino Sánchez: *Corpus lingüístico del español contemporáneo,* Madrid, SGEL, 39-70.

Carter, Ronald. (1996): *Vocabulary. Applied Linguistic Perspectives*, London, Routledge.

Carter, Ronald y McCarthy, Michel (1996): *Vocabulary and Language Teaching*, New York, Longman.

Conrad, S.M. (1999): "The importance of corpus-based research for language teachers", *System* 27/1, 1-18.

Fox, Gwyneth (1992): "Using corpora for lexicographic purposes", *Revista Canaria de Estudios Ingleses* 25, 45-53.

Gómez Molina, José Ramón (2001): "La competencia léxica en la enseñanza-aprendizaje de español como L2 y LE", *Mosaico* 5.

Juilland, Alphonse y Chang-Rogríguez, E. (1964): *Frequency Dictionary of Spanish Words*, The Hague, Mouton & Co.

López Morales, Humberto (1996): "*Corpora* orales hispánicos", en Antonio Briz, José Ramón Gómez Molina y María José Martínez (Eds.): *Pragmática y gramática del español hablado,* Universidad de Valencia, 137-145.

Mallikamas, Prima (1999): "Applications of Corpora in Language Teaching", ThaiTESOL Bulletin, 12/1. (http://www.thaitesol.org/bulletin/1201/120101.html)

McCarthy, Michael (1990): *Vocabulary,* Oxford, Oxford University Press.

ASELE. Actas XII (2001). M.ª CARMEN IZQUIERDO GIL. Los córpora electrónicos en l...

Centro Virtual Cervantes

McCarthy, Michael (1999): "What constitutes a basic vocabulary", *Studies in English Language and Linguistics* 1, 233-249.

Morales, Amparo (1986): *Léxico básico del español de Puerto Rico*, San Juan de Puerto Rico, Academia Puertorriqueña de la lengua española.

Nation, I.S.P. (1990): *Teaching and Learning Vocabulary*, Boston, Heinle Publishers.

Nation, I.S.P. (2001): *Learning Vocabulary in Another Language,* Cambridge, Cambrige University Press.

O'Dell, Felicity (1997): "Incorporating vocabulary into the syllabus", en Norbert Schmitt and Michael McCarthy (Eds.): *Vocabulary : Description, Acquisition and Pedagogy*, Cambridge, Cambridge University Press, 258-278.

Picoche, Jacqueline (1992): *Précis de lexicologie française. L'étude et l'enseignement du vocabulaire,* Paris, Nathan Université, [1977].

Picohe, Jacqueline (1993): *Didactique du vocabulaire français*, Paris, Nathan Université.

Sánchez Pérez, Aquilino (1995): *Corpus lingüístico del español contemporáneo. Fundamentos, metodología y aplicaciones*, Madrid, SGEL.

Sánchez Pérez, Aquilino (2000): "Language Teaching before and after 'Digitalized corpora'. Three main sigues", *Cuadernos de Filología Inglesa*, 9/1, 5-37.

Schmitt, Norbert y McCarthy, Michael (Eds.) (1997): *Description, Acquisition and Pedagogy*, Cambridge, Cambridge University Press.

Schmitt, Norbert (2000): *Vocabulary in Language Teaching,* Cambridge, Cambridge University Press.

Sinclair, John (1991): *Corpus, Concordance, Collocation,* Oxford, Oxford University Press.

Sinclair John y Renouf Antoinette (1996): "A lexical syllabus for language learning", en Ronald Carter y Michael McCarthy (Eds.): *Vocabulary and Language Teaching*, London, Longman, 140-160.

Willis, Dave (1999): "Syllabus Design and the Pedagogic Corpus", en *Vocabulary learning in a foreign language,* Paris, Bristish Council, Goethe-Institut y ENS Fontenay/St Cloud, 115-148.

ASELE. Actas XII (2001). M.ª CARMEN IZQUIERDO GIL. Los córpora electrónicos en l...

Centro Virtual Cervantes

Centro Virtual Cervantes