

LA FRECUENCIA LÉXICA Y SU UTILIDAD EN LA ENSEÑANZA DEL ESPAÑOL COMO LENGUA EXTRANJERA

Manuel Alvar Ezquerro
Universidad Complutense de Madrid

Me veo en la necesidad de aclarar, como ya he hecho en otras ocasiones, que no me siento especialista en la enseñanza del español como lengua extranjera, y siempre que he abordado estas cuestiones ha sido desde la perspectiva general del estudioso de la lengua española, y más concretamente de su léxico y fuentes lexicográficas, por lo que mi intervención de hoy va a estar marcada por esas limitaciones.

En el instante los organizadores de este congreso me invitaron a participar en él ya sabían muy bien todo eso y me pidieron que hablara sobre alguna cuestión concerniente a la enseñanza del léxico, pues tenían muy presente la reciente publicación de mi librito sobre *La enseñanza del léxico y el uso del diccionario*¹. En él incluí un capítulo sobre “La riqueza del léxico y la frecuencia de palabras” que hubiera deseado más amplio, pero para el que no había más espacio, por las características de la colección en que apareció la obra. Inmediatamente pensé en él, y lo que sigue no es más que consecuencia de lo que quedó sin desarrollar entonces, de la voluntad de quienes me invitaron, y de las deudas de generosa amistad y colaboración.

Cuando nos disponemos a enseñar una lengua, la materna u otra, especialmente a la hora de enseñar el léxico, se nos plantean diversas preguntas, que requieren contestaciones específicas: qué enseñar, cuándo hacerlo y cómo hacerlo. En las páginas que siguen quiero intentar si no dar contestación cabal a todas esas preguntas, por lo menos mirar los caminos que se pueden recorrer para averiguar qué léxico debemos enseñar. Otra cosa son el cuándo y el cómo, que dependen de la edad y conocimientos adquiridos por los alumnos, y del método seguido para hacerlo, aunque, necesariamente, habrá que hacer alguna referencia a ello.

No son muchos los medios de que disponemos para saber cuáles son las palabras que tienen que conocer nuestros alumnos, pues el diccionario, entendido como diccionario de lengua, tal y como lo concebimos habitualmente, pese a la riqueza de su contenido, no nos proporciona informaciones sobre la frecuencia del léxico, ya que su finalidad no es ésa, por un lado, y, por otro, al tratarse de una descripción de la lengua, todas las palabras y acepciones constan en él por igual, o al menos de una manera similar: cada palabra aparece una sola vez, pues en la lengua las unidades léxicas son únicas. El uso es asunto muy diferente; pertenece al plano del hablar, como todos bien sabemos.

El conocimiento de la frecuencia del léxico puede valernos enormemente para saber cuáles son las palabras que deben aprender nuestros alumnos, y para organizar su enseñanza. Pienso ahora en las palabras desde el punto de vista formal, pues las redes y relaciones de carácter semántico deben establecerse de otro modo. Es más, el léxico es sólo una parte de los conocimientos lingüísticos que debe adquirir el alumno, por lo que la enseñanza de determinadas formas no dependerá de su frecuencia de uso, sino de otros aspectos lingüísticos. Es lo que nos sucede, por ejemplo, con los derivados regulares no lexicalizados, cuyo conocimiento adquirirá el estudiante mientras progresa en el aprendizaje de las estructuras gramaticales, no de las léxicas. De este apartado debemos separar en nuestras tareas docentes las formas que se han lexicalizado, con la pérdida del recuerdo de su motivación lingüística (como sucede con *cenicero*, *lenguado*, *mantilla*, *pañuelo*, etc., y con algunos compuestos como *mariposa* y otros), por lo que deberán ser enseñadas cuando les corresponda en el progreso de la adquisición del léxico. Esto es, la información sobre la frecuencia de uso no puede ser un valor rígido al que debemos sujetarnos, pues hay otros factores que influyen en el aprendizaje de las palabras, como los señalados, o las mismas relaciones formales y designativas que puedan tener. Las palabras no son hechos aislados

¹ Arco-Libros, Madrid, 2003.

como nos las presentan los diccionarios, sino que mantienen relaciones de muy diversa naturaleza con otras palabras.

La pregunta es la de cómo saber cuál es la frecuencia de las palabras. En la actualidad disponemos de dos vías para averiguarlo:

- a) las indicaciones que nos proporcionan los corpus.
- b) los diccionarios de frecuencias.

Cada uno de estos caminos ofrece sus ventajas y sus inconvenientes. Los corpus están ligados a las nuevas tecnologías y es de esperar que en un futuro inmediato nos ofrezcan unas posibilidades de consulta que se adecuen a nuestras necesidades. Por el contrario, los diccionarios de frecuencias son unos instrumentos que no se han renovado y presentan el inconveniente de su antigüedad, a pesar de que los ordenadores ofrecen grandes posibilidades para su renovación. No son dos vías de consulta que estén separadas, sino que pueden confluir, por más que parezcan divergentes debido al desarrollo que ha tenido cada una de ellas.

Un corpus está formado por una gran cantidad de datos lingüísticos —léxicos para nuestros intereses— que se pretenden representativos de la lengua en alguna o varias de sus manifestaciones (lengua hablada, lengua escrita), niveles (culto, literario, coloquial, etc.), medios por los que se transmite (novela, teatro, prensa diaria, radio, etc.), especialidades o ámbitos (economía, política, deporte, historia, aficiones y coleccionismo, etc.), etc. Los datos contenidos en los corpus pueden consultarse mediante unas herramientas en las que se especifican los criterios de búsqueda. No son, pues, índices léxicos en sentido estricto, sino que contienen palabras, millones de palabras, a las que podemos acceder en nuestras consultas. Esa gran cantidad de palabras que registran les confiere una objetividad y representatividad de la realidad lingüística no logradas hasta nuestros días.

Por ahora, no son muchos los corpus que existen para el español, aunque cada vez hay más, de contenidos más ricos, y de mayores posibilidades de acceso, aunque algunos de ellos no son de dominio público.

En general, los corpus presentan las palabras como concordancias, esto es, cada voz con un contexto más o menos largo, pero sin estadísticas globales, lo que ya es un escollo para nuestros intereses. Ello no quiere decir que no existan corpus en los que se ofrezcan frecuencias, o que no se hayan desarrollado programas específicos para obtenerlas. Sin embargo, no es éste el mayor inconveniente que se puede señalar, ya que lo que nos presentan son las formas que se encuentran en los textos, sin el menor análisis ni agrupación. Pese a lo cual, como veremos, no carecen de utilidad, y esperemos que no pase mucho tiempo antes de que en nuestras consultas nos presenten las formas agrupadas de acuerdo con nuestros intereses, con lo que serían un instrumento verdaderamente útil para la preparación de los materiales didácticos. Otras dificultades con que podemos tropezarnos en la consulta de los corpus es la limitación en el número de accesos, o en el de elementos presentados. No pretendo aquí exponer una enumeración de las dificultades, que por sí solas bastarían para hacernos desistir de nuestro empeño, sino que quiero ver cómo pueden ayudarnos y por dónde deben mejorar.

En cuanto a los diccionarios de frecuencias, son viejos, por lo que nos presentan un caudal de palabras que no responde al uso actual de la lengua. Incluso alguna obra de análisis estadístico del léxico se ha confeccionado a partir de diccionarios, por lo que no puede ayudarnos para saber la frecuencia de las palabras, por más que sea de utilidad en otros aspectos del estudio del léxico; pueden servirnos para conocer algunas estructuras formales de las unidades léxicas, pero no para saber cuál es su frecuencia de uso. Pienso, por ejemplo, en el excelente trabajo —para otros fines— de Patterson y Urrutibéheity², por el que ha pasado ya un buen número de años, y que no resulta de utilidad para su aplicación en la enseñanza del léxico español, pues considera a las palabras como clases de distintos tipos: genealógicas, cronológicas, funcionales, físicas y estadísticas.

¿PARA QUÉ SIRVEN LAS LISTAS DE FRECUENCIAS LÉXICAS?

Desde que se elaboraron los primeros recuentos de léxico, se vio la utilidad práctica, pedagógica, que tenían, cuando no se confeccionaron con esa finalidad, como las diversas listas elaboradas por Thorndike³ para el inglés hace ya muchos años, o las que se hicieron bajo la dirección de Rodríguez Bou para el español de Puerto Rico⁴. Thorndike expuso claramente los objetivos

² William Patterson y Hector Urrutibéheity, *The lexical structure of Spanish*, La Haya-París, Mouton, 1975.

³ La primera de ellas fue Edward L. Thorndike, *The Teacher's Word Book*, 1921, revisado en *A Teacher's Word Book of Twenty Thousand Words Found Most Frequently and Widely in General Readings for Children and Young People*, Columbia University, Nueva York, 1931; y por último, Edward L. Thorndike e Irving Lorge, *The Teacher's Word Book of 30,000 Words*, Columbia University, Nueva York, 1944.

⁴ *Recuento de vocabulario de preescolares*, Universidad de Puerto Rico, Río Piedras, 1967.

que perseguía con sus listas, y que fueron tomados más adelante por Rodríguez Bou⁵. Aunque en esas exposiciones se pensaba en la enseñanza de la lengua materna, son igualmente aplicables a la enseñanza de segundas lenguas, tanto es así que cabe adaptar con facilidad aquellos objetivos para la enseñanza del español como lengua extranjera, pues los índices de frecuencias pueden servir para:

- 1º Conocer la importancia de las palabras de acuerdo con su frecuencia y distribución.
- 2º Decidir qué palabras deben enseñarse o no en cada nivel, en cuáles hay que insistir más y cuáles deben enseñarse para que los alumnos entiendan los materiales que manejan.
- 3º Poseer una información objetiva sobre las palabras y cuándo deben enseñarse.
- 4º Extraer el inventario de palabras del ámbito lingüístico en que se mueven los alumnos para facilitarles la comprensión de su entorno.
- 5º Saber cuáles son las palabras de ámbitos concretos cuando se enseña la lengua con fines específicos.
- 6º Seleccionar los grupos de palabras y formas que deben enseñarse conjuntamente, y en qué orden.
- 7º Determinar el grado de dificultad del léxico de los materiales de lectura y aprendizaje.
- 8º Que quienes confeccionan los materiales pedagógicos puedan seleccionar el léxico que debe aparecer en ellos.

Al hablar de estas cuestiones, Rodríguez Bou exponía que Michael West, en *The New Method Dictionary*, dio la pauta al definir cerca de 18 000 palabras y 6000 frases idiomáticas con unas 1500 palabras de las más frecuentes del inglés, por lo que, señalaba, «en español hace falta un diccionario de esta naturaleza para fines escolares. Basta con seleccionar cualquier palabra de un diccionario corriente de español y examinar su definición para darse cuenta de ello. A veces se definen palabras en términos de otras más difíciles aún que las que se trata de definir»⁶. Lo cierto es que aunque han pasado más de 50 años desde que escribió esas líneas, y de que en los últimos años se han realizado grandes avances en la lexicografía pedagógica del español, todavía no disponemos de un diccionario escolar de español lengua materna realizado con unos criterios parecidos a esos. El único intento, que yo conozca, en ese sentido, aunque dirigido a estudiantes de español lengua extranjera –sin olvidar a los hablantes nativos–, es el *DIPELE*⁷, con sus 2000 definidores, que aparecen relacionados al final de la obra, seleccionados no solamente por criterios estrictamente estadísticos, sino también sociolingüísticos.

Como se entiende con facilidad, las listas de frecuencias no son la solución de los problemas de la enseñanza del léxico, aunque su utilidad es evidente. En palabras del propio Rodríguez Bou, «una lista de frecuencia de palabras no es por sí un ábrete sésamo para la enseñanza de vocabulario. Creemos, sin embargo, que sí es una excelente ayuda»⁸.

Entre las aplicaciones didácticas de las listas de frecuencias de las palabras, se ha señalado la enseñanza de la lengua, pero solamente como un subproducto de estas listas⁹, ya que la ayuda principal se dirige hacia la estructuración del léxico a través de sus propiedades estadísticas.

Para enseñar la lengua, especialmente la nuestra, a estudiantes extranjeros, resulta de una gran utilidad saber cuál es su vocabulario, y cuáles son las voces más frecuentes, pues de ningún modo puede pretenderse que una persona que aprende una segunda lengua, cuando la domine, conozca las cerca de cien mil voces que hay en los diccionarios generales –que representan el vocabulario de la lengua–, pues ni siquiera los hablantes nativos nos aproximamos a esa cantidad, ni tampoco puede estar entre los objetivos de la enseñanza que se alcancen las 30 000 palabras que constituyen la competencia de un hablante culto. Las voces que se utilizan habitualmente son muchas menos, y el léxico medio de una comunidad, el que emplean todos sus miembros, puede establecerse entre 3000 y 5000 términos, si bien no disponemos de datos fiables a este respecto. Por ejemplo, para el francés, se cifra esa cantidad en unas 7000 u 8000 palabras¹⁰. Hacia esas cifras es a las que debe apuntar la enseñanza de segundas lenguas para que el estudiante sea capaz de entender un texto de un nivel lingüístico sin grandes pre-

⁵ Véase la "Introducción" de Ismael Rodríguez Bou, *Recuento de vocabulario español*, I, Universidad de Puerto Rico, Río Piedras, 1952, y en especial las págs. 9-10.

⁶ En la "Introducción" de su *Recuento*, pág. 10.

⁷ *Diccionario para la enseñanza de la lengua española*, Bibliograf-Universidad de Alcalá, Barcelona, 1995.

⁸ De nuevo en la "Introducción" de su *Recuento*, pág. 12.

⁹ Alphonse Juilland y Ernesto Chang Rodríguez, *Frequency Dictionary of Spanish Words*, Mouton, Londres-La Haya-Paris, 1964, pág. XXII.

¹⁰ Cfr. Jacqueline Picoche, *Précis de lexicologie française. L'étude et l'enseignement du vocabulaire*. Nueva edición revisada y actualizada, Nathan, Paris, 1994, pág. 48.

tensiones. Para ello hay que saber cuáles son esas palabras, pues sin saber cuáles son no podremos enseñarlas.

Independientemente de la competencia léxica que puedan tener en nuestra lengua los alumnos extranjeros a los que debemos enseñar, es conveniente saber cuáles son las voces más frecuentes, y para ello no disponemos sino de unos repertorios de carácter desigual, algunos de ellos ciertamente viejos, realizados con unos fines alejados de los que podemos tener en la enseñanza de español como lengua extranjera.

Pero la utilidad no solamente radica en saber cuáles son las palabras más frecuentes, sino también otra no muy desarrollada en el español, y de un enorme interés. Si los estudiantes que comienzan a aprender una segunda lengua no poseen un conocimiento amplio del léxico, y si además deben emplear un diccionario específicamente dirigido hacia ellos, éste no podrá contener en las definiciones palabras que no conozcan, pues si no, no se enterarán de lo que se quiere decir. Este tipo de diccionarios, en el que pensaba Rodríguez Bou, resulta imprescindible en cuanto hayamos avanzado un poco en la enseñanza de la lengua, y ello es así porque, normalmente, los grupos de alumnos son heterogéneos debido al número de lenguas que hablan, y el profesor no puede traducir a todas las lenguas las palabras, ni es conveniente que los alumnos manejen continuamente diccionarios bilingües, pues les limitan su capacidad de comprensión del español, y pueden proporcionarales un exceso de información, lo cual no contribuye a que aumenten sus conocimientos.

Llegados a este punto, debemos diferenciar lo que es la riqueza del vocabulario de una lengua, y el conocimiento que los hablantes tienen de las palabras, de lo que es el empleo que realizamos de esas palabras, pues las que utilizamos de forma habitual no son muchas, alrededor de mil¹¹. Pero hay más. No solamente son las de mayor de uso, sino que también aparecen distribuidas por todos los tipos de textos y situaciones de comunicación, ya que con ellas damos cuenta de las relaciones lingüísticas y de las nociones generales.

Disponemos de algunos datos del español, pero no parece haber demasiada concordancia entre todos ellos, por lo que habremos de consultarlos con precaución. Así, según las informaciones de Patterson y Urrutibéheity nuestras 100 palabras más frecuentes constituyen más del 30% del material léxico de cualquier texto; con las 1000 se alcanza el 50% de todo texto, y con las 5000 más abundantes se sobrepasa el 90%¹². Esas cifras son suficientes para comenzar a reflexionar, si bien no coinciden con las que nos ofrecen otros autores ni con los que yo mismo he podido recabar. Difieren sustancialmente de lo expuesto por Juilland y Chang Rodríguez, en cuyo diccionario de frecuencias sólo las diez primeras son ya el 36.55% de la frecuencia total, las cien primeras el 66.20%, y las mil primeras el 85.50% del total. Estos números hablan por sí solos, pese a sus diferencias, y nos muestran bien a las claras qué cantidad de palabras habremos de enseñar según los fines perseguidos, pues entre esas mil, o aquellas 5000, y las cien mil de los diccionarios generales de la lengua queda un enorme espacio que apenas conocemos (por eso, entre otros motivos, son necesarios los diccionarios).

Las palabras más frecuentes son, principalmente, las que sirven para mantener la estructura del mensaje, mientras que las voces realmente portadoras de información se repiten relativamente poco en el discurso.

En otro lugar¹³ ya he dado cuenta de mis recuentos realizados a partir del corpus Vox-Bibliograf, cuyo contenido es de 10 352 337 ocurrencias. Las formas más frecuentes son:

¹¹ Cfr. Aurélien Sauvageot, *Portrait du vocabulaire français*, Larousse, Paris, 1964, pág. 33.

¹² *Op. cit.*, pág. 9.

¹³ En *La enseñanza del léxico y el uso del diccionario*, citado, págs. 100-101.

de	691 341	6.68%
la	445 107	4.30%
que	319 892	3.09%
el	314 572	3.04%
en	301265	2.91%
y	284 619	2.75%
a	209 775	2.03%
los	186 827	1.80%
se	140 253	1.35%
del	129 081	1.25%
las	127 647	1.23%
un	111 381	1.08%
por	108 133	1.04%
no	96 402	0.93%
una	92 790	0.90%
con	92 729	0.90%
es	81 250	0.78%
su	72 013	0.70%
para	64 062	0.62%
al	60 006	0.58%
como	54647	0.53%
más	48691	0.47%
o	47992	0.46%
ha	31213	0.30%
sus	29235	0.28%
pero	28648	0.28%
si	22968	0.22%
este	21788	0.21%
son	21117	0.20%
entre	21106	0.20%
sobre	20564	0.20%
le	20563	0.20%
ya	20232	0.20%
sín	19102	0.18%
esta	17854	0.17%
muy	17844	0.17%
dos	17403	0.17%
todo	17190	0.17%
también	17098	0.17%
ser	16901	0.16%
Total (40 formas)	4 441 301	42.9%

Como vemos, las veinte más abundantes suman el 37.96%—cantidad consonante con los datos ofrecidos por Juilland y Chang Rodríguez, y no con los de Patterson y Urrutibéheity—de todas las apariciones, descendiendo a partir ahí de forma apreciable el número de ocurrencias. De acuerdo con los datos del corpus que manejo, con las cien formas más abundantes se llega al 49.06%, con las primeras mil alcanzamos el 67.54%, y con las cinco mil más frecuentes el 82.76%, cifras estas menores que las proporcionadas por los otros autores, seguramente porque en mi recuento he partido de formas, no de palabras lematizadas.

Las discrepancias que se observan en los distintos recuentos de léxico se deben a que están realizados de manera muy dispar. Bien es cierto que las palabras más frecuentes son siempre las mismas, pues no puede ser de otro modo, ya que constituyen el armazón de la lengua. Sin embargo, en no todos los recuentos aparecen en el mismo orden, ya que no siempre se deslindan los homógrafos o las funciones: unos proporcionan informaciones sólo sobre las formas, y otros sobre las palabras discriminadas por su categoría, función, significación, o agrupadas en paradigmas.

Pero ésa no es la única lamentación que se nos escapa al ver esas listas de palabras. Hay otra mayor que nos surge cuando queremos manejarlas para nuestros fines: la fecha de su realización. El léxico es algo cambiante, entre otros motivos porque refleja la realidad extralingüística, también cambiante. Hay palabras que no figuran en algunos de los recuentos, o que constan con una frecuencia que no es la que nos esperaríamos hoy, ya que en su momento no existía lo designado, o apenas se tenía conocimiento de ello. Así, no encontraremos en el recuento de Rodríguez Bou (1952) palabras como *televisión*, *video* u *ordenador*, pues aún no había llegado el momento de su generalización social. Además, entre sus materiales acoge los anteriores de Buchanan¹⁴, junto a los propios, lo que hace que su contenido sea ciertamente heterogéneo (Buchanan partió de 1.200.000 palabras, y Rodríguez Bou sobrepasa los 7.000.000, mientras que García Hoz¹⁵ se queda en las 400.000, material suficiente, según él, para determinar el vocabulario usual).

Por otra parte, la manera de presentar los materiales tampoco es igual, por lo que las comparaciones pueden resultar difíciles. Véase, por ejemplo, cómo el mismo Rodríguez Bou nos ofrece las 199 palabras más frecuentes:

LAS DIEZ MIL UNIDADES LÉXICAS MÁS FRECUENTES EN ORDEN DE RANGO

LAS PRIMERAS 87 UNIDADES LÉXICAS MÁS FRECUENTES DE LAS PRIMERAS 500 (111)

<i>Unidad léxica</i>	<i>Unidad léxica</i>	<i>Unidad léxica</i>
a	ése	no
agua	eso	nos
ahora	estar	otro
al	este	papá
algún	flor	para
allí	gato	pero
aquí	grande	perro
así	gustar	por
bien	haber	porque
bueno	hacer	que
casa	he aquí	se
como	ir	ser (V.)
con	jugar	sí
consigo	la	sí
cuando	le	su
de	libro	tal vez
decir	lo	también
del	maestro	tan
después	mamá	te (Pron.)
día	más	tener
donde	me	todo
dos	mí	tres
el	mí	tú
él	mío	tuyo
ella	mirar	uno
en	mucho	ver
entonces	muy	vez
escuela	nada	y
ese	niño	yo

En esas primeras 87 vemos que no solamente no se nos proporciona información alguna sobre su frecuencia, sino que, además, las palabras aparecen por orden alfabético —lo cual mermaría mucho las posibilidades de aplicación práctica ulterior si no fuesen las de frecuencia más alta—, lo cual no sucede con los siguientes grupos de voces. Si nos fijamos con detenimiento en esa lis-

¹⁴ Milton A. Buchanan, *A Graded Spanish Word Book*, University of Toronto Press, Toronto, 1927.

¹⁵ Victor García Hoz, *Vocabulario usual, vocabulario común y vocabulario fundamental*, CSIC, Madrid, 1953.

ta, vemos aparecer palabras que designan la realidad más inmediata, aunque desde fuera no parece que deban constar entre las 100 más frecuentes de la lengua, como, por ejemplo *he aquí, tal vez, maestro, escuela, gato, perro* o *tres*. Las siguientes palabras de su lista son:

LAS SIGUIENTES 112 UNIDADES LÉXICAS MÁS FRECUENTES DE LAS
PRIMERAS 500 (112)

<i>Unidad léxica</i>	<i>Frecuencia ponderada</i>	<i>Unidad léxica</i>	<i>Frecuencia ponderada</i>	<i>Unidad léxica</i>	<i>Frecuencia ponderada</i>
dar	35958	vivir	9599	comprar	7260
poder	29188	dejar	9573	leche	7089
querer	25048	blanco	9406	siempre	7066
árbol	19601	salir	9393	traje	7060
nuestro	17542	cosa	9239	nuevo	7058
bonito (Adj.)	17274	usted	9232	nombre	7019
hombre	15566	madre	9222	vaca	6935
saber (V.)	15550	tierra	9189	azul	6913
venir	14728	cuadro	9116	tiempo	6890
comer	13664	mismo	9070	leer	6886
poner	13538	lápiz	8985	mundo	6878
llamar	13527	sin	8966	color	6855
correr	13345	encontrar	8930	mañana	6844
pasar	13114	noche	8700	lindo	6785
llevar	12978	papel	8687	calle	6772
señor	12935	sol	8662	aquel	6759
alguno	12776	quien	8594	luna	6716
llegar	12736	hermano	8474	traer	6709
pájaro	12367	sala	8393	mesa	6644
Dios	11940	juego	8143	malo	6636
pueblo	11820	quedar	8133	carro	6578
cantar (V.)	11725	parte	8004	río	6522
éste	11555	rey	7986	mejor	6518
campo	11516	estrella	7976	muchacho	6433
ellos	11478	hasta	7749	mar	6327
año	11468	hablar	7678	poco	6321
amigo	11075	silla	7670	trabajar	6318
o	10712	nosotros	7668	persona	6306
caballo	10620	gente	7613	creer	6270
padre	10534	gallina	7588	mano	6266
animal	10326	vida	7578	muñeca	6253
ya	10202	oír	7521	alto	6251
hijo	10178	pequeño	7496	clase	6217
coger	10158	pues	7447	cielo	6163
sobre	10143	buscar	7407	echar	6100
escribir	10011	ayudar	7382	jardín	6081
deber (V.)	9979	dulce	7372	tomar	6066
primero	9689				

El más reciente de nuestros diccionarios de frecuencias¹⁶ es el que presenta un conjunto léxico más moderno, como es lógico, ya que recoge textos de los 15 años anteriores a su publicación, y es también el que se basa en un corpus amplio, al menos más amplio que el de algunos de sus predecesores: 2.000.000 de palabras. No se limita a ofrecer las frecuencias léxicas, sino también de los grafemas, de los bigramas y de las sílabas en español, todo ello de un interés innegable por las aplicaciones didácticas que puede tener, así como en logopedia, psicología del aprendizaje y otros dominios que se interesan por la lengua. Por lo que respecta al léxico, ofrece tanto las palabras ordenadas alfabéticamente como por frecuencias, como había hecho Rodríguez Bou, aunque tampoco incluye la frecuencia relativa de aparición de cada una de las formas registradas. En general, sus datos, al menos en las frecuencias más altas, son equiparables a los otros recuentos que estamos viendo; así, las 160 palabras más frecuentes que registra son:

de:	116480	qué:	3447	estaba:	1925	hacer:	1292
la:	78770	ni:	3395	menos:	1891	cosas:	1274
que:	67636	entre:	3370	ese:	1854	cómo:	1271
y:	60094	muy:	3233	sino:	1841	veces:	1249
el:	54027	hasta:	3146	día:	1798	todas:	1232
en:	49542	nos:	3049	casa:	1775	ver:	1219
a:	44456	porque:	3019	mundo:	1775	madre:	1209
los:	31399	ella:	2974	han:	1740	noche:	1209
se:	28254	son:	2897	esa:	1736	sido:	1209
un:	25119	sólo:	2883	poco:	1725	mejor:	1194
no:	24135	dos:	2865	después:	1695	he:	1174
las:	21768	vez:	2837	donde:	1692	esto:	1164
con:	20405	también:	2789	aunque:	1687	luego:	1151
una:	19973	él:	2725	tanto:	1673	toda:	1150
por:	19723	tan:	2702	tiene:	1671	aquí:	1148
del:	19170	desde:	2683	pues:	1645	hecho:	1140
su:	16533	así:	2515	sí:	1621	aquella:	1133
lo:	15375	te:	2498	dijo:	1607	momento:	1110
es:	14453	este:	2459	(50%)		tres:	1106
al:	12290	bien:	2454	hace:	1601	mano:	1099
como:	11958	hay:	2428	antes:	1585	quien:	1080
para:	11493	siempre:	2425	entonces:	1564	mis:	1073
más:	9401	esta:	2368	les:	1532	ante:	1072
o:	8945	vida:	2293	hacia:	1525	misma:	1053
me:	8648	todos:	2287	unos:	1512	otras:	1039
pero:	8226	fue:	2265	mucho:	1504	tu:	1039
le:	7908	años:	2255	ojos:	1466	sea:	1033
sus:	7057	tiempo:	2226	e:	1441	forma:	1031
si:	5856	eso:	2163	nunca:	1423	padre:	1027
sin:	5324	ahora:	2145	tenía:	1417	fin:	1023
era:	5201	mismo:	2138	decir:	1416	durante:	1018
había:	5161	nada:	2097	casi:	1410	mas:	1011
ya:	4772	otro:	2096	otros:	1383	nadie:	1006
cuando:	4738	está:	2050	ellos:	1369	habían:	1003
mi:	4499	cada:	1993	mientras:	1354	alli:	998
todo:	4389	uno:	1988	tal:	1350	cuerpo:	998
ha:	4144	puede:	1974	aquel:	1336	jugar:	998
sobre:	4106	algo:	1969	mujer:	1335	contra:	974
o:	3697	hombre:	1965	parte:	1300	aún:	964
ser:	3467	otra:	1953	gran:	1292	tarde:	960
						caso:	953

En este recuento, las 20 primeras palabras suman un total 739 602 apariciones, lo que representa un 36.98% de todo el léxico (en mi corpus son un 37.96%), mientras que las 40 primeras suman 868 592 apariciones, esto es, un 43.43% (en mi corpus eran un 42.9%). Para alcanzar el 50% de las palabras de un texto hay que llegar al rango 98. Si tuviéramos que representar con un

¹⁶ José Ramón Alameda y Fernando Cuetos, *Diccionario de frecuencias de las unidades lingüísticas del castellano*, 2 vols., Universidad de Oviedo, Oviedo, 1995. Para facilitar la consulta, las negritas son mías, así como las cursivas (que aparecen cada veinte palabras).

gráfico el descenso de frecuencia conforme avanzamos en el rango, nos ofrecería una curva con una caída muy fuerte en el inicio, y un descenso suave en el resto de las apariciones.

Son las primeras formas de esos elencos, y las que les siguen inmediatamente en orden de frecuencia, las que tendríamos que enseñar en primer lugar, pues no solamente son las más abundantes, sino también las que aparecen en todos los textos, esto es, son de un alto rendimiento funcional.

Creo que no es necesario advertirlo: cuando hablo de enseñar el léxico no me estoy refiriendo a hacer que el alumno se aprenda de memoria listas de palabras más o menos organizadas, de 10, 15 o 20 unidades, de un día para otro, como hemos tenido que sufrir más de uno de nosotros cuando hemos aprendido idiomas extranjeros, y no solamente los modernos. Estos índices de frecuencia deben servirnos para buscar palabras de un mismo orden que enseñáremos a nuestros alumnos de la manera que creamos más conveniente, mediante juegos, ejercicios variados, a partir de textos que, en la mayor parte de las ocasiones, habremos creado nosotros mismos, pues aquí radica una de las tareas que nos corresponden como profesores, y para la que han de ser de gran ayuda estas listas.

Pero no olvidemos que no todas las palabras tienen la misma distribución. Si consideramos el léxico de la lengua no como un continuo, sino como conjuntos de vocabularios, no nos vale con saber cuáles son las palabras más frecuentes, sino que también deberíamos saber cuáles son las que tienen una mayor dispersión, por lo que la enseñanza del léxico no puede enfocarse, sin más, hacia las más frecuentes, sino que deberemos atender a las que cubran un mayor número de campos, o que figuren con mayor intensidad aquel campo que deseamos enseñar, sobre todo cuando nuestra docencia va orientada hacia lenguajes con fines específicos. La motivación y las necesidades de los alumnos son las que nos obligarán a dar cuenta de las palabras que sirven para nombrar los objetos más inmediatos, las que designan los hechos y actividades cotidianos, etc., no siempre en los primeros rangos de frecuencia (por ejemplo, *perro, gato, pan o coche*).

Un primer planteamiento de lo que es la dispersión en la estadística léxica para nuestra lengua fue el realizado por García Hoz, quien, al buscar los textos para constituir su fuente de información y establecer lo que él consideraba como vocabulario usual, delimitó cuatro ámbitos: la vida familiar, la vida social indiferenciada, la vida social regulada y la vida cultural, lo cual le llevó a recoger el léxico, para cada uno de esos ámbitos de diferentes tipos de textos, cartas, periódicos, documentos oficiales y libros, respectivamente, si bien esa distribución se quedó ahí, como podemos ver en la página 28.

Al comentar esos datos, escribe García Hoz que «las palabras que se hallan en todos tipos de vocabulario son ciertamente las que pueden considerarse más generales; en ellas está el fondo común del vocabulario español, ya que son las palabras que necesariamente ha de utilizar el español en cualquier manifestación de su vida, tanto en la vida familiar cuanto en la cultural y la social. Este conjunto de palabras que figuran en todos los tipos de vocabulario constituyen el que llamo *vocabulario común*»¹⁷. Una vez que hubo determinado el vocabulario usual, quiso llegar a ese vocabulario común, que estaría constituido por las palabras del vocabulario usual que aparecen en todos los tipos de vocabulario, y que, por ello, son las más importantes, pues son las más generales, y, por tanto, las más usuales; valga como muestra la página 29.

Sin embargo, sucede que palabras muy usuales no figuran en todos los tipos de vocabulario, motivo por lo que no pertenecen al vocabulario común, lo cual no quiere decir que sean menos importantes; unas de ellas son las que tienen una frecuencia absoluta muy elevada (mayor de 40 ocurrencias), para las que establece una lista separada; como muestra, la página 30.

Además, hay otro grupo de palabras que sin poseer una frecuencia elevada sí merece la pena tener en consideración, ya que expresan “ideas afines” a otras, unidas a las cuales acumulan unas frecuencias dignas de atención, y que aparecen en la misma lista, con la frecuencia acumulada en la más frecuente, sin que se indique cuáles son las formas que se unen (así, por ejemplo, *acentuar* figura en esa lista con 41 apariciones ya que *acentuar* tiene 29 y *acento* —que no está— 12).

Por otro lado, no todas las palabras pueden tener el mismo valor, pues hay un factor muy importante que se debe considerar: la presencia que tienen en los diferentes tipos de vocabularios, esto es, su distribución. García Hoz, habla de la correlación entre los distintos grupos del vocabulario común. Así, mediante unos cálculos estadísticos, llega a establecer el vocabulario fundamental, «constituido por las palabras que tienen sus frecuencias prácticamente distribuidas en

¹⁷ García Hoz, *Vocabulario* ..., pág. 385.

Número	PALABRA	Cartas	Periódicos	Documentos oficiales	Libros	TOTAL
4.181	Dominar...		18	3	6	27
4.182	Domingo...	62	46	17	3	128
4.183	Dominica...			5		5
4.184	Dominical...		2	2		4
4.185	Dominicano...	1	2			3
4.186	Dominico...	3				3
4.187	Dominio...		7	2	5	14
4.188	Don...	135	224	209	139	707
4.189	Donar...		3	1		4
4.190	Donativo...		6	3	1	10
4.191	Doncel...				1	1
4.192	Doncella...	1				1
4.193	Donde...	76	81	62	91	310
4.194	Donjuán...				3	3
4.195	Donjuanesco...				9	9
4.196	Donjuanismo...				2	2
4.197	Donostiarra...		1	1		2
4.198	Dofia...	1	16		46	63
4.199	Doquiera...		1	1	2	4
4.200	Dorado...	5			10	15
4.201	Dormir...	18	4	1	22	45
4.202	Dormitar...				1	1
4.203	Dormitorio...	1	5		1	7
4.204	Dorso...			3	1	4
4.205	Dos...	156	146	70	248	620
4.206	Doscientos...	1	9	1	6	17
4.207	Dosel...		1			1
4.208	Dosificar...		1	1		2
4.209	Dosis...			1	8	9
4.210	Dotación...			5		5
4.211	Dotar...	1	6	21		28
4.212	Dote...	2			2	4
4.213	Dozavo...			1		1
4.214	Dracma...				1	1
4.215	Dragón...			2		2
4.216	Drama...	1	7	1	5	14
4.217	Dramáticamente...		1		1	2
4.218	Dramático...		20		6	26
4.219	Dramatismo...				1	1
4.220	Dramaturgo...		1		2	3

Numero	PALABRA	Cartas	Periódicos	Documentos oficiales	Libros	TOTAL
391	Comprobar...	3	6	1	7	17
392	Comprometer...	3	6	2	4	15
393	Común...	2	13	16	15	46
394	Comunicación...	4	7	16	4	31
395	Comunicar...	50	27	24	10	111
396	Comunidad...	3	14	13	6	36
397	Comunista...	1	80	9	1	91
398	Con...	1.001	853	924	826	3.604
399	Conceder...	6	27	56	11	100
400	Concentración...	3	5	6	4	18
401	Concepto...	8	9	22	22	61
402	Conciencia...	1	9	16	7	33
403	Concretar...	2	1	3	6	12
404	Concreto...	2	4	5	9	20
405	Condición...	5	38	42	22	107
406	Conducir...	3	16	6	24	49
407	Conducta...	2	3	9	4	18
408	Conducto...	2	4	3	1	10
409	Conferencia...	6	62	22	1	91
410	Confesión...	3	2	5	1	11
411	Confesor...	11	2	3	4	20
412	Confianza...	8	12	10	11	41
413	Confiar...	10	10	6	10	36
414	Confidencia...	14	2	1	1	18
415	Confirmación...	1	3	2	1	7
416	Confirmar...	5	10	6	3	24
417	Conforme...	4	3	18	2	27
418	Confundir...	1	1	2	5	9
419	Congreso...	4	28	16	1	49
420	Conjunto...	10	27	13	11	61
421	Conocer...	85	32	37	48	202
422	Conocido...	13	16	3	12	44
423	Conocimiento...	6	11	46	20	83
424	Conquista...	4	6	5	2	17
425	Consecuencia...	4	24	25	7	60
426	Consciente...	1	1	1	3	6
427	Conseguir...	13	22	37	39	111
428	Consejo...	9	59	67	13	148
429	Conservar...	3	5	13	28	49
430	Considerable...	2	3	6	2	13

Número	PALABRA	Cartas	Periódicos	Documentos oficiales	Libros	TOTAL
J						
123	Juzgado... ..	2	21	38		61
K						
124	Kilo... ..	18	15		11	44
L						
125	Laboral... ..	2	21	31	4	58
126	Legal... ..	1	9	30	2	42
127	Límite... ..	3	13	20	40	76
128	Literario... ..		1	14	32	47
129	Lo... ..		19		280	299
130	Lote... ..	3	18	24		45
131	Lunes... ..	22	37		3	62
M						
132	Madrina... ..	43	5			48
133	Mamá... ..	88	2		19	109
134	Matar... ..	20	17	1	20	58
135	Mayo... ..	24	23	56		103
136	Mina... ..		17	19	9	45
137	Moderno... ..		9	17	14	40
138	Moral... ..		10	29	12	51
139	Mostrar... ..		8	10	25	43
N						
140	Nene... ..	87			11	98
141	Novedad... ..	32	11		3	46

partes iguales en uno y otro tipo de vocabulario, mientras que habrá otras que, aun siendo comunes, son más propias de un tipo de vocabulario que de otros, porque la distribución de sus frecuencias es claramente desigual en unos vocabularios que en otros»¹⁸. Ese vocabulario fundamental está constituido por 208 palabras, que él presenta en tablas alfabéticas.

El corolario que se desprende de todo ello, al menos para nuestros intereses actuales, salvando las distancias del tiempo transcurrido, y la necesaria actualización de esos datos, aunque no dejen de ser un buen punto de partida, es que en los primeros momentos de la enseñanza de nuestro léxico a estudiantes extranjeros, los alumnos tendrán que aprender las 208 palabras del vocabulario fundamental. A continuación, y durante la primera etapa, en el primer nivel, habrán de enseñarse las restantes palabras que constituyen el vocabulario común, a las que se podrían unir las 97 más frecuentes de las que no aparecen en todos los tipos de vocabulario, así como las 115 que expresan ideas afines y cuyas frecuencias se suman. Aquí, por otro lado, se ve la importancia de la concepción de la enseñanza del vocabulario, que no es tan sólo la enseñanza de unidades aisladas, como si las unas no tuvieran relaciones con las otras. Durante el proceso de enseñanza hay que ir introduciendo las relaciones formales y semánticas de las palabras, aunque de una manera muy limitada en los primeros momentos: no se trata de enseñar campos léxicos o semánticos de forma intensa, sino palabras estrechamente relacionadas. Por eso mismo, por ejemplo, cuando enseñemos un verbo, no podemos enseñar el infinitivo, o no solamente el infinitivo (nuestros alumnos tienen que aprender a reconocer la palabra), sino aquella forma del paradigma que sea más frecuente (y cuando se enseñe la morfología verbal se irán introduciendo las demás formas).

Tal vez el diccionario de frecuencias más conocido del español sea el de Alphonse Juilland y Ernesto Chang Rodríguez¹⁹, cuya muestra no parece muy extensa, 500.000 palabras, cantidad que se ha tomado en otros recuentos de ámbito más restringido, como el del español puertorriqueño²⁰, que no voy a tener en cuenta aquí, si bien no se me esconde la importancia que pueden tener ese tipo de trabajos en el ámbito dialectológico o sociolingüístico, sino también en la enseñanza del español como lengua extranjera, pues nos informan de las voces más empleadas en el dominio o ámbito en que se ha efectuado la recogida, y que nos pueden mostrar los términos particulares que también habremos de enseñar para que los alumnos no se sientan desvinculados del entorno en que se mueven²¹. Sin embargo, esto no deja de plantear muchos otros problemas que nos obligarían a un análisis de esos repertorios, y a señalar la necesidad de su elaboración para la mayor parte de las zonas geográficas del español, por no hablar, de nuevo, de los fines específicos. En este caso, los conocimientos y experiencia del profesor son necesarios para enseñar aquellas palabras que pueden vincular al alumno con su entorno más inmediato.

No tomaron Juilland y Chang Rodríguez en consideración textos hispanoamericanos, como hicieron Buchanan o Rodríguez Bou, ni dialectales, como este último o García Hoz, lo cual no se podrá olvidar cuando los empleamos como fuentes para preparar materiales u organizar las clases. Los suyos son textos de entre 1920 y 1940, de todos modos más modernos que los de Buchanan, quien partía del siglo XVII.

Cuando Juilland y Chang Rodríguez realizaron su diccionario de frecuencias tuvieron en cuenta los trabajos anteriores, y establecieron unas listas de palabras algo más elaboradas que las que hallamos en los demás recuentos, basándose en tres criterios, similares a los de García Hoz: la frecuencia absoluta, el uso y la dispersión, como puede verse en la siguiente página:

Como se aprecia a través de las muestras, no son grandes las diferencias que existen al considerar las listas de palabras por frecuencia absoluta y por el uso, aunque las hay. Sin embargo, es muy diferente la lista de acuerdo con la dispersión, tanto que llega a ser desconcertante, al menos en la idea apriorística que tenemos del léxico de la lengua. Y ello es así por las fórmulas empleadas para calcular la dispersión: una palabra cuyas ocurrencias se distribuyen de la misma manera en los ámbitos considerados, esto es, tiene la misma frecuencia en cada uno de los universos, alcanza el 100% de dispersión, con lo que se priman las voces de menor frecuencia: si se consideran cinco universos (como hacen Juilland y Chang Rodríguez), una palabra que tuviera sólo 5 ocurrencias, pero con una en cada uno de los ámbitos, nos daría una dispersión del 100%. Eso es

¹⁸ *Op. cit.*, pág. 468.

¹⁹ Citado.

²⁰ Amparo Morales, *Léxico básico del español de Puerto Rico*, Academia Puertorriqueña de la Lengua Española, San Juan de Puerto Rico, 1986. Sigue los principios de Juilland y Chang Rodríguez, para establecer, a partir del uso, un léxico básico de 4456 palabras. En el *Recuento de vocabulario de preescolares* dirigido por Rodríguez Bou, citado, fueron 470 026 palabras.

²¹ En este sentido puede ser útil el trabajo de Rosario Guillén Sutil, *El habla culta de Sevilla*, Alfár, Sevilla, 1987.

opaco aj.	3.43	9	38.14
	0	2	0 1 6
opaca	2	0	1 0 0 1
opacas	1	0	0 0 0 1
opaco	2	0	1 0 1 0
opacos	4	0	0 0 0 4

operación n.	22.21	33	67.29
	1	5	5 14 8
operación	15	1	3 4 2 5
operaciones	18	0	2 1 12 3

operar v.	6.15	11	55.93
	1	1	6 1 2
opera	5	0	0 5 0 0
operamos	1	0	0 0 0 1
operando	1	0	0 0 1 0
operar	3	0	1 1 0 1
operé	1	1	0 0 0 0

opinar v.	15.68	19	82.54
	6	2	4 3 4
opina	2	1	0 0 1 0
opinaba ^a	3	1	0 0 0 2
opinaban	1	0	0 1 0 0
opinabas	1	1	0 0 0 0
opinado	1	0	0 0 1 0
opinan	2	0	0 1 1 0
opinar	4	2	0 2 0 0
opinaran	1	0	0 0 0 1
opinarán	1	0	1 0 0 0
opinaria ^a	1	0	0 0 0 1
opino	1	1	0 0 0 0
opinó	1	0	1 0 0 0

opinión n.	64.44	87	74.07
	8	6	26 19 28
opinión	62	7	6 18 15 16
opiniones	25	1	0 8 4 12

oponer v.	26.70	31	86.13
	7	3	8 7 6
opondría ^a	1	0	0 0 0 1
opone	1	0	0 0 1 0
-opone	5	1	1 2 0 1
oponen	5	0	0 3 2 0

-oponen	1	0	0 0 0 1
oponer	9	4	1 2 0 2
-opongas	1	1	0 0 0 0
opongo	2	0	0 0 2 0
-oponía	1	0	1 0 0 0
oponía ^a	1	0	0 0 1 0
oponiendo	1	0	0 0 0 1
opuesto	1	0	0 1 0 0
-opuse	1	1	0 0 0 0
opusieron	1	0	0 0 1 0

oportuno aj.	16.96	23	73.73
	3	5	2 9 4
oportuna	5	0	2 1 2 0
oportunas	5	1	0 0 3 1
oportuno	12	2	2 1 4 3
oportunos	1	0	1 0 0 0

oposición n.	11.92	24	49.65
	0	4	14 4 2
oposición	17	0	3 8 4 2
oposiciones	7	0	1 6 0 0

opresión n.	4.42	9	49.08
opresión	9	0	2 2 5 0

optimismo n.	5.82	10	58.17
optimismo	10	0	4 3 3 0

optimista aj.	5.56	12	46.32
	0	2	3 7 0
optimista	8	0	2 3 3 0
optimistas	4	0	0 0 4 0

opuesto aj.	18.02	31	58.13
	0	1	11 6 13
opuesta	8	0	0 4 2 2
opuestas	1	0	0 0 0 1
opuesto	16	0	1 5 3 7
opuestos	6	0	0 2 1 3

oración n.	10.53	15	70.19
	0	4	4 5 2
oración	10	0	1 4 4 1
oraciones	5	0	3 0 1 1

lo que explica los resultados algo sorprendentes, y que la mayor parte de las palabras tengan un índice de dispersión entre el 45% y el 75%²². Por esto es conveniente, en la enseñanza del léxico a partir de estos índices de frecuencias, considerar la frecuencia absoluta, y el uso, informaciones que los autores ofrecen en la parte alfabética de su diccionario, tal y como se aprecia en la página reproducida antes.

Antes de calcular el uso de las palabras, Juilland y Chang Rodríguez pasan revista a todas las posibilidades que hay para hacerlo, entre otras la de García Hoz. Su pretensión es la de proporcionar un coeficiente que nos sirva para admitir o rechazar una palabra en el vocabulario básico de la lengua, que llegan a establecer en 5024 palabras, tras la aplicación de una serie de fórmulas. A nosotros no nos interesa tanto cómo se llega a esos resultados, cuáles son las técnicas empleadas, incluso si son correctas o no —yo no estoy en condiciones de juzgarlo—, sino los resultados mismos y sus posibilidades de aplicación. Y en este sentido, igual que decía al hablar del recuento de García Hoz, parece que son esas 5024 palabras las que deberán constituir la base de la enseñanza del léxico, y no creo que exclusivamente en el primer nivel, sino también en el siguiente, por lo que habrá que hacer una distribución de ese conjunto de palabras, y comenzar la enseñanza por las de mayor uso.

Bien es cierto que entre los factores distorsionantes en cualquier recuento que hagamos está la representatividad de la muestra, la adecuada elección de las fuentes para los fines perseguidos. De todos modos, en la enseñanza del léxico no podemos desviarnos de nuestros objetivos ni ignorar las necesidades de los alumnos, por lo que tendremos que seleccionar aquellas palabras que se adecuen a los ámbitos lingüísticos que queremos enseñar, esto es, tendremos que saber cuál es la dispersión en los campos de nuestro interés, y eso, por ahora, no nos lo proporcionan los diccionarios de frecuencias ni los corpus, aunque en algunos de estos, por ejemplo, en el CREA de la Real Academia Española, se pueden buscar las palabras en los diferentes dominios en que se han distribuido los textos que lo constituyen.

Los índices de frecuencias nos proporcionan, como vemos, valiosas informaciones sobre el uso de las palabras, por más que no siempre se hayan tenido presentes en la organización de materiales para la enseñanza de español a extranjeros. Pero no solamente son listas de palabras y números. Si las analizamos con un poco de atención obtendremos de ellas otras informaciones en las que habitualmente no pensamos, y que pueden indicarnos sobre qué puntos insistir, o pueden ayudarnos a organizar la enseñanza, sobre todo si ponemos en relación unas palabras con otras y cotejamos sus frecuencias. En mi librito sobre *La enseñanza del léxico y el uso del diccionario* advertía cómo en las series bipolares los dos términos no aparecen con la misma frecuencia, por lo que consideraba que, desde el punto de vista formal, debemos tomar al más frecuente como no marcado, mientras que el menos frecuente sería el marcado, lo que nos debería llevar a enseñar en primer lugar aquél, y a partir de él dar cuenta del otro, independientemente de su rango de frecuencia²³; veamos unos cuantos de estos pares entresacados de los datos que nos proporciona el corpus Vox-Bibliograf:

²² Véase el gráfico que ofrecen en la pág. LX.

²³ Pág. 102 del libro. Tomo los ejemplos que ofrecía allí y añadido alguno más en la muestra que sigue.

PAREJAS	OCURRENCIAS	PAREJAS	OCURRENCIAS
gran, grande	15704	duro, -a	1568
pequeño, -a	4321	blando, -a	210
chico, -a	1460	verdadero, -a	1526
mucho, -a	14372	falso, -a	665
poco, -a	9117	simple	1396
nuevo, -a	13185	complicado, -a	414
antiguo, -a	2543	cerca	1610
buen, -o, -a	11443	lejos	1017
mal, -o, -a	3931	rico, -a	1322
vida	7798	pobre	992
muerte	2600	encima	1309
después	6815	debajo	723
antes	6308	salir (sólo inf.)	1235
bajo, -a	5443	entrar (sólo inf.)	974
alto, -a	4526	rápido, -a	1096
claro, -a	4926	lento, -a	571
oscuro, -a	735	estrecho, -a	913
oscuro, -a	4	ancho, -a	394
largo, -a	4500	abrir (sólo inf.)	463
corto, -a	1163	cerrar (sólo inf.)	321
fuera	3439	delgado, -a	365
dentro	4733	flaco, -a	64
tarde	3105	grueso, -a	303
pronto	1599	gordo, -a	234
joven	2953	subir (sólo inf.)	317
viejo, -a	2254	bajar (sólo inf.)	261
ir (sólo inf.)	1916	guapo, -a	194
venir (sólo inf.)	509	feo, -a	148
abierto, -a	1714	levantarse (sólo inf.)	83
cerrado, -a	935	acostarse (sólo inf.)	25

Esa misma distribución asimétrica se repite también con algunos heterónimos, como:

hombre	11039
mujer	6005
caballo	836
yegua	31
materno, -a	121
paterno, -a	94
yerno	41
nuera	19

si bien en algunos casos las diferencias desaparecen, como en:

oveja	181
cordero	180

Y si nos acercamos a algunas series, veremos que no todos sus elementos constan con la misma frecuencia, y que las diferencias pueden ser notables; por ejemplo, los colores o los días de la semana:

blanco	2166
negro	1942
rojo	1409
verde	829
azul	760
amarillo	519
gris	333
marrón	90

domingo	1262
sábado	722
viernes	686
lunes	652
jueves	482
martes	476
miércoles	422

Dentro de las series de grados, no todas las palabras aparecen con la misma frecuencia, destacando aquellas que se toman como representativas de la serie, que, de hecho, funcionan como opuestas, como se puede apreciar en:

frío, -a	1365
caliente	339
helado, -a	271
templado, -a	192
tibio, -a	80
ardiente	73

seco, -a	699
húmedo, -a	414
mojado, -a	93
empapado, -a	50

Esa falta de homogeneidad entre los elementos se aprecia, igualmente, fuera del ámbito estrictamente léxico, en los paradigmas, como el de los pronombres personales:

yo	11577
ellos,-as	8750
Él	8174
usted,-es	3193
Tú	2711
nosotros,-as	2582
vosotros,-as	279

Podemos hacer búsquedas similares con otras piezas léxicas y averiguar cómo se comportan, desde el punto de vista estadístico, dentro de algunas series, especialmente si pensamos las asociaciones que se producen por lo nombrado –la realidad extralingüística–; valgan como ejemplo algunos nombres de parentesco:

hijo	3681
hija	1029
padre	3388
madre	2404
hermano	1275
hermana	519
tío	989
tía	380
prima	485
primo	245
abuelo	368
abuela	236
nieto	196
nieta	45
sobrino	107
sobrina	54
cuñado	67
cuñada	29

En otros grupos de palabras se aprecian diferencias similares, como en:

sierra	1293 ²⁴
montaña	1021
monte	669
cerro	257
colina	156

²⁴ Es la frecuencia de la forma *sierra*, donde confluyen la elevación del terreno y el instrumento. No se me oculta que en muchas de las formas presentadas confluyen homónimos y sentidos diferentes.

mano	4166
frente	3736 ²⁵
ojo	2509
cabeza	2287
pie	1707
brazo	958
corazón	931
boca	925
oreja	630
espalda	593
dedo	570
pierna	540
hombro	404
labio	370
cuello	333
diente	330
pecho (sólo sing.)	314
pulmón	219
rodilla	266
nariz	198
cintura	126
cadera	95
muslo	93
codo	92
mejilla	92

Todo ello puede constituirse en una guía interesante no sólo para la enseñanza del léxico y de algunos sistemas léxicos, sino también para la enseñanza de las formas de los paradigmas e, incluso, de la morfología, como acabamos de ver. Por supuesto, hemos de saber qué palabra es más frecuente para enseñarla en primer lugar, pero no podemos ignorar cuál de las formas de su paradigma es la más común, la menos marcada, la que le aparecerá antes al alumno, o un mayor número de veces, y que será la que debemos enseñarle por delante de las demás, o la que debemos presentarle en primer lugar en los ejercicios o lecturas; por ejemplo, si consideramos que hemos de enseñar la palabra *mucho* no estaría de más echar un vistazo en un corpus para saber cuál es la frecuencia de sus diferentes formas:

mucho	5756
muchos	4092
muchas	3477
mucha	1047

Más problemas nos plantea la enseñanza de los paradigmas verbales, ya que no se enseña por separado cada una de sus formas, sino el sistema. No es una cuestión léxica, sino morfológica. De todos modos, consultando las frecuencias de las diferentes formas podremos saber cuáles son las más empleadas, y así plantear la enseñanza del paradigma y elaborar los ejercicios con aquellas personas y tiempos más frecuentes. Por ejemplo, para el verbo *acabar*, éstas son las frecuencias, por orden decreciente, de las formas simples más empleadas, de acuerdo con los datos del corpus que vengo manejando:

²⁵ En esta cantidad se incluye tanto la frecuencia del sustantivo como la de los usos adverbiales.

acabar	478
acabó	409
acabo	112
acabaron	110
acabe	87
acabará	85
acabaría	52
acabara	31
acabarán	26
acabarían	26
acaben	24
acabas	22
acabaré	13

Si vamos a trabajar con un ámbito designativo, sería conveniente que conociéramos la frecuencia de las distintas piezas léxicas que lo integran para organizar adecuadamente su enseñanza y preparar los ejercicios apropiados; así, por ejemplo, si nos interesan los medios de transporte veremos que algunas de las palabras que se utilizan para nombrarlos poseen el siguiente uso:

coche	1407
barco	872
avión	687
tren	606
automóvil	543
autobús	385
camión	301
taxi	224
bicicleta	214
moto	185
helicóptero	123
tranvía	81
motocicleta	55

Hasta hace bien poco nos encontrábamos con la falta de este tipo de materiales, pero hoy comienzan a ser una realidad que tenemos bien próxima. Tal vez nada de ello sea excesivamente relevante, pero son datos objetivos y elementos que se encuentran a nuestra disposición, de los que debemos obtener el mayor rendimiento posible, y utilizar de la manera más adecuada. Es nuestra obligación como profesores.

Pero también es nuestra obligación decir cuáles son nuestras necesidades para que, en el futuro, quienes se encargan de constituir los corpus y las herramientas para su consulta pongan a nuestra disposición unos elementos que respondan a la situación actual de la lengua, de los cuales podamos extraer de una forma sencilla y rápida las informaciones en las que basar nuestra enseñanza, o con los que preparar lecturas graduadas o los ejercicios que deseemos; son las ventajas que nos ofrecen las nuevas tecnologías. Sin duda, los corpus pueden prestarnos una gran ayuda, como hemos podido ver, pero si las posibilidades de consulta fuesen más adecuadas a nuestras necesidades podrían constituirse en una herramienta ineludible; para ello habrían de mostrarnos no sólo los índices de frecuencia de voces (tanto generales como por tipos de textos, absolutas y relativas), sino los de las palabras lematizadas, e, incluso, los de las familias léxicas, amén de permitirnos seleccionar los ámbitos de nuestro interés —para el caso en que nos dediquemos a la enseñanza con fines específicos—, o el establecer relaciones, siquiera elementales, de tipo morfológico, léxico, semántico o pragmático. Disponemos de suficientes corpus de la lengua española, y el software para hacer los recuentos no reviste dificultades mayores; además, se están desarrollando las herramientas apropiadas para realizar los otros tipos de consultas, para lo que contamos ya con excelentes analizadores y generadores morfológicos, y se avanza en el aná-

lisis automático del contenido. No se trata de ciencia ficción. Falta únicamente la claridad en las ideas para saber qué queremos y para qué, y la voluntad de ponerlo en marcha. Ojalá pronto podamos preparar nuestras clases sin tener que recurrir a viejos libros, por muy certeros que sean, que contienen datos que no reflejan la realidad de nuestros días, y que podamos hacer las consultas de una manera ágil y sencilla. Lo demás tendrá que ser de nuestra cosecha, ideando cómo queremos efectuar la enseñanza en cada momento para transmitir los conocimientos deseados. Esperemos que así sea.

BIBLIOGRAFÍA CITADA

- Alameda, J. R. y Cuetos, F. (1995): *Diccionario de frecuencias de las unidades lingüísticas del castellano*, 2 vols., Oviedo, Universidad de Oviedo.
- Alvar Ezquerro, M. (2003): *La enseñanza del léxico y el uso del diccionario*, Madrid, Arco-Libros.
- Buchanan, M. A. (1927): *A Graded Spanish Word Book*, Toronto, University of Toronto Press.
- DIPELE, *Diccionario para la enseñanza de la lengua española*, Barcelona, Bibliograf-Universidad de Alcalá, 1995.
- García Hoz, V. (1953): *Vocabulario usual, vocabulario común y vocabulario fundamental*, Madrid, CSIC.
- Guillén Sutil, R. (1987): *El habla culta de Sevilla*, Sevilla, Alfar.
- Juilland, A. y Chang Rodríguez, E. (1964): *Frequency Dictionary of Spanish Words*, Londres-La Haya-París Mouton.
- Morales, A. (1986): *Léxico básico del español de Puerto Rico*, San Juan de Puerto Rico, Academia Puertorriqueña de la Lengua Española.
- Patterson, W. y Urrutibéheity, H. (1975): *The lexical structure of Spanish*, La Haya-París, Mouton.
- Picoche, J. (1994): *Précis de lexicologie française. L'étude et l'enseignement du vocabulaire*. Nueva edición revisada y actualizada, París, Nathan.
- Rodríguez Bou, I. (1952): *Recuento de vocabulario español*, I, Río Piedras (Puerto Rico), Universidad de Puerto Rico.
- Rodríguez Bou, I., (dir.) (1967): *Recuento de vocabulario de preescolares*, Río Piedras, Universidad de Puerto Rico.
- Sauvageot, A. (1964): *Portrait du vocabulaire français*, París, Larousse.
- Thorndike, E. L. (1931): *The Teacher's Word Book*, 1921, revisado en *A Teacher's Word Book of Twenty Thousand Words Found Most Frequently and Widely in General Readings for Children and Young People*, Nueva York, Columbia University.
- Thorndike, E. L. y Lorge, I. (1944): *The Teacher's Word Book of 30,000 Words*, Nueva York, Columbia University.