



UNIVERSIDAD DE SALAMANCA
FACULTAD DE FILOLOGÍA
DEPARTAMENTO DE LENGUA ESPAÑOLA

**MÉTODOS DE SELECCIÓN LÉXICA
APLICADOS A LA ENSEÑANZA.
EL VOCABULARIO FUNDAMENTAL DEL
ESPAÑOL**

Tesis doctoral presentada por Xandra Santos Palmou

Dirigida por el Prof. Dr. D. Julio Borrego y por la Prof. Dra. D^a. Inmaculada Mas

Vº Bº del director,

Vº Bº de la directora,

Salamanca, junio de 2017

Índice

RESUMO	9
INTRODUCCIÓN.....	11
CAPÍTULO 1. LA COMPETENCIA LÉXICA	15
1.1. El lexicón mental	15
1.2. El concepto de competencia léxica	20
1.3. ¿Qué significa conocer o aprender una palabra?	23
1.4. Vocabulario receptivo y vocabulario productivo	26
1.5. Léxico y vocabulario	28
1.6. Unidades léxicas y palabras	29
CAPÍTULO 2. EL DESARROLLO DE LA COMPETENCIA LÉXICA	37
2.1. Métodos y enfoques didácticos	40
2.2. Tipos de enseñanza del léxico.....	46
2.3. Características del vocabulario.....	52
CAPÍTULO 3. CRITERIOS Y MÉTODOS DE SELECCIÓN LÉXICA. ¿QUÉ VOCABULARIO ENSEÑAR?.....	63
3.1. Críticas a la selección léxica en los materiales de español como lengua extranjera (ELE)	66
3.2. Criterios de selección léxica	71
3.2.1. ¿Qué criterios deben guiar nuestra selección?.....	75
3.3. Métodos de selección léxica.....	83
3.4. Métodos objetivos de selección léxica.....	85

3.4.1. La frecuencia léxica	90
3.4.1.1. Los vocabulario básicos y los corpus de frecuencias	105
3.4.1.2. Puntos débiles de la frecuencia y los corpus textuales para la selección léxica	112
3.4.2. La disponibilidad léxica	114
3.4.2.1. Ventajas de la disponibilidad léxica en la selección del vocabulario	116
3.4.2.2. Algunas propuestas de aplicación de la disponibilidad léxica	138
3.4.3. El vocabulario fundamental	140
3.4.4. El vocabulario instrumental	143
3.4.5. La lingüística contrastiva	144
CAPÍTULO 4. EL NÚMERO DE UNIDADES LÉXICAS: ¿CUÁNTO VOCABULARIO SELECCIONAR PARA CADA NIVEL?	147
4.1. ¿Cuánto vocabulario seleccionar para cada nivel?	147
4.1.1. Cobertura léxica textual	150
4.1.2. Palabras, lemas, vocablos (<i>types</i>), palabras (<i>tokens</i>) y familias de palabras	156
4.2. Conocimiento léxico y comprensión lectora	166
4.3. Cobertura textual y tipos de texto	168
4.4. Umbral léxico de comprensión lectora y comprensión oral	172
4.4.1. Vocabulario visual	176
4.4.2. Inferencias léxicas	177
4.5. Palabras de baja frecuencia, cobertura léxica y lectura efectiva de textos	178
4.5.1. Número de palabras frecuentes para el nivel inicial	179

4.5.2. Número de palabras, cobertura léxica y umbral de lectura para los diferentes niveles del <i>Marco común europeo de referencias para las lenguas (MCERL)</i>	182
CAPÍTULO 5. LOS VOCABULARIOS BÁSICOS, EL LÉXICO FRECUENTE Y EL LÉXICO USUAL	187
5.1. Los corpus de frecuencias	188
5.1.1. Corpus lingüísticos, léxico frecuente, léxico usual y léxico básico	189
5.1.2. Rango, dispersión e índice de uso	200
5.1.3. Recuentos léxicos y lematización	202
5.2. Los listados de frecuencias más destacados	203
5.3. Léxicos frecuentes y léxicos básicos del español: obras escogidas	235
CAPÍTULO 6. DISPONIBILIDAD LÉXICA: HISTORIA, METODOLOGÍA Y APLICACIONES ...	237
6.1. Orígenes y definición	237
6.2. Panorámica de los trabajos de disponibilidad léxica	242
6.2.1. Cuestiones metodológicas: la fórmula matemática y los índices de disponibilidad léxica	245
6.2.2. El Proyecto Panhispánico de Estudio de la Disponibilidad Léxica (PPHDL)	249
6.3. Aplicaciones de la disponibilidad léxica	255
6.3.1. Disponibilidad léxica, sociolingüística y contacto de lenguas	255
6.3.1.1. Adaptaciones metodológicas	259
6.3.1.2. Cambios y ampliación de los centros de interés	263
6.3.1.3. Trasvases léxicos entre lenguas	264
6.3.2. Disponibilidad léxica y dialectología: cotejo de resultados entre estudios y sintopías	267
6.3.3. Disponibilidad léxica y enseñanza-aprendizaje del español como lengua extranjera (ELE)	270

6.3.3.1. Los centros de interés: ampliación y nuevas propuestas	276
6.3.3.2. Propuestas de selección del vocabulario	282
6.3.3.3. Propuestas para la enseñanza-aprendizaje del español para fines específicos (EpFE) desde la disponibilidad léxica	285
6.3.3.4. La disponibilidad léxica como medio de evaluación y diagnóstico de la competencia léxica	290
CAPÍTULO 7. EL VOCABULARIO FUNDAMENTAL	295
7.1. Origen del vocabulario fundamental y su metodología:	
el Francés Fundamental	295
7.1.1. El vocabulario fundamental y la didáctica de lenguas extranjeras	296
7.1.2. El Francés Fundamental	300
7.2. El Alemán Fundamental	307
7.3. El Español Fundamental	308
7.4. El corpus del Portugués Fundamental	313
7.5. Los vocabularios fundamentales del francés, del español y del portugués:	
semejanzas y diferencias	316
7.6. Metodología y criterios de selección de las palabras en los proyectos de	
vocabulario fundamental	323
CAPÍTULO 8. HACIA UN VOCABULARIO FUNDAMENTAL DEL ESPAÑOL	333
8.1. Sobre el concepto de vocabulario fundamental: definición	338
8.2. Análisis y comparación de cuatro repertorios: dos vocabularios	
disponibles, el vocabulario básico temático de Davies y las nociones	
específicas del Plan Curricular del Instituto Cervantes (PCIC)	345
8.3. Aportaciones para obtener el vocabulario	
fundamental del español	356
8.3.1. Fuentes usadas	358

A) Proceso de selección de las obras de vocabulario básico	358
B) Proceso de selección de las obras de vocabulario disponible	360
8.3.2. Metodología y criterios de selección	362
8.3.2.1. Criterios de selección del vocabulario básico	363
8.3.3.2. Criterios de selección del vocabulario disponible	365
8.3.4.3. Selección de las multipalabras	404
CAPÍTULO 9. ANÁLISIS DE LOS RESULTADOS Y DISCUSIÓN	415
9.1. Selección y organización del vocabulario	415
9.2. Propuesta de vocabulario fundamental por orden	
alfabético (Anexo 1)	419
9.3. Propuesta de vocabulario fundamental del español organizado	
en series, clases de palabras y temas (Anexos 2 y 3)	429
9.4. Organización temática del vocabulario fundamental	439
9.5. Observaciones sobre los datos de nuestro vocabulario fundamental	443
9.6. Futuros trabajos e investigaciones	448
CONCLUSÕES	451
REFERENCIAS BIBLIOGRÁFICAS	461
ÍNDICE DE CUADROS	
ÍNDICE DE FIGURAS	
ÍNDICE DE TABLAS	
ÍNDICE DE ABREVIATURAS Y SIGLAS	
ANEXOS (EN CD)	
Anexo 1. Vocabulario fundamental en orden alfabético	
Anexo 2. Vocabulario fundamental: series, clases de palabras y multipalabras	
Anexo 3. Vocabulario fundamental temático	

Anexo 4. Cotejo del léxico disponible, el léxico frecuente y el *PCIC*

Anexo 5. Límites establecidos en los histogramas para los datos del léxico disponible de Tomé Cornejo (2015)

Anexo 6. Límites establecidos en los histogramas para los datos del léxico disponible de López Meirama (2008)

Anexo 7. Gráficos realizados con los datos de Tomé Cornejo (2015)

Anexo 8. Gráficos realizados con los datos de López Meirama (2008)

RESUMO

O objetivo principal deste trabalho é achar um método de seleção e união do *léxico frequente* e o *disponível* para elaborar o *vocabulário fundamental*. Para isso, estudamos ambas linhas de trabalho, as listagens de frequência lexical derivadas da linguística do corpus e os estudos sobre disponibilidade lexical, com a intenção de entender quais são os seus resultados e as palavras às quais chegamos em cada caso. Consideramos que a aplicação principal deste vocabulário fundamental é o ensino-aprendizagem do espanhol como língua estrangeira, pelo que nos dois primeiros capítulos estudamos questões relacionadas com este aspeto e com as características do vocabulário, a competência lexical, os métodos e enfoques didáticos, etc.

Como o quadro em que se insere a elaboração de um vocabulário fundamental é a seleção lexical, estudamos os principais modelos que existem e o modo no qual esta que se costuma levar a cabo na prática. Centramo-nos, assim, nos métodos objetivos de escolha léxica, já que procuramos sempre critérios científicos e eficazes. Conseguimo-lo graças a procedimentos estatísticos e matemáticos.

Para determinar o número de palavras que são recomendáveis para cada nível do *Quadro Europeu Comum de Referência para as Línguas (QECR)*, recorreremos a estudos sobre percentagens de cobertura lexical das palavras frequentes, os resultados sobre a compreensão de textos em estudantes de inglês como língua estrangeira e a hipótese sobre a existência de um limiar lexical necessário para que se dê a compreensão oral e escrita de forma ótima. Com isso estabelecemos um intervalo de unidades que se devem ensinar em cada nível de língua.

Por último, determinamos o número de palavras que deve ter o vocabulário fundamental (orientado a um nível B1 do *QECR*) e os parâmetros para seleccionar os termos mais usuais e os mais disponíveis, que obtivemos de três obras previamente definidas: uma listagem com as 5.000 palavras mais frequentes e duas listagens de disponibilidade lexical de diferentes características. O resultado é o vocabulário fundamental do espanhol, ao qual se acrescenta a definição do seu conceito e a história das suas origens e anteriores propostas.

Introducción

Una de las cuestiones que debe resolver quien crea materiales y enseña un idioma es qué unidades léxicas seleccionar para cada nivel. Cuanta más centralidad se ha ido concediendo al vocabulario en el aprendizaje de una lengua, mayor importancia ha cobrado este tema. Así lo atestiguan las nuevas aproximaciones léxicas como el *lexical approach* (Lewis 1993: 90) o los estudios sobre el lexicón mental (§ 1.1.) y la competencia léxica (§ 1.2.).

Al mismo tiempo, quien haya afrontado alguna vez la creación de materiales didácticos ha comprobado la falta de estudios que nos aporten suficientes datos objetivos y empíricos sobre las unidades léxicas que debemos incluir. Sin duda, documentos como el *Marco común europeo de referencia para las lenguas* y el *Plan curricular del Instituto Cervantes* son de consulta obligada y resultan de gran ayuda para guiar este trabajo en una fase inicial. Sin embargo, cuando se profundiza en la cuestión, nos encontramos igualmente ante muchas dudas e interrogantes, ya que estos textos solo apuntan direcciones, pero no completan caminos. Quien pretenda hacer este recorrido hasta el final, debe buscar, pues, otros recursos. La gran mayoría de los que existen para el español, que son muy pocos en comparación con los realizados para otras lenguas, como el inglés, no explicitan la metodología que han seguido para su elaboración ni tampoco manifiestan que hayan seguido métodos científicos. Muchas investigaciones han avalado que la falta de rigurosidad en el modo de proceder de los creadores de manuales de español resulta en contenidos poco adecuados al nivel educativo y en incongruencias en el vocabulario seleccionado (§ 3.1.).

Por otro lado, somos conscientes de que el estudio del léxico puede resultar mucho más complejo que las cuestiones relacionadas con la sintaxis o la

gramática en general, debido a sus peculiaridades (§ 2.3.), ya que su carácter abierto, variable e ilimitado hace que sea difícil abarcarlo en su totalidad, como también dificulta la selección del más útil o rentable (§ 3.2. y 3.3.). Además de la existencia de múltiples y variadas unidades léxicas, sabemos que para llegar a aprenderlas no es suficiente con la mera presentación, sino que la repetición y la práctica son imprescindibles para conseguir memorizarlas y alcanzar un dominio tal que nos permita usarlas y combinarlas correctamente (§ 1.3. y 2.2.).

Por todos estos motivos, en el Capítulo 3 de nuestra investigación nos interesamos por los diferentes métodos de selección léxica y nos centramos en los frutos que pueden darnos aquellos que demuestran objetividad en su modo de proceder (§ 3.3. y 3.4.). Este es el marco general de nuestro trabajo, aunque el objetivo principal es encontrar la mejor metodología para aplicarla a la elaboración de un *vocabulario fundamental*, pues partimos de la hipótesis de que este refleja de forma más fiel y completa la competencia léxica de los hablantes. Si bien desde los estudios de disponibilidad léxica en español se ha hablado ya sobre las diferentes unidades léxicas de las que se nutre el vocabulario fundamental, en nuestros días carecemos de aportaciones que profundicen en su búsqueda, en la actualización de los procedimientos empleado y en su aplicación a la enseñanza de una lengua extranjera. Por eso, por una parte hemos buscado los orígenes del concepto y los vocabularios fundamentales que se han hecho en diferentes lenguas, como el francés, el alemán, el portugués, el español, etc. (Capítulo 7) y, por otro lado, hemos estudiado en profundidad las dos líneas de investigación de las que debemos nutrirnos para crearlo: la lingüística del corpus (Capítulo 5) y la disponibilidad léxica (Capítulo 6).

En los Capítulos 8 y 9 hacemos las aportaciones centrales para responder a nuestros objetivos: ofrecemos una definición sobre lo que entendemos por vocabulario fundamental (§ 8.1.), seguimos una metodología y criterios de selección de las palabras frecuentes y disponibles (§ 8.3.2.) y, finalmente, presentamos un proyecto provisional de lo que sería nuestro *vocabulario fundamental*, con las unidades léxicas que lo componen organizadas alfabética (§ 9.2.) y temáticamente (§ 9.3. y 9.4.). Dos pilares centrales que sustentan este proyecto son: conseguir delimitar el número de palabras que incorporar al

vocabulario y establecer los mejores procedimientos para escogerlas, siguiendo criterios objetivos y científicos en ambos casos. En el Capítulo 4 tratamos la primera cuestión, para lo cual recurrimos a los datos sobre cobertura léxica textual (§ 4.3.) y a la teoría sobre la existencia de un umbral léxico de comprensión lectora (§ 4.4.). Después de analizar diferentes estudios sobre estos temas, ponemos en relación el número de palabras, la cobertura léxica que nos proporcionan y el umbral de lectura para los diferentes niveles del *Marco común europeo de referencias para lenguas (MCERL)*. Las conclusiones que extraemos las exponemos en los últimos epígrafes del Capítulo 4, que culmina con la elaboración de una tabla que ilustra, a modo de resumen, el número de términos necesario para los diferentes niveles del *Marco* y, por consiguiente, para el vocabulario fundamental del español (§ 4.5.2.). Estos datos han sido los que han orientado nuestra selección del léxico frecuente y los que han ayudado a definir cuál sería el volumen total de nuestro vocabulario fundamental: entre 3.000 y 5.000 unidades, tomando como meta el nivel B1 del *Marco*.

Para delimitar el léxico disponible buscamos un procedimiento que, además de eficaz, se rigiera por criterios objetivos y empíricos. Si sobre frecuencia léxica y comprensión textual hay estudios abundantes para la lengua inglesa, sobre la selección del léxico disponible no se encuentran apenas aportaciones, y las que hay son parciales y/o poco explícitas. A pesar de la dificultad que supone este vacío, encontramos las respuestas de la mano de métodos estadísticos y matemáticos (§ 8.3.3.2.). Con ellos comprendimos que debíamos analizar el conjunto de unidades de cada centro de interés mediante gráficos (histogramas), con el fin de encontrar cortes naturales. De este modo, la elaboración de histogramas de frecuencia nos ofrece una radiografía mucho más clara sobre el comportamiento de los centros de interés que su mera disposición en una lista de palabras. Por otra parte, para los límites poco claros en los cortes que debíamos establecer, hemos hallado respuestas en los modelos matemáticos de la lógica difusa, que brindan la oportunidad de operar de un modo objetivo, pero sin caer en el estatismo o la rigidez.

Aunque el objetivo principal de esta investigación no es alcanzar un vocabulario fundamental del español como un producto final y acabado, sino

descubrir la metodología para hallarlo y probar su eficacia, en los Anexos 1, 2 y 3 se pueden consultar los listados resultantes de este trabajo, cuyo análisis realizamos en el Capítulo 9.

Finalmente, cabe indicar que en este estudio se pueden encontrar capítulos de diversa índole. Así, por ejemplo, las cuestiones tratadas en los tres primeros tienen como finalidad conseguir una visión multifacética y plural de la cuestión principal. Con esto intentamos obtener una panorámica global, sin que eso suponga alejarnos de los temas centrales que nos ocupan, sino complementarlos con nuevas soluciones, como puede ser el uso de gráficos para secuenciar el vocabulario disponible o los análisis matemáticos de la teoría de los conjuntos difusos. De este modo, en los primeros apartados nos centramos en estudiar dos aspectos importantes a la hora de hablar de la selección léxica: la complejidad del léxico y la importancia del vocabulario en el aprendizaje de una lengua. Esto se debe a que no queremos entender el vocabulario fundamental como un mero resultado de dos campos de investigación y acometer la suma de unidades sin más. Todo lo contrario, intentamos conjugar la experiencia que ofrece el trabajo práctico docente con el conocimiento teórico sobre los avances en el campo del español como lengua extranjera (ELE) y las teorías sobre la enseñanza-aprendizaje de lenguas. Creemos que esta óptica más integral ayuda a determinar mejor cuáles son las necesidades concretas existentes en cuanto al desarrollo de la competencia léxica y a la demarcación de los contenidos léxicos. Así podemos tener una visión más amplia, pero a la vez analítica y crítica.

También por este motivo, no nos hemos centrado exclusivamente en las corrientes de la lingüística del corpus y de los listados de frecuencias, ni tampoco en la línea de trabajo de la disponibilidad léxica. Es verdad que hemos profundizado en estos campos (Capítulos 5 y 6), para conocer su evolución, sus contribuciones, sus resultados y su metodología, pero también hemos querido salir de ellos, estudiar nuevos caminos y, finalmente, retornar a nuestro objetivo con otras perspectivas.

Capítulo 1. La competencia léxica

En este trabajo estudiamos los métodos de selección léxica, el conocimiento del lexicón mental de los hablantes y el desarrollo de la competencia léxica en estudiantes de una segunda lengua (L2) o de una lengua extranjera (LE). Pero ¿a qué nos referimos en concreto con estos conceptos? Cuando hablamos de *léxico*, *vocabulario* o *palabra*, ¿qué queremos decir? Este primer capítulo está dedicado a delimitar los conceptos básicos a nivel teórico y a explicar la terminología que empleamos de manera recurrente a lo largo de la tesis. La noción de competencia léxica es aglutinadora del resto de conceptos, de manera que es la elegida para dar título al capítulo. En los seis apartados que siguen exponemos cómo se explican las nociones básicas relacionadas con la enseñanza y el aprendizaje del léxico en la bibliografía especializada.

1.1 . El lexicón mental

Aunque se habla de *lexicón mental* como un concepto interdisciplinar que puede estudiarse en diferentes campos, como el de la Psicolingüística o el de la Lingüística computacional, si nos centramos en el ámbito de la Lingüística, se define como “el conocimiento que un hablante tiene del vocabulario” (Martín Peris *et al.* 2008, s. v. *lexicón mental*). En este sentido, el concepto se considera estrechamente relacionado con el de *competencia léxica*, que se concibe como “la capacidad de entender y utilizar unidades léxicas”, así como morfemas que permiten al hablante interpretar o generar unidades no percibidas ni producidas anteriormente, pudiendo combinarlas con otras (Martín Peris *et al.* 2008, s. v. *lexicón mental*).

Según Luque Durán, el lexicón mental es “una realidad mental paralela a otra realidad física que conocemos como mundo”, en la que distingue dos

componentes: unidades, elementos que han cristalizado como realidades peculiares, únicas y discretas; y redes, todas las relaciones que mantienen entre sí las unidades: sustitución, oposición, parecido, redes semánticas, semiotización, etc. (Luque Durán 1998: 124-125).

Baralo nos presenta otra definición de lexicón:

Llamamos 'lexicón' a la parte de la competencia lingüística que contiene las piezas léxicas formantes, es decir, las raíces y temas, los afijos flexivos y derivativos y las reglas que regulan su combinación. El concepto es cognitivo, dinámico y procesual, a diferencia del concepto de "léxico", como sinónimo de "vocabulario", entendido como el simple listado de palabras, o como una organización de campos semánticos. El lexicón es un constructo que da cuenta de la capacidad creativa del lenguaje, y por lo tanto de las lenguas. Permite comprender y explicar los fenómenos de generación de nuevas palabras, que pueden ser entendidas, procesadas y recreadas de manera novedosa, sin información explícita, por cualquier hablante nativo e, inclusive, por un hablante no nativo de la lengua. En este sentido, el adjetivo "mental" puede ser redundante, pero mantenemos la calificación con el propósito de recalcar su carácter cognitivo. [Baralo 2001: 26]

Se suele comparar, de forma más práctica, con una base de datos mental donde se encuentran almacenadas las unidades léxicas, interrelacionadas de muy diferentes maneras. Es decir, en este almacén residen los conocimientos que integran la competencia léxica, guardamos las representaciones mentales de las palabras, junto con las unidades léxicas correspondientes y su información semántica o conceptual y sintáctica (Aitchison 1987). Frente a la memoria semántica, que funciona como un almacén de conceptos a corto plazo y con una capacidad bastante limitada, el lexicón mental se ocupa de conservar de modo más permanente y estructurado las formas de las palabras.

Algunas de las características que definen este inventario léxico mental, y que se aplican tanto a hablantes nativos como a no nativos, son su fluidez y dinamismo, ya que se puede añadir nueva información, modificar la existente o establecer relaciones distintas. Así, se puede reorganizar y reestructurar continuamente, ampliando sus elementos y las relaciones establecidas entre ellos de modo constante mediante el aprendizaje lingüístico y cultural. Se ha estudiado la gran capacidad de los niños para aprender palabras nuevas y ampliar el lexicón mental, algo que se relaciona con la teoría lingüística de

campos léxicos, pues en ella se basa la organización lingüística del primero, de forma que en el proceso de aprendizaje, el niño asigna la palabra nueva a una categoría semántica en un primer momento, y luego aprende a distinguirla de las demás palabras de la misma categoría. Gracias a que esta base de datos se encuentra organizada, podemos reconocer y usar sus unidades rápidamente. Aitchison (1987) propone, para explicar el proceso de producción léxica, un primer momento, en el que tras la activación de un campo léxico entero se escoge entre una serie de alternativas contextualmente apropiadas, para en un segundo momento activarse el componente fonológico, comparando las palabras que contienen sonidos parecidos, de modo que se consideran solo las que fonológica y semánticamente son apropiadas y se desechan las otras progresivamente, hasta que queda sólo una palabra, y esta es la que usa el hablante.

En cuanto a la organización del lexicón mental, la información que se guarda de cada elemento es de diverso tipo (fónica, gráfica, morfológica, semántica, discursiva, etc.), así como las relaciones y asociaciones que se establecen entre sus unidades. Se describen estos procesos mediante el modelo de redes semánticas (Quillian 1968; Aitchison 1987: 72-85), que explica que la información se almacena en el cerebro formando redes que ayudan a organizar su estructura de modo económico, es decir, recurriendo a tipos básicos de relaciones semánticas: como las de hiperonimia y meronimia, sinonimia y antonimia, entre otras. Se trata de un sistema asociativo en el que las relaciones entre las unidades léxicas son muy diversas y complejas, al contrario de lo que se pueda pensar, esto permite recuperar las palabras con gran rapidez. Así, más que como un diccionario, tenemos que ver el lexicón como una red, o una red de redes, con relaciones de distintos tipos, que se van ampliando a lo largo del tiempo, gracias a la exposición a *input*.

Lahuerta y Pujol (1996) describen, por su parte, cuatro características principales del lexicón: 1) Es una red con múltiples asociaciones —fonéticas, gramaticales, semánticas, enciclopédicas ... con diferentes tipos de relaciones a su vez; por ejemplo, dentro de las semánticas hay: a) relaciones de coordinación (*arriba / abajo, sal y pimienta*); b) relaciones de combinación (*salado /mar: vivo/rojo*); c) de hiperonimia, d) de sinonimia, etc.—. 2) La cantidad de

asociaciones de las entradas léxicas es variable, cada unidad establece una serie de relaciones que no tiene por qué coincidir con la cantidad y naturaleza de otras. 3) Muchas de las asociaciones son personales y otras están ya fijadas en la lengua (*blanco y negro, marido y mujer, cada dos por tres*). Y 4) la relación entre las palabras es más importante que su localización absoluta.

No cabe duda de que conocer el lexicón mental y su funcionamiento resulta muy útil para la selección y la enseñanza del vocabulario. Estos conocimientos deben tenerse en cuenta en la enseñanza de español, con el fomento de las asociaciones entre palabras, como se observa en los métodos de enseñanza actuales con los mapas de ideas, redes, gradaciones, pares, etc. De esta forma el nuevo conocimiento es más fácilmente absorbido cuando es asimilable a algo ya conocido y cuando se ha activado el marco de conocimiento adecuado en el alumnado. Estos marcos de conocimiento o *scripts* son un conjunto de estructuras cognitivas que se basan en nuestras experiencias pasadas, filtrando y dando forma a las percepciones; su función principal es ayudar a procesar, organizar y comprender la información nueva (Martín Peris *et al.* 2008, s. v. *marcos de conocimiento*). Un ejemplo de cómo responden alumnos de ELE a la pregunta de qué tipo de asociaciones vienen a su cabeza para la palabra *coche* es la siguiente imagen (Higuera García 2004a: 14):

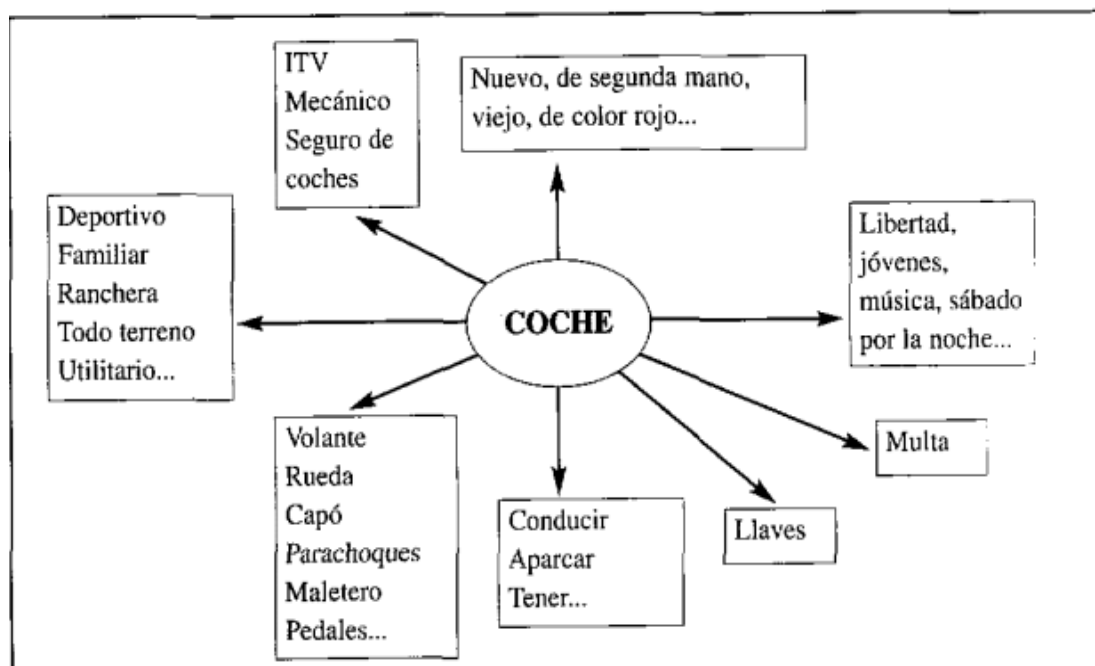


IMAGEN 1. Asociaciones de la palabra *coche* (Higuera García 2004a: 14)

En consecuencia, parece lógico que el profesorado deba presentar y practicar las palabras y unidades léxicas mediante técnicas y contextos que permitan relacionar unas con otras, tal y como se almacenan en el lexicón mental, con el objetivo de mejorar y agilizar su aprendizaje.

Además, al explicar el lexicón mental en relación con una segunda lengua, surgen varias dudas. En un primer lugar nos preguntamos cómo actúa la lengua materna (LM) en relación con la L2 u otras lenguas, y si existe una influencia entre los diferentes elementos o sobre la organización de los mismos. La respuesta parece afirmativa, ya que una de las formas de verificar la organización mental del léxico y su influencia sobre la L2 son las pruebas de disponibilidad léxica, en las que se aprecian los elementos que ofrecen los individuos, junto con su forma y el orden en que se activan. Por otro lado, tenemos que explicar qué relación tiene el lexicón mental con la lengua materna y una L2, para lo cual existen varias hipótesis: desde las que hablan de la existencia de un solo lexicón para cada lengua, la materna y la L2, hasta las que optan por un sistema de almacenamiento compartido. Se suele hablar de la hipótesis del sistema extendido, que entiende que las palabras de la lengua materna y la L2 se almacenan en un único lugar; la hipótesis del sistema dual, que explica esta cuestión con dos espacios distintos; la tripartita, que apuesta por un mismo espacio para palabras similares, mientras que las más divergentes estarían en lugares separados; y, por último, la de subconjuntos, que indica la existencia de un mismo lexicón que está formado por dos subconjuntos (Pérez Basanta 1999). Weinreich (1953) había planteado una clasificación en tres modelos de almacenamiento de las unidades léxicas: modelo coordinado (dos lexicones), modelo compuesto (dos lexicones diferentes, pero comparten un mismo sistema conceptual) y modelo subordinado (un único sistema conceptual y dos lexicones, pero la vía de información es a través de la primera lengua). Posteriormente se añadió a la clasificación de Weinreich el modelo de la representación mixta, según el cual en ocasiones se accede a la información de la L2 a través de la L1, pero también se da el caso de que se recupera directamente la información sin el paso intermedio de la L1.

1.2. El concepto de competencia léxica

En la enseñanza de segundas lenguas, la investigación sobre los procesos de aprendizaje y desarrollo del vocabulario ha pasado a ocupar una posición privilegiada, por lo que al concepto de competencia léxica se le ha prestado mayor atención en los últimos años. Conocer cómo se adquiere el vocabulario en una L2, distinguir entre adquisición y aprendizaje o definir qué significa conocer una palabra son algunos de los temas a los que tratan de responder las publicaciones que abordan este concepto. La noción de competencia se usa en el ámbito de la lingüística aplicada y de la enseñanza de lenguas extranjeras como una habilidad (saber hacer), no solo como un conocimiento (saber). Fue Hymes quien acuñó el término de *competencia comunicativa* por primera vez en el área de la etnografía y pronto pasó a ser imprescindible en la sociolingüística y la pragmática. En contraste con la idea de competencia que defendía Chomsky desde la teoría generativa-transformacional, para Hymes la competencia comunicativa consiste en la capacidad de usar la lengua de forma adecuada en cada situación concreta de comunicación. Según el modelo de Canale (1995), esta capacidad comprende varias subcompetencias: lingüística, sociolingüística, discursiva y estratégica. Así pues, dentro de la subcompetencia lingüística es donde se sitúan el conocimiento léxico, así como el de la ortografía, fonología, morfología, semántica y sintaxis.

Para otros autores, como Lahuerta y Pujol (1996: 121), es “la capacidad para relacionar formas con significados y utilizarlas adecuadamente se denomina competencia léxica, y al lugar donde reside el conocimiento de esa competencia se le denomina lexicón mental”. Para Cassany *et al.* (1998: 379) “El dominio del vocabulario consiste tanto en el conocimiento de las palabras y de los conceptos a los que se refieren como en las diversas estrategias para usarlos con eficacia y adecuación”. Y el *Marco común europeo de referencia para las lenguas (MCERL)*, Consejo de Europa 2001: 126-131) la define como “el conocimiento del vocabulario de una lengua y la capacidad para utilizarlo”; a continuación, en el *MCERL* se señala que “la competencia semántica comprende la conciencia y el control de la organización del significado que posee el alumno”.

En el *MCERL* encontramos también la distinción entre *competencia léxica* y *semántica* (la primera es el conocimiento del vocabulario y la capacidad de usarlo, mientras que la segunda es la conciencia y control de la organización del significado que posee el alumno), mientras que Marconi (2000: 79-99) habla de *competencia inferencial* (conocimiento de la red de conexiones, que atañen a destrezas como la inferencia semántica, la paráfrasis, la definición y el hallazgo de sinónimos) y *competencia referencial* (que tiene que ver con la capacidad de nombrar un objeto y aplicarle una palabra, es decir, la capacidad para activar las palabras en el mundo real).

Aplicado a la docencia de una LE y de forma más global, sirvan las palabras de Cassany *et al.*:

Resumiendo, la competencia léxica va más allá de las listas exhaustivas y de la competencia gramatical. Además de conseguir y ampliar el bagaje léxico necesario (activo y pasivo), los alumnos tienen que adquirir criterios de uso para cada palabra en relación con el registro, la precisión, las connotaciones, etc. de la situación comunicativa en la que se quiera emplear. [Cassany *et al.* 1998: 382]

No obstante, la cuestión no resulta tan simple. La naturaleza multifacética de la competencia léxica y de sus unidades nos impide dar una definición sencilla. De hecho, al explorar los distintos modelos existentes, parece que su definición debe sustentarse en diferentes dimensiones. En primer lugar, encontramos diversas dimensiones lingüísticas (fonológica, morfológica, sintáctica y semántica) y en un segundo momento otras dimensiones pragmáticas (discursiva, referencial, sociocultural y estratégica), que también deben abordarse en el estudio del léxico en el aula de español LE/2. Este cambio de paradigma se dio en la enseñanza de las LE a partir de los años 70, sobre todo de la mano del enfoque comunicativo. La enseñanza del vocabulario de una LE reside, por lo tanto, en desarrollar al mismo tiempo el dominio de todos estos componentes, que están relacionados con los distintos aspectos que implica conocer una palabra. Así lo defienden autores como Tréville y Duquette (1996), que adaptan la descripción de la competencia comunicativa de Canale (1995) a la competencia léxica. Retomamos la explicación de Izquierdo Gil (2004), quien trata el tema en su tesis doctoral. Para esta autora, ampliar la competencia léxica del alumnado de ELE supone atender no solo a la dimensión lingüística, sino también a los saberes que engloba la pragmática. Más concretamente, al enseñar

vocabulario atendiendo a su contenido lingüístico trabajamos conocimientos como la forma oral de las unidades, su forma escrita, su significado, su morfología, su sintaxis y los contextos de uso. Sin embargo, hay otros aspectos que no engloba este conocimiento. Para complementarlos, debemos recurrir a saberes relacionados a) con la competencia discursiva, que nos ayuda a dar coherencia y cohesión a un texto usando correctamente los marcadores discursivos, los pronombres, artículos y otros elementos catafóricos y anafóricos, y nos permite interpretar y construir los diferentes tipos de textos; b) con la competencia referencial, que tiene que ver con el conocimiento extralingüístico y atiende al desarrollo de dos líneas: los campos semánticos (que se asocian a un tema concreto) y las cadenas léxicas (que se asocian a determinados esquemas de comportamiento social o *scripts*); c) con la competencia sociolingüística o sociocultural, que desarrolla las habilidades para usar la lengua en su aspecto social, ajustarse a los diferentes registros, y manejar las fórmulas convencionales, y, por último, d) con la competencia estratégica, que desarrolla capacidades responsables de resolver las dificultades o deficiencias de conocimiento que se puedan manifestar en los actos comunicativos, activando procedimientos como las inferencias, las paráfrasis u otras técnicas de aprendizaje o reglas nemotécnicas.

En relación con estas habilidades estratégicas está lo que se suele llamar *vocabulario instrumental*, también conocido como “*verba omnibus*”, “palabras comodín” o “palabras herramienta”. Izquierdo Gil (2004: 101) usa la primera expresión y lo define como “vocabulario compuesto fundamentalmente por unidades léxicas de significado muy general, que permiten explicar otras palabras más específicas, sustituirlas, parafrasearlas, definir las, estructurarlas y organizar su significado”. Se trata de palabras o expresiones de significado no específico que son de utilidad para expresar términos o conceptos que el alumno desconoce o que no acuden a la boca de hablante nativo en determinadas circunstancias. Hablamos de palabras como “lugar, persona, cosa, instrumento, etc.”, de gran utilidad para suplir las carencias que los estudiantes de español como lengua extranjera (ELE) puedan tener y, por lo tanto, recomendables para ser incluidas en la planificación de la enseñanza del léxico a nivel inicial.

Al preguntarnos por los elementos que integran la competencia léxica, podemos pensar en las respuestas que ofrecen algunos autores de léxicos frecuentes, que recurren a imágenes de diferentes círculos concéntricos para explicarlos. Así, por ejemplo, Ávila Muñoz (1999) y Terrádez Gurrea (2001) acompañan sus explicaciones con un modelo de tres círculos para ilustrar los diferentes tipos de vocabulario: un primer núcleo compuesto por el léxico funcional (las unidades básicas y usuales que aparecen en todo intercambio comunicativo, como son los determinantes, pronombres, preposiciones, conjunciones, verbos copulativos, verbos auxiliares, etc.), un segundo grupo con las expresiones usadas en situaciones comunicativas ordinarias (elementos léxicos que poseen un carácter polifuncional: familia, viajes, comidas, relatos, ocio, etc.), mientras que el último círculo está compuesto por los vocablos producidos en circunstancias particulares (áreas temáticas específicas, lenguas especiales de grupo, términos, etc.). Aplicado al ámbito del ELE, las unidades centrales y del segundo círculo serían objeto de estudio en un curso de lengua general, mientras que las terceras se estudiarían en un curso de español de los negocios (ENE) o español para fines específicos (EpFE) o bien en un nivel avanzado de lengua. Por nuestra parte, nos centraremos sobre todo en las unidades del núcleo.

1.3. ¿Qué significa conocer o aprender una palabra?

Como hemos explicado, la información léxica puede estar almacenada a corto plazo en la memoria o a largo plazo en el lexicón mental, por lo que uno de los objetivos del aprendizaje del vocabulario en ELE es intentar el trasvase de información desde la primera hacia el segundo. La memoria la usamos en el proceso de manipulación del lenguaje, mientras que se supone que el lexicón interioriza las formas léxicas ya de modo permanente. Pero ¿a qué nos referimos cuando hablamos de aprender una palabra?

En el caso del inglés como lengua extranjera, Harmer (1991: 158) esquematiza lo que significa aprender una nueva palabra de la siguiente forma:

WORDS	MEANING	Meaning in context Sense relations
	WORD USES	Metaphor and idiom Collocation Style and register
	WORD INFORMATION	Parts of speech Prefixes and suffices Spelling and pronunciation
	WORD GRAMMAR	Nouns: countable and uncountable, etc. Verb complementation, phrasal verbs, etc Adjectives and adverbs:

CUADRO 1. Aspectos que entran en juego al conocer nuevas palabras (Harmer 1991: 158)

Dentro del ámbito hispánico, Cassany *et al.* (1998: 380) apuntan que conocer una palabra implica tener conocimientos sobre: su pronunciación y ortografía (saber los sonidos y letras que la forman, saber reconocerla, pronunciarla y escribirla), sobre su morfología (conocer su composición, flexión y las palabras con las que puede combinarse), su sintaxis (conocer su categoría y las reglas sintácticas que rigen su uso), su semántica (saber su valor denotativo y connotativo, así como sus diferentes acepciones según el contexto y ser capaz de relacionarla semánticamente con otras unidades léxicas), su pragmática (usarla como parte de un texto en relación a un contexto y utilizarla con un propósito determinado) y sociolingüística (conocer su valor dialectal y su registro, además de usarla de forma adecuada a la situación comunicativa).

Algunas respuestas a lo que supone el conocimiento de las unidades léxicas vienen también de la mano de Izquierdo Gil (2004: 46-8), para quien los aspectos que somos capaces de resolver cuando aprendemos una unidad léxica son los siguientes:

- a) reconocer la unidad léxica cuando se oye y saber pronunciarla,
- b) reconocer la forma escrita de la unidad léxica y ser capaz de escribirla,
- c) reconocer la morfología de la unidad léxica, es decir, los morfemas que la forman, relacionar dichas partes con su significado, así como ser capaz de formar la unidad léxica utilizando los morfemas correctos,
- d) reconocer las diferentes acepciones o significados y ser capaz de producir la unidad léxica para expresar su significado según el contexto,
- e) reconocer su categoría gramatical,
- f) conocer las estructuras sintácticas en las que puede aparecer y sus

- restricciones,
- g) reconocer y ser capaz de producir otras unidades con las que se relacione desde el punto de vista del significado (sinónimos, antónimos, cohipónimos, hiperónimo, etc.) (relaciones paradigmáticas),
 - h) reconocer y ser capaz de producir las unidades léxicas con otras unidades con las cuales típicamente suele combinarse (“colocaciones”, relaciones sintagmáticas),
 - i) conocer la adecuación pragmática de una unidad léxica a la situación o contexto comunicativo (según el lugar, el interlocutor, la intención, etc.),
 - j) conocer su frecuencia de uso,
 - k) conocer a qué registro pertenece y utilizarla en una situación adecuada,
 - l) saber qué información cultural transmite para una comunidad lingüística,
 - ll) saber si pertenece a alguna expresión idiomática o institucionalizada,
 - m) reconocer y saber qué unidades están restringidas al discurso oral o escrito,
 - n) conocer sus equivalentes en otras lenguas.

Higueras García (2009: 114-7) responde a esta cuestión con once aspectos que implican el conocimiento de una unidad léxica para un nativo:

1. La denotación y referencia.
2. El sonido o grafía, según el canal.
3. Las estructuras sintácticas en las que aparece.
4. Las peculiaridades morfológicas.
5. Las relaciones paradigmáticas con las unidades que podrían haber aparecido en su lugar.
6. Las combinaciones sintagmáticas o palabras con las que normalmente se asocia.
7. El registro y si tiene más probabilidades de aparecer en la lengua hablada o escrita o ambas.
8. El contenido cultural.
9. Los usos metafóricos.
10. La frecuencia de uso.
11. La pertenencia a expresiones institucionalizadas.

Y, siguiendo a Carter (1987), añade dos más, que se activan en el caso concreto de una L2:

12. Saber usarla productivamente, como consecuencia de la suma de toda esa información, aunque para algunos fines sea necesario sólo entenderla.
13. Conocer las relaciones que tiene con unidades de su lengua materna.

A la vista del número de ítems que sirven para dar respuesta, en los acercamientos mencionados, a la cuestión inicial, parece claro que el proceso de conocimiento de las unidades léxicas es complejo, por lo que debe ser gradual. De esta forma, nos podemos centrar en unos aspectos que interesen más en un nivel inicial y podemos dejar otros para ir desarrollando progresivamente, ya que el conocimiento sobre las unidades léxicas no es estático, sino que se va aumentando y sumando a los conocimientos previos. El docente puede

seleccionar aquellos aspectos que considere más importantes para introducir antes o después en función de cada caso y muy probablemente el alumno interiorizará los diferentes aspectos a largo plazo, después de varias exposiciones a cada unidad, observar diferentes contextos y realizar diferentes tareas y prácticas. Como vamos a ver más abajo, la defensa de la enseñanza cualitativa del vocabulario es uno de los principios del llamado “enfoque léxico”.

Otro aspecto a tener en cuenta, a propósito de la enseñanza del vocabulario, es que no solo tenemos que enseñar nuevas palabras, sino también ampliar el conocimiento que tenemos sobre ellas, pues conocer una unidad léxica implica poseer mucha información de distinta naturaleza sobre ella. A nivel práctico, cuando el profesorado introduce una palabra o quiere repasarla, como *trabajo*, aunque sea en un nivel inicial, se pueden explicar también sus colocaciones, es decir, los verbos con los que suele coaparecer: *encontrar, buscar, perder, tener, estar en el, ir al, volver del...*, así como los adjetivos que pueden calificar a este nombre: *intelectual, físico, interesante, mecánico*, etc. Conviene comentar el registro, los contextos de uso en los que puede o debe usarse, ciertas fórmulas o expresiones institucionalizadas, así como el significado contextual y cotextual. Higuera García (2004a: 11) ejemplifica este tipo de enseñanza con un cuadro dedicado a los matices que podemos tratar para la palabra *majo*:

	Persona	Sueldo	Padres	Sitio	Piso
MAJO-A	Agradable, simpático, amable...	Alto, es un trabajo bien remunerado...	Comprensivos, amables, tolerantes	Cómodo, acogedor...	Grande, coqueto, bonito...

CUADRO 2. Matices que se pueden enseñar de la palabra *majo-a* según Higuera García (2004a:11)

1.4. Vocabulario receptivo y vocabulario productivo

Si bien deben considerarse dentro de un *continuum* y no como dos componentes discretos y separados, cuando se hace referencia al lexicón mental se suele distinguir entre vocabulario receptivo —o pasivo— y vocabulario productivo —o activo—. La distancia entre uno y otro debe entenderse como grados de conocimiento o de familiaridad, no como polos independientes y

opuestos. Ambos están relacionados directamente con las habilidades receptivas y productivas y forman parte del lexicón mental, aunque de forma complementaria, ya que el primero corresponde a las palabras que un individuo es capaz de interpretar en diferentes situaciones de recepción de mensajes. No solo engloba las palabras almacenadas en la memoria, sino también las que podría comprender si se presenta la ocasión gracias a las relaciones con otras palabras o a la capacidad de asociación de ideas e inferencias (Cassany *et al.* 1998: 381). Al contrario, el productivo es aquel que el hablante efectivamente emplea al hablar y escribir en su vida cotidiana. Consecuentemente, el vocabulario pasivo es mucho más amplio que el activo, puesto que incluye toda palabra o expresión que entendemos, pero que no solemos emplear, desde léxico dialectal de otras variedades, hasta palabras que designan realidades lejanas, terminología específica, etc.

Aunque no existe consenso sobre el tamaño de esta diferencia, sí sabemos que el proceso de comprensión es anterior al de producción, que se necesita menos información para interpretar una palabra que para generarla y que este orden influye en el proceso de adquisición del vocabulario, de tal manera que, independientemente de la distancia que existe entre la cantidad de un tipo de vocabulario y el otro, en la didáctica de una L2 o de una lengua extranjera (LE) debemos contribuir a pasar del conocimiento receptivo al productivo. Ambos están muy relacionados con los diferentes grados de conocimiento de las unidades léxicas que hemos comentado en el apartado anterior, ya que el grado de control que debemos tener para interpretarlas es menor que el que necesitamos para producirlas. En la docencia conviene facilitar este camino para ir consolidando la información que posee el aprendiz sobre las palabras y para que la pueda añadir a los conocimientos previos. También debemos ser conscientes de que sólo las palabras que se usan pasan a formar parte del léxico activo.

1.5. Léxico y vocabulario

Si bien en algunas disciplinas lingüísticas se establece distinción entre ambos, en la didáctica de la lengua los términos *vocabulario* y *léxico* suelen usarse como equivalentes. Sin embargo, si queremos ser más específicos y meticulosos debemos definir qué entendemos en cada caso. De esta forma, sabemos que el léxico forma parte del sistema lingüístico y como tal es independiente del conocimiento particular de cada hablante, lo que lo distingue del vocabulario, que depende de cada individuo. Estos dos términos, léxico y vocabulario, se relacionan respectivamente con dos planos diferentes, el de la lengua en su sistema (*langue*) y el del uso de la lengua (*parole*). En consecuencia, el léxico hace referencia al sistema de palabras que componen un idioma, esto es, todo el caudal léxico de una lengua. Mientras que, cuando esas unidades virtuales de la lengua se actualizan en el discurso de un hablante concreto, reciben el nombre de *vocablos* y al conjunto de todos estos elementos lo llamamos *vocabulario*. Este es, pues, el conjunto de términos que el hablante actualiza en el discurso y que constituye el *vocabulario potencial* de un individuo. Por otra parte, el *léxico del español* es el conjunto de unidades léxicas de la lengua española.

En este trabajo usamos como sinónimos las palabras “léxico” y “vocabulario”, aunque, como acabamos de explicar, somos conscientes de que existen autores que entienden que el primero es más amplio y se refiere a todo el caudal léxico de una lengua, mientras que el segundo se refiere solo a aquellas palabras que los hablantes conocen y que actualizan en la práctica. En nuestro caso, la acepción que nos interesa es esta última.

Acerca del vocabulario suelen destacarse dos aspectos, que se reflejan también en su didáctica y que ya hemos comentado: las múltiples dimensiones de las palabras y el gran número de unidades que lo conforman, aunque estas sean menos que si nos referimos al léxico. Por lo que, en el caso de la enseñanza, es conveniente disponer de herramientas y criterios para su selección, por ahora aplicados muy escasamente al español. Estos aspectos marcan algunas de las principales directrices de su didáctica, como la que quizá constituya la cuestión fundamental: enseñanza explícita o implícita, pues se trata de responder a la

pregunta de si es posible enseñar de forma explícita (en qué medida y de qué modo) un componente del lenguaje con tantos elementos integrantes y tantas dimensiones para cada uno de ellos. Algunas respuestas las veremos a continuación.

1.6. Unidades léxicas y palabras

Otra cuestión epistemológica que conviene mencionar se refiere a los términos a los que se recurre habitualmente para nombrar a las unidades que forman el lexicón mental, concepto diferente a su vez, como hemos visto, tanto del de léxico, como del de vocabulario. Recurrentemente en estas líneas han aparecido las expresiones “unidad léxica”, “palabra”, “vocablo”, etc. para referirnos a ellas, aunque en función de los autores, se pueden encontrar también otras, como “categoría cognitiva”, “lexía”, “locución”, “palabra gráfica”, etc. En general las usamos como expresiones sinónimas, sin embargo, podemos también establecer distinciones en este sentido, por lo que dedicaremos algunas líneas a esta tarea.

Quizá el término de “palabra” sea el menos recomendable, por el motivo de que se puede identificar con “palabra gráfica”, mientras que la naturaleza de estas unidades es de diferente tipo. En este sentido, la expresión “unidad léxica” resulta eficaz por ser más amplia y menos específica, por lo que conseguimos así englobar bajo un mismo paraguas los diferentes tipos de unidades y sus diversos aspectos. Gómez Molina (2003, 2005a) trata la cuestión del siguiente modo:

Consideramos la unidad léxica (*lexical item* para Lewis 1993, 1997; *lexical unit* para Cruse 1986 y Bogaards 2001) como la unidad de significado en el lexicón mental, la cual sirve como elemento vehiculador de la cultura del español y puede estar formada por una palabra, simple o compuesta (*coche, agridulce, camposanto,...*) o por varias, denominada unidad léxica pluriverbal (*buenos días, encantado de conocerte, llover torrencialmente, odio mortal, hora punta, pedir la mano, por la boca muere el pez, ...*). Es evidente que entre los ejemplos citados no todos presentan un significado unitario, pero es la perspectiva pedagógica, no lingüística, la que nos permite considerarlos como unidades léxicas. [Gómez Molina 2005a: 85-6]

Lo que realmente nos interesa es identificar cuáles son estas unidades que conforman el lexicón mental y cómo funcionan. Sabemos que son imprescindibles para que los hablantes de cualquier lengua puedan categorizar y

organizar la información, descomponerla en segmentos léxicos, tanto simples como compuestos, procesarlas y, si no lo están ya, almacenarlas en el lexicón mental. También parece lógico que en el tratamiento del vocabulario que debe enseñarse en una L2 no se haga hincapié en la distinción convencional dentro de la lingüística entre el subsistema léxico (palabras) y el subsistema fraseológico (colocaciones, locuciones, etc.), ya que ambos deben ser objeto de estudio. Unido a ello está el cambio de perspectiva sobre la enseñanza-aprendizaje del vocabulario a aprendices de lenguas extranjeras que se ha dado en los últimos años y que tiene como lanzadera la teoría del “enfoque léxico” (*lexical approach*), acuñada por Lewis (1993: 90), a la que ya hemos aludido con anterioridad. Aunque no nos vamos a detener en ello, lo mencionamos para defender lo eficaz que resulta trabajar de manera sistemática los distintos tipos de unidades léxicas de forma explícita y planificada, algo que se dejó de valorar durante algunos años con el apogeo del enfoque comunicativo. Según el enfoque comunicativo, el uso de textos auténticos y la presentación de los ítems léxicos en contexto debería ser suficiente para el aprendizaje del léxico, por lo que no se veía necesario un tratamiento especial y explícito del vocabulario, adaptado a los diferentes niveles.

La dicotomía existente entre gramática y vocabulario y la prioridad de la primera frente al segundo parecen hoy en día superadas y sustituidas por la visión de un *continuum* entre ambos, como ha mostrado la lingüística cognitiva. De este modo, resulta mucho más eficaz un tratamiento sistemático e integrado de los diferentes elementos que componen la competencia léxica, con la finalidad de trasladar al aula los múltiples aspectos que supone el funcionamiento de las unidades léxicas de forma conjunta: semánticos, discursivos, gramaticales, relaciones paradigmáticas (de hiperonimia, sinonimia, antonimia, etc.) y sintagmáticas (combinaciones, colocaciones, expresiones fijas, etc.). Así, se justifica el creciente interés por conocer el repertorio de las unidades léxicas mediante diferentes métodos y se promueven las investigaciones que van encaminadas a ello. Y ¿qué tipo de unidades integran la competencia léxica teniendo en cuenta la enseñanza de una LE desde esta perspectiva? Desde luego, el contenido léxico abarca un amplio elenco de unidades en un *continuum* difícil

de segmentar, por lo que delimitar sus componentes no es tarea simple. Veamos a continuación algunas propuestas sobre la tipología de unidades léxicas.

El *MCERL* (Consejo de Europa 2001: 126-128) especifica que la competencia léxica está compuesta por dos tipos de elementos, los léxicos y los gramaticales, mientras que la competencia semántica tiene que ver con el significado de las unidades léxicas, tanto en el contexto general (referencia, connotación, nociones específicas) como en lo que afecta a las relaciones semánticas que se establecen entre ellas. Explica, además, que los elementos gramaticales forman el “vocabulario cerrado” y limitado (los artículos, cuantificadores, demostrativos, posesivos, pronombres personales, pronombres relativos, adverbios interrogativos, preposiciones, verbos auxiliares y conjunciones) y que los elementos léxicos se dividen en dos bloques:

1. El “vocabulario abierto”: sustantivos, verbos, adjetivos y adverbios, así como los diversos sentidos que pueden adquirir. También integra aquí los conjuntos léxicos cerrados (días de la semana, meses del año, pesos y medidas, etc.).

2. Las fórmulas fijas, que comprenden: fórmulas de interacción social y de cortesía, refranes, proverbios y arcaísmos; modismos (agrupan clichés, locuciones idiomáticas y metáforas lexicalizadas); estructuras fijas, aprendidas y utilizadas como conjuntos no analizados (entre ellas, las iniciadoras de frase); otras frases hechas (combinaciones formadas por palabras que se utilizan juntas frecuentemente); y combinaciones de régimen semántico o colocaciones.

Desde su enfoque, Lewis (1993, 1997a, 2000) propone una taxonomía que intenta organizar este conjunto ilimitado estableciendo las unidades que deben componer el *vocabulario mínimo adecuado*. En 1993 propuso diferentes clases de elementos léxicos (*lexical items*): palabras (*words*) y unidades léxicas complejas o multipalabras (*multi-word items*), palabras complejas (*polywords*, ej. *record player*), colocaciones (*collocations*, ej. *unemployment rise*), expresiones institucionalizadas (*institutionalized expressions*, ej. *just a moment, please*). En trabajo posterior agrupa en cuatro tipos: 1) *Words*, 2) *Collocations*, 3) *Fixed expressions* 4) *Semi-fixed expressions* (Lewis 1997a). Añade que, mientras las colocaciones ya deben trabajarse desde el nivel inicial, las expresiones se

estudian en los niveles intermedio y avanzado como una de las formas de desarrollar la fluidez.

Por su parte, McCarthy (1999: 238-247) entiende que el vocabulario básico o esencial, que él denomina *core vocabulary*, lo forman diez tipos de unidades léxicas, todas importantes por igual como componentes básicos de la comunicación: modalizadores, verbos de régimen, palabras interactivas, marcadores del discurso, sustantivos básicos, deícticos, adjetivos básicos, adverbios básicos, verbos básicos de acción y sucesos, y por último otras unidades léxicas y fraseológicas. Posteriormente (McCarthy 2001: 149-163), establece el contenido léxico para los niveles superiores (*advanced level vocabulary*) y señala como tareas: el desarrollo de redes asociativas (semánticas o psicológicas) que pueden establecer las palabras ya conocidas en los niveles inferiores; la ampliación de nuevos significados de un mismo vocablo (connotaciones en sus contextos típicos de uso); las expresiones fijas, especialmente, las semánticamente opacas; y el conocimiento fraseológico, principalmente, las colocaciones, las cuales considera, junto con las expresiones idiomáticas, como el componente principal de la enseñanza del léxico en este nivel. Destaca que estas unidades deben trabajarse usando el contexto, así como la necesidad de propiciar actividades sobre conocimientos y destrezas léxicos que permitan al alumno convertirse en un aprendiz autónomo de vocabulario. Este mismo autor (McCarthy 1984) comentaba la necesidad de saber usar las unidades léxicas no sólo desde el punto de vista sintáctico y semántico, sino también pragmático o discursivo. Es el uso del léxico a un nivel superior del de la frase, es decir, al nivel del discurso, lo que permite un verdadero desarrollo de la competencia comunicativa. Además, señala que años atrás las actividades para aprender y aumentar el léxico consistían en textos escritos, por lo que el único vocabulario que se incrementaba era el receptivo.

Del mismo modo, la distinción que describe Gómez Molina se aproxima mucho a las propuestas anteriores (2004a: 4-20). En su caso, propone como metodología avanzar del subsistema léxico al subsistema fraseológico siguiendo el esquema tripartito coseriano sobre grados de abstracción y, una vez presentadas las palabras, secuenciar las fórmulas rutinarias (habla), las colocaciones (norma) y las locuciones (sistema). Así, establece cinco tipos de

unidades, que van orientados directamente a los docentes de ELE, y que revisa un año más tarde fusionándolos en cuatro, aunque en realidad cambia apenas cuestiones terminológicas (2005b: 89-93).

- Para él, en un primer lugar también están las palabras simples (*lunes, leer, bien...*) y compuestas o compuestos léxicos (de diverso tipo, *termómetro, polideportivo, coche-cama, parachoques...*), a las que suma los elementos cerrados (series cerradas de palabras “gramaticales”, artículos, demostrativos, preposiciones, etc.) y las abreviaturas y siglas.
- En un segundo nivel sitúa las colocaciones léxicas y gramaticales (*esfuerzo ímprobo, concertar una cita, odio mortal; maúlla el gato, relinchar el caballo; confiar en, estar seguro de, deseo de...*) y los compuestos sintagmáticos (*lengua de trapo, lágrimas de cocodrilo, merienda cena*), que fusiona más tarde en lo que denomina combinaciones y compuestos sintagmáticos.
- En otro nivel está lo que denomina fórmulas rutinarias o fórmulas de interacción social (*Encantado de conocerte...*), que engloban también las palabras interactivas (*De acuerdo. ¡Un momento!*), los marcadores de relaciones sociales (*Usted. Muy señor mío. Atentamente...*) y las estructuras iniciadoras de frase (*¿Podría decirme ...?*).
- Por último están las locuciones idiomáticas, nominales (*abogado del diablo, mosquita muerta, coser y cantar*), verbales (*poner los cuernos, echar chispas, lavarse las manos, nadar y guardar la ropa*), adjetivas (*loco de remate, de chicha y nabo, de pelo en pecho, estar sin blanca*), adverbiales (*a la chita callando, de uvas a peras, sin ton ni son, a ciegas, en ayunas*), prepositivas (*por medio de, en lugar de, gracias a*), conjuntivas (*no sea cosa que, de ahí que...*).

Encontramos, también, en su taxonomía otro tipo de estructuras, como las comparaciones (*sordo como una tapia*), las muletillas (*y eso, y tal*), las paremias y las frases hechas o clichés.

A nivel práctico, lo interesante de las aportaciones de este lingüista es que complementa su clasificación con la nómina de actividades que se podrían realizar para desarrollar cada tipo de elementos y opina sobre el nivel de lengua en el que introduciría casa uno de ellos. Así, en las palabras del primer tipo tenemos que concentrarnos desde un nivel inicial, junto con las fórmulas rutinarias, que deben ser tratadas en todos los niveles, debido a que son de gran

rentabilidad comunicativa y abarcan las funciones lingüísticas de cualquier situación comunicativa. El estudio de las colocaciones y de los compuestos sintagmáticos puede iniciarse en el aprendizaje de ELE desde un nivel elemental, si bien es preferible introducirlas en el nivel intermedio y profundizar en ellas en el avanzado. Del mismo modo, pueden ofrecerse algunas combinaciones ya en niveles iniciales, mientras que las muletillas y paremias son más propias de los intermedios.

Finalmente, Higuera García (1997b) entiende también que existen dos tipos de unidades léxicas, las palabras (simples y compuestas) y las unidades pluriverbales lexicalizadas y habitualizadas. Dentro de estas últimas, encontramos las combinaciones sintagmáticas, a las que más adelante llama colocaciones (2004a: 15). Define la combinación sintagmática como “una unidad léxica formada por dos lexemas, que tiene un sentido unitario, semánticamente transparente, con cierta estabilidad y cohesión interna, que no equivale a un elemento oracional desde el punto de vista sintáctico ni morfológico y que sirve para dar cuenta de las restricciones combinatorias de las palabras desde el punto de vista semántico” (Higuera García 1997b: 36). Frente a las expresiones idiomáticas, las combinaciones sintagmáticas son semánticamente transparentes, es decir, su significado es la suma de los significados de las palabras que la integran, tienen cierta estabilidad y presentan cohesión interna, la cual se manifiesta, por poner un ejemplo, en la concordancia en género y número cuando aparecen sustantivos y adjetivos, permiten la inclusión de modificadores, la transformación a pasiva y los cambios de gradación del adjetivo, por ejemplo, pero no la alteración del orden. Por último, concluye afirmando que las combinaciones sintagmáticas sirven para dar cuenta de las restricciones combinatorias de las palabras, hecho crucial en el aprendizaje del léxico como lengua extranjera¹.

Una idea especialmente importante desde la perspectiva de la didáctica del léxico a extranjeros que comenta Higuera García es que la combinación sintagmática tiene un sentido unitario, de ahí que semánticamente se pueda

¹ Puede encontrarse un estudio más completo sobre las colocaciones y la enseñanza de español en Higuera García (2006, 2007).

considerar una unidad. Ya Lewis (1993) defendía que enseñar *combinaciones sintagmáticas* es decisivo porque ayuda a formar redes de significados entre palabras, con lo cual se favorece su aprendizaje y es una estrategia que ayudará después al aprendiz a recuperar más fácilmente las unidades necesarias en una situación concreta. Lo fundamenta la idea de que así se almacenan las unidades léxicas en el lexicón, es decir, que en la memoria no se almacenan palabras aisladas, sino unidades de significado más amplias. En la práctica consiste en enseñar a segmentar el discurso en unidades significativas (*pedagogical chunking*).

Por otra parte, son miles las combinaciones de este tipo que existen en cada lengua, así que debemos pensar también en qué combinaciones sintagmáticas se deben enseñar. Las combinaciones sintagmáticas que son propias de cada lengua, y por lo tanto son diferentes en la lengua materna y en la extranjera, son las que crean problemas y son las que fundamentalmente debemos enseñar, mientras que las que derivan del conocimiento del mundo no es tan preciso enseñarlas (Bahns 1993).

En el mismo nivel se encuentran las expresiones idiomáticas, o frases hechas, y las *expresiones institucionalizadas*. Esta autora no comenta en qué fase se deben introducir estas últimas aunque, siguiendo la línea de Lewis, defiende la importancia de las colocaciones desde un grado inicial debido a que tienen un papel decisivo en la enseñanza de una lengua extranjera. Su carácter pragmático deja clara la intención del hablante, por lo que su enseñanza sirve para agilizar las partes más rutinarias de la conversación y mejorar la competencia comunicativa. En lo que se refiere a su didáctica, lo importante es que están fijadas por el contexto y por la intención del hablante. Además, el alumnado reconoce su carácter de unidad o fórmula, ya que en sus lenguas maternas también se utilizan procedimientos parecidos.

También las *expresiones idiomáticas*, precisamente por ser idiomáticas, es decir, que su significado no se puede entender a pesar de conocerse el significado de las palabras que las integran, suponen una dificultad para un aprendiz extranjero, tanto a la hora de comprenderlas como de producirlas, por lo que deben ocupar un papel destacado en la enseñanza del léxico. De todos los tipos

señalados, ocuparán un papel privilegiado las frases hechas y, en menor medida, las combinaciones fijas.

No obstante, no se deben establecer fronteras nítidas entre unos tipos y otros de unidades léxicas, sino que resulta preferible adoptar una clasificación no discreta desde el marco teórico de la lingüística cognitiva, que permite ver en ella conjuntos borrosos, ejemplares que pertenecen a categorías intermedias, intersecciones entre conjuntos o, como también se ha denominado, elementos prototípicos y otros no prototípicos.

Capítulo 2. El desarrollo de la competencia léxica

Como hemos comentado en el capítulo anterior, la principal ventaja de enseñar unidades léxicas es que, al utilizar un concepto más amplio que el de palabra, ayudamos al alumnado a tomar conciencia de la combinación de los vocablos, de este modo favorecemos el proceso de aprendizaje y recuperación y mejoramos la competencia comunicativa. No obstante, convertir la enseñanza del léxico en algo más que enseñar palabras tiene consecuencias didácticas y es importante que el docente sepa distinguir las y elija las técnicas precisas para enseñarlas.

Existen diversas propuestas metodológicas para el desarrollo de la competencia léxica en español², aunque nos parecen especialmente completas las de Gómez Molina (2004a: 4-20 y 2004b: 804-807) e Higuera García (2009: 117-125). Ambos beben, sobre todo, del enfoque comunicativo³ y del enfoque léxico y ofrecen un amplio repertorio de tipos de actividades, técnicas y consejos prácticos para el profesado de lenguas extranjeras que desee potenciar el componente léxico en sus clases.

² Propuestas metodológicas para la enseñanza del léxico en español: Lahuerta & Pujol (1996), Gómez Molina (1997, 2000a, 2000b, 2005a), Higuera García (1996, 1997a, 1997b, 2004a, 2004b), Cervero & Pichardo (2000), Alvar Ezquerro (2003). Y en el ámbito anglosajón: MacCarthy (1990), Lewis (1993, 1997b, 2000), Nation (2001), Bogaards (2001).

³ Algunos principios y técnicas del enfoque comunicativo son: la atención a las necesidades concretas del alumnado, la atención a las cuatro destrezas, la incorporación del componente lúdico en el aprendizaje, la enseñanza explícita de estrategias para conseguir su autonomía, el respeto a distintos estilos de aprendizaje. Y en concreto en el caso del léxico, su enseñanza contextualizada (aunque también pueda ser lícita la enseñanza descontextualizada para determinado léxico de objetos concretos y en niveles iniciales) y la necesidad de relacionar los contenidos léxicos con otros que integran la competencia comunicativa (Salazar García 1996; Nation 2001: 382).

Siguiendo su estela y de modo general exponemos algunas consideraciones básicas y prácticas para el desarrollo de la competencia léxica que ambos autores proponen en diferentes publicaciones. Así, Higuera García propone algunas técnicas para el tratamiento del léxico que se basan en tres pilares: presentar el significado de las unidades léxicas (con apoyo de imágenes, definiciones, ejemplos, traducciones, uso de gestos, etc.), mostrar las semejanzas y diferencias entre las palabras y ayudar a crear redes entre ellas (Higuera García 2004a: 17). Algunas de estas técnicas son las siguientes:

- presentar el léxico nuevo en series cerradas de elementos (días de la semana, colores, meses del año, etc.)
- uso de escalas para presentar las palabras (ejemplo: *helado, frío, templado, caliente, ardiente*);
- presentarlas por categorías;
- presentarlas por campos semánticos o por relaciones de sinonimia y antonimia, trabajar los sinónimos y antónimos en contexto y buscar colocaciones sinónimas en un texto, resaltar relaciones de inclusión (mediante árboles, por ejemplo, tipos de casas: *chalet, piso, estudio, adosado*, etc.);
- elaborar mapas semánticos, que pueden servir para presentar nombres, verbos, colocaciones y expresiones sobre ese mismo tema;
- uso de tablas o parrillas para mostrar colocaciones, también en el caso de palabras sinónimas (por ejemplo, *mirar y ver*);
- seleccionar las unidades léxicas más importantes de un tema o texto y trabajar sus colocaciones.

Y añade otros aspectos, como que el desarrollo del léxico sirve para implementar la carga cultural del aprendizaje, pues ambos están ligados, que la presentación de palabras sencillas descontextualizadas es una práctica legítima en los niveles elementales, y que lo importante cuando se presenta el léxico es el cotexto (es decir, las palabras que están inmediatamente antes y después), más que el contexto o la situación comunicativa. En un trabajo posterior expone de forma detallada sus sugerencias metodológicas para la enseñanza del léxico, con

las que desarrolla cada una de estas técnicas mediante tipos de actividades concretas y variadas (Higueras García 2009: 117-125).

Por su parte, Gómez Molina indica que es importante que el alumnado use la lengua como lo haría un hablante nativo, experimentando y deduciendo las reglas. Aporta una serie de técnicas y estrategias, algunas de ellas casi idénticas a las anteriores, como son la exposición a textos auténticos, la búsqueda en todo tipo de diccionarios, inclusión de vocabulario en contexto, explorar campos semánticos y construir mapas mentales, practicar estructuras léxicas (formación y descomposición) y tipos de unidades léxicas (combinaciones) con actividades de práctica controlada, memorización, actividades relacionadas con las cuatro destrezas y complementariedad de léxico y gramática (Gómez Molina 2005a: 93). En otro trabajo enumera las que considera premisas fundamentales que deben orientar las actividades que desarrollen el conocimiento y uso del vocabulario (Gómez Molina 2004a: 4):

- aprendizaje en diferentes grados: exposición, interpretación, práctica contextualizada, retención (memoria a corto plazo), uso y mejora (memoria a largo plazo, lexicón);
- uso de estrategias (a las que nos referiremos más adelante);
- el docente debe poner en práctica diferentes técnicas, sin olvidar el uso de las Tecnologías de la Información y la Comunicación (TIC);
- en los niveles iniciales apoyar el aprendizaje en métodos lúdicos (sopas de letras, crucigramas, letras desordenadas de una palabra, describir imágenes,...) y técnicas asociativas (mapas mentales, pares asociados imagen-palabra, asociaciones contextuales, tarjetas, parrillas, etc.);
- en los niveles avanzados, trabajar mediante el procesamiento verbal (asociaciones lingüísticas y gramaticales, buscar o interpretar los significados mediante claves lingüísticas y no solo contextuales, etc.)

Resulta especialmente útil la propuesta de este autor ya que, al tiempo que define cuáles son para él las diferentes unidades léxicas de la lengua, va aportando tipos de actividades con las que se pueden trabajar cada una de ellas, añadiendo además consideraciones y distinciones para cada nivel de la lengua (2004a: 4-20).

2.1. Métodos y enfoques didácticos⁴

La enseñanza del vocabulario siempre ha sido relevante en la historia de la enseñanza de lenguas. No obstante, no siempre se le ha dado la misma importancia ni se ha tratado de la misma forma. En el *método clásico*, el estudio del vocabulario tomaba las etimologías como forma para explicar los significados y la ortografía. Esta había sido, por ejemplo, la base del aprendizaje del latín y el griego. En el *método gramática y traducción*, que surgió en Prusia a finales del siglo XVIII, el vocabulario era presentado en listas de palabras, con sus respectivas traducciones para memorizar, y cuando el alumnado preguntaba en clase el significado de una palabra desconocida se le respondía ofreciendo la traducción del vocablo. La memorización de las palabras se afianzaba mediante ejercicios de traducción, directa e inversa, y se trabajaba el reconocimiento de los falsos amigos.

En el siglo XIX, con el movimiento reformista, surgió el *método directo*, que desechó las listas de palabras descontextualizadas a favor de una enseñanza que priorizaba el vocabulario en lugar de la gramática y apostaba por una didáctica situacional, en la que el significado depende del contexto y la situación comunicativa. Así, se emplean objetos y dibujos en lugar del término en la lengua materna, presentando los significados mediante asociaciones con la lengua meta. Esta opción tenía como base que los niños aprendieran la lengua inicial haciendo asociaciones directas entre los objetos y las palabras, por lo que esa debía de ser la forma más natural de aprender una lengua, aunque fuese una L2. Las palabras dejan de enseñarse de forma aislada y pasan a contextualizarse, por ejemplo en frases, se enfatiza la comunicación oral, por lo que se prioriza el vocabulario cotidiano mediante demostraciones, objetos, dibujos. Esto provoca un aumento de las ilustraciones en los manuales. En la siguiente imagen podemos observar un ejemplo de un conocido manual para enseñar inglés de Eckersley (1938):

⁴ Una visión más completa y amplia sobre estas cuestiones se puede consultar en Richards & Rogers (2011), autores a los que seguimos en estos breves apuntes.

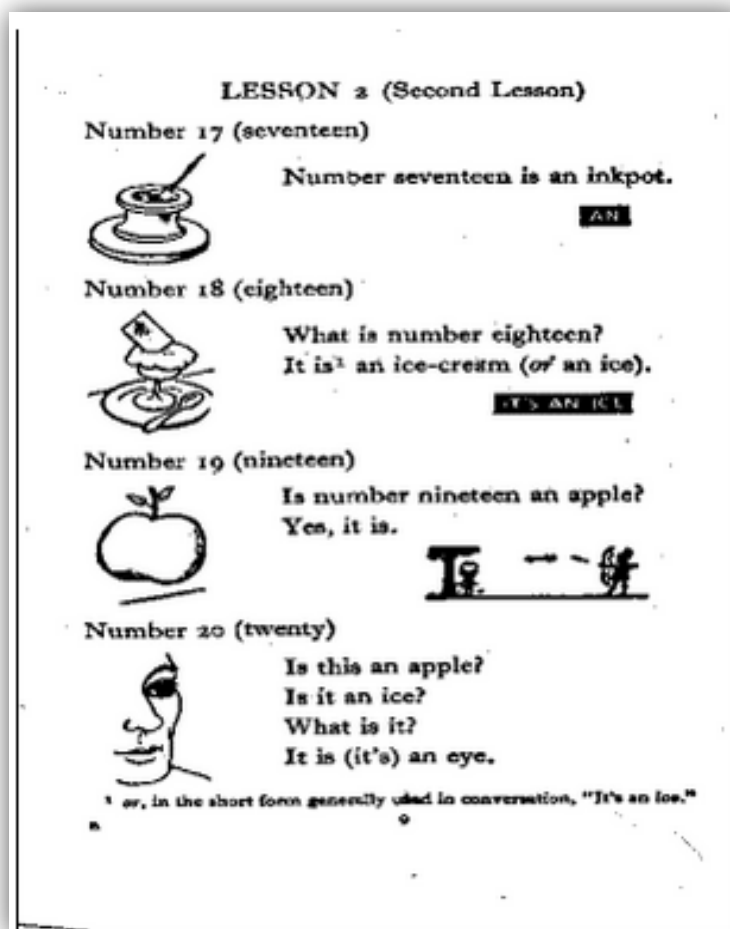


IMAGEN 2. Página interior del manual de Eckersley (1938)

Aunque este método defiende que el profesorado debe crear formas para que la lengua se use como en la vida real, se puede observar que en los enunciados no se guarda relación con la lengua coloquial, sino que sirven solo para introducir el vocabulario nuevo y las estructuras sintácticas. Las frases “Is this an apple?” y la respuesta “It is an eye”, tienen como objetivo el aprendizaje de los demostrativos, del verbo *to be* y de las palabras *apple* y *eye*, pero no resultan de ningún modo espontáneas.

Posteriormente, en el *método audiolingual* también se enfatizaba la necesidad de aprender las palabras en su contexto, pero se pensaba que el número de voces debía ser limitado, ya que la adquisición del vocabulario tendría que ser posterior a la del sistema gramatical. Los principales objetivos de este método eran las estructuras lingüísticas, por eso sus materiales didácticos estaban organizados en una secuencia de estructuras gramaticales seleccionadas

en un nivel progresivo de dificultad. Estas estructuras eran repetidas varias veces para que el alumnado pudiese automatizarlas. Véase en la Imagen 3 un ejemplo tomado de uno de los manuales más usados en todo el mundo para la enseñanza del inglés en los años 7:

Teaching Unit 2

Example:

Is this your umbrella?

No. It isn't my umbrella. It's your umbrella.

Now you do the same.

- 1 Is this your pen?
- 2 Is this your pencil?
- 3 Is this your book?
- 4 Is this your watch?
- 5 Is this your coat?
- 6 Is this your dress?
- 7 Is this your skirt?
- 8 Is this your shirt?
- 9 Is this your car?
- 10 Is this your house?

IMAGEN 3. Ejemplo de ejercicio para repetición (Alexander 1967)

Como se puede observar, la estructura que se practica en esta unidad es "Is it your + nombre?". Se pretendía su memorización mediante la repetición de la estructura y la sustitución del sustantivo que la acompañaba. En los enfoques estructurales, por lo tanto, el vocabulario está subordinado a la gramática, pues la adquisición del lenguaje es entendida como el dominio de estructuras sintácticas.

En español también tuvimos aportaciones al método audiolingual de la mano de Rojo Sastre, Rivenc y Ferrer en 1968 con su *Vida y diálogos de España*. En la siguiente imagen podemos ver la portada de uno de los tomos de esta pionera obra de español para extranjeros que incluía un libro de imágenes.

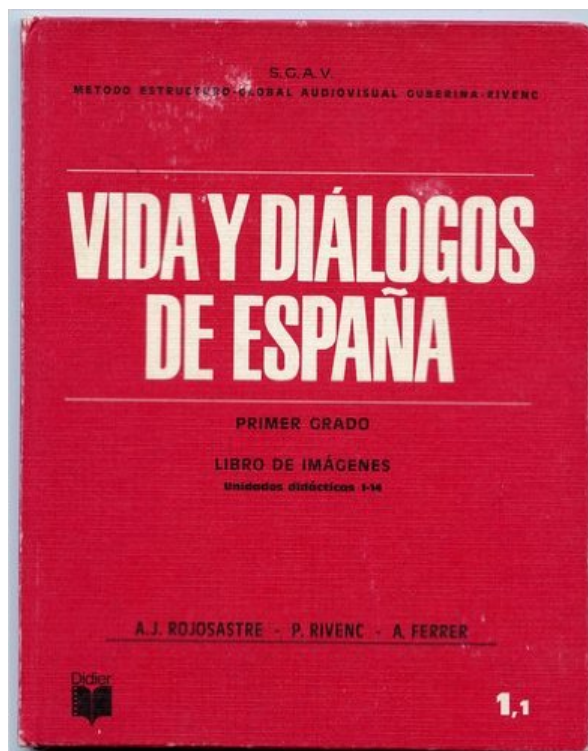


IMAGEN 4. Portada del libro de Rojo Sastre, Rivenc y Ferrer (1968)

Con la llegada del *método comunicativo* el objetivo es presentar el uso de la lengua auténtica y el idioma meta pasa a ser, no solo el objeto de estudio, sino también el medio de comunicación. La gramática y el vocabulario que el alumnado aprende deriva de la función, del contexto situacional y del papel de los interlocutores. De este modo, se deja de concentrar la enseñanza en el léxico y en estructuras concretos, en favor de textos auténticos. Posteriormente este método fue complementado por el *enfoque léxico* de Lewis, que recusa el principio inicial de que la lengua se divide en estructura y vocabulario, y propone observar la gramática subordinada al léxico. En la contracubierta de su primer libro *The lexical approach*, encontramos la síntesis de los principios por los que se rige su propuesta (Lewis 1993):

- The grammar/vocabulary dichotomy is invalid.
- Collocation is used as an organic principle.
- Successful language is a wider concept than accurate language.
- The Observe-Hypothesise-Experiment cycle replaces the Present-Practice-Produce paradigm.
- Most importantly, language consists of grammaticalised lexis – not lexicalized grammar.

En conclusión, la adquisición del vocabulario es, pues, un elemento fundamental para los aprendices de lenguas extranjeras y, como tal, requiere una atención especial⁵. Incluso se puede argumentar que, si bien hasta cierto punto es posible transmitir información sin tener en cuenta la gramática de una lengua, no podemos decir lo mismo del léxico. Aunque discutido y criticado en una fase inicial, el principio que Lewis enunció en su primera obra era que la gramática puede explicarse a partir del léxico. Para él, la lengua es léxico gramaticalizado, no gramática lexicalizada y, en consecuencia, se debe dedicar más tiempo a la enseñanza del vocabulario que a la de la gramática. Más recientemente esto es defendido por diversos autores de las últimas décadas, quienes señalan que en la adquisición del lenguaje se pueden aprender ítems o unidades individuales (el ejemplo de formas como *amado, bebido*) o bien aprender el sistema (en este mismo caso sería ver la cuestión como *amar + ado= amado, beber + ido= bebido*). Pues bien, mientras que en la visión tradicional se defendía que el aprendizaje de las unidades léxicas se conseguía a través del primer principio y la gramática era consecuencia del segundo, en la actualidad se afirma que los dos principios no se excluyen mutuamente. En la práctica, estamos diciendo que una vez que se adquiere un segmento léxico, el aprendiz puede analizarlo, adquirir el significado de cada palabra y aprender la gramática mediante el léxico (Schmitt 2000: 127).

Sin embargo, es muy probable que nuestro alumnado no sea consciente de la importancia de su aprendizaje hasta que ha sobrevivido a la tormenta del estudio de la gramática, que suele ocupar la parte inicial del proceso. Muy al contrario, en el mundo real en el que usarán la lengua, habitualmente son los errores léxicos los que dificultan la comprensión, mientras que los gramaticales se toleran mucho mejor. Por otra parte, el léxico no es un fin en sí mismo, sino que resulta un medio para comunicar y mejorar la competencia comunicativa. Si comparamos el proceso de aprendizaje del léxico en nuestra lengua materna con el de la gramática o la fonología, observamos que el primero llega hasta la edad adulta y que nunca se da por finalizado, mientras que los otros dos concluyen antes. Pero aprender léxico es un proceso lento y complejo que requiere esfuerzo y dedicación por parte del profesorado y alumnado, por lo que corresponde al

⁵ Diversos autores han revisado el tratamiento que ha recibido el componente léxico en los distintos métodos (Salazar García 1994; Zimmerman 1997; Cervero & Pichardo 2000: 12-21).

profesorado subrayar la necesidad de trabajar la competencia léxica, lo cual no solo ayudará a nuestros estudiantes a comunicarse mejor, sino que también ayudará a aumentar su motivación. El estudio del léxico no debe ser entendido solo como retención de ítems léxicos aislados, sino también como el trabajo con colocaciones, fórmulas, mecanismos de formación de palabras, cuestiones sobre su uso, etc.

Con todo, el tiempo en el aula es limitado e insuficiente para un aprendizaje extenso, y, para ampliar el vocabulario, es necesario, además de las actividades de clase, dedicación individual y puesta en práctica de estrategias de aprendizaje eficientes por parte del alumnado. Cabe también al docente esta labor informativa de las posibilidades y estrategias individuales que pueden ayudar en cada caso, pues las personas son diferentes y, en consecuencia, los estilos de aprendizaje también. Muchos profesores y profesoras comentan que sus estudiantes desconocen otras estrategias más allá del uso del diccionario y las listas de palabras, por lo que debemos incentivarlos a desarrollar otras estrategias complementarias, que en ocasiones pueden resultar más productivas.

Otra cuestión a tener en cuenta en el aprendizaje de una L2 es que nuestro alumnado ya domina al menos otra lengua, en el caso de los adultos, o bien está en proceso, en el caso de los niños, por lo que desde el primer momento del aprendizaje intentan organizar el léxico nuevo de alguna manera que les permita almacenarlo y recuperarlo con rapidez y efectividad. También se puede afirmar que hay diferencias significativas entre niños y adultos a la hora de aprender su lengua materna, ya que se aprecia en la infancia un desarrollo de la capacidad de generalización y abstracción, que en adultos se manifiesta con una mayor comprensión y producción de metáforas y una mayor expresividad. Así, también debemos tener en cuenta esta diferencia en nuestras prácticas educativas, pues en el caso de los adultos, potenciaremos estrategias de aprendizaje que permitan este almacenamiento y recuperación del vocabulario mediante ejercicios que enseñen no solo a comprender las palabras, sino a relacionarlas y diferenciarlas (Anglin 1970).

2.2. Tipos de enseñanza del léxico

Así pues, parece que tras años de abandono frente a la enseñanza de la gramática, los estudios sobre la enseñanza del vocabulario están actualmente de moda, pero ¿hasta qué punto los docentes debemos dedicar tiempo y tareas dentro del aula a enseñarlo?, ¿en qué medida debe ser explícito su aprendizaje y cómo? Las respuestas a estas preguntas han sido diferentes, como hemos visto, a lo largo de la historia de la didáctica en función de los diferentes métodos y enfoques, pues, mientras algunos —como el gramatical o el directo— abogaban por una enseñanza explícita del vocabulario, el enfoque comunicativo prefirió en sus orígenes la enseñanza implícita, principalmente usando técnicas de inferencia del vocabulario en contexto. A pesar de que con la enseñanza comunicativa de la lengua había disminuido la enseñanza explícita del léxico, actualmente se concilian e integran ambas posturas gracias sobre todo al enfoque léxico. Este enfoque ofrece las aportaciones más novedosas sobre la enseñanza del léxico y defiende que el profesorado debe estar atento a los dos tipos de aprendizaje y programar los cursos de tal forma que haya oportunidades tanto para el aprendizaje directo como para el indirecto, lo cual no supone una ruptura con el método anterior, sino, todo lo contrario, su mejor desarrollo, ya que investigaciones recientes confirman que es preciso cierto estudio explícito del léxico para dominar una lengua extranjera.

Con todo, no debemos minusvalorar el potencial del aprendizaje implícito, como lo demuestran los estudios de Saragi, Nation y Meister (1978) y más tarde Horst, Cobb y Meara (1998). Estos autores hicieron un test a varias personas que habían leído la novela *La naranja mecánica*, en la lengua “nadsat”, un idioma inventado en esa obra y al que, en principio, ninguno de ellos había prestado atención especial durante la lectura. Los resultados fueron que la mayoría de los entrevistados había aprendido unas 70 palabras de las 241 que tenía ese idioma. De este curioso experimento se desprende que se adquiere léxico a pesar de que no esté dirigida la atención del lector a comprender esas palabras nuevas. A pesar de ello, los estudios sobre adquisición del léxico abogan por usar ambas formas de enseñanza-aprendizaje para mejorar y ampliar estas cifras. En términos de Sökmen, la enseñanza explícita y el trabajo en profundidad sobre las

unidades (*deep level of processing*) contribuye a que las palabras se aprendan mejor (se adquieren más elementos de distintas dimensiones y se favorece su utilización), ya que cuanto más se manipula o procesa el vocabulario más se desarrolla la competencia léxica (Sökmen 1997). Por otro lado, la inferencia es uno de los principales procedimientos para su adquisición y por ello es altamente recomendable favorecer su utilización, pues el gran número de unidades y sus múltiples dimensiones no hacen posible una enseñanza explícita de una cantidad considerable del vocabulario de la lengua meta.

A esto se suma el hecho de que cuando los aprendices de un idioma utilizan buenas estrategias aprenden más, disponen de más información y producen mensajes de mejor calidad y, en consecuencia, se puede afirmar que su manejo favorece ambas formas de aprendizaje: aprendizaje explícito —que implica procesamiento de las unidades en profundidad en situación de enseñanza explícita— y aprendizaje implícito —el que tiene lugar implícitamente en la realización de tareas que no tienen el vocabulario como objetivo—. Consecuentemente, las propuestas actuales sobre didáctica del vocabulario recomiendan el trabajo sobre tres ejes que, conjuntamente, han de favorecer el aprendizaje del vocabulario (Martín Peris *et al.* 2008, *s. v. vocabulario*):

1. Enseñanza explícita y procesamiento en profundidad de parcelas de vocabulario.
2. Enseñanza-aprendizaje implícitos (la lectura extensiva es uno de los principales medios de adquisición de vocabulario).
3. Desarrollo de estrategias.

Los dos primeros puntos son tipos de aprendizaje de naturaleza distinta y persiguen objetivos diferentes, por lo que se potencian mediante distintos modelos de actividades. Ya en la década de los noventa, Nation había distinguido los dos tipos de aprendizaje en relación con la enseñanza del inglés y, así, indica que las dos mil palabras más frecuentes y las ochocientas que componen el vocabulario técnico deben intentar trabajarse de forma explícita, mientras que el aprendizaje indirecto se podría ocupar de las menos frecuentes (Nation 1990: 19). Este autor especifica en una tabla muy ilustrativa para nuestro trabajo el tipo de vocabulario, el número de palabras que lo componen y la *cobertura léxica textual* que alcanzan en un texto, así como su frecuencia (*vid. infra* § 4.1.1.). Para

cada uno de los tipos ofrece observaciones sobre las implicaciones para su enseñanza y aprendizaje. A las unidades más frecuentes se les debe dedicar gran cantidad de tiempo, mientras que a las unidades léxicas técnicas y a las de baja frecuencia se recomienda dedicarles menos tiempo y enseñar estrategias al alumnado para que las adquieran por su cuenta.

Tipo de vocabulario	Número de palabras	Frecuencia	Cobertura, alcance en un texto	Origen	Implicaciones para su enseñanza y aprendizaje
Palabras de alta frecuencia	2.000	Frecuentes en todo tipo de textos.	Cerca del 87% de un texto.	Aproximadamente la mitad proceden del latín, francés o griego.	Invierte mucho tiempo en estas palabras y asegúrate de que las aprenden.
Vocabulario académico	800	Frecuentes en la mayoría de textos académicos.	Cerca del 8% de un texto académico.	Dos tercios proceden del latín, el francés o el griego.	Si los alumnos están en esa situación, dedica mucho tiempo a estas palabras y asegúrate de que las aprenden.
Vocabulario técnico	Entre 1.000 y 2.000 para cada materia.	Aparecen, a veces con frecuencia, en textos especializados.	Cerca del 3% de un texto especializado.		Aprender una materia implica aprender su vocabulario. Los profesores de esas materias pueden trabajar con ese vocabulario y los profesores de inglés pueden ayudar enseñando estrategias.
Palabras de baja frecuencia	Cerca de 123.000	No aparecen frecuentemente.	Cerca del 2% o más de las palabras de cualquier texto.		Enseña estrategias para tratar con estas palabras. Las palabras en sí mismas no merecen que se les dedique tiempo de enseñanza.

TABLA 1. Relación entre el tipo de vocabulario, cantidad, frecuencia, cobertura textual e implicaciones didácticas (Nation 2000, *apud* Higuera García 2004a: 8)

El mismo autor comenta también algunas técnicas que se relacionan con cada tipo de aprendizaje (Nation 2001: 16):

Enseñanza directa ==>	Explicaciones del profesorado Enseñanza entre compañeros
Aprendizaje directo ==>	Estudiar a partir de tarjetas que contienen palabras Utilizar el diccionario
Aprendizaje incidental ==>	Deducir por el contexto en lecturas extensivas Utilizar el léxico en actividades comunicativas
Encuentros planificados ==>	Lecturas graduadas Ejercicios de vocabulario

CUADRO 3. Relación entre tipos de aprendizaje y algunas de sus técnicas ((Nation 2001, *apud* Higuera García 2004a)

Schmitt (2000) mantiene opiniones parecidas y explica que el aprendizaje explícito centra su atención en la información que hay que aprender y garantiza su incorporación, pero requiere invertir gran cantidad de tiempo, por lo que resulta imposible aprender así todas las palabras que necesitamos para hablar una lengua extranjera. Por el contrario, el aprendizaje incidental o no intencionado resulta de usar la lengua con fines comunicativos y, consecuentemente, el tiempo empleado en él es doblemente beneficioso, pero tiene el inconveniente de que es más lento, en cuanto se trata de un aprendizaje gradual y menos sistemático. Este autor recomienda también diferentes técnicas y tipos de actividades y, así, indica que para el tratamiento incidental del léxico las actividades de comprensión oral y escrita son las adecuadas, ya que se expone al alumnado a gran cantidad de *input*. En la misma línea de Nation, apuesta para el aprendizaje explícito en el caso de enseñar los significados más frecuentes de cada palabra, ofrecer información sobre familias de palabras y pensar cuáles son las palabras más problemáticas para cada grupo meta en concreto (2000: 116-141). Por su parte, Sökmen (1997) expone también las técnicas que él considera para la enseñanza explícita del léxico: crear un gran lexicón; integrar las nuevas palabras con las ya conocidas; proporcionar varios encuentros con las palabras enseñadas; trabajar con las palabras en tareas de distinta complejidad, lo que supone no sólo memorizar, sino conseguir además

un nivel de procesamiento profundo; facilitar la creación de imágenes; hacer que las palabras nuevas parezcan reales para intentar conectarlas con el mundo real del alumnado; usar varias técnicas, y animar a los aprendices a que empleen técnicas de aprendizaje autónomo.

Con respecto al desarrollo de estrategias, no hay un consenso entre los autores que han tratado el tema sobre la necesidad de enseñarlas o no, o sobre el lugar que las estrategias ocupan en la planificación de la formación. Las propuestas que abogan por la enseñanza explícita de estrategias para el aprendizaje del léxico las aprueban como forma de suplir el tiempo que no se puede dedicar en clase y paliar el hecho de que las unidades léxicas sean tan numerosas. Entre estas propuestas se encuentran las aportaciones de Chamot *et al.* (1999: 11 y 140), que apuestan por la enseñanza explícita de estrategias en la clase presencial y proponen cinco pasos para llevarla a cabo: preparación (para activar el conocimiento previo), presentación de la estrategia, práctica, evaluación del empleo de esa estrategia y expansión a otras situaciones. Por su parte, Higuera García (2004a: 20-22) defiende que deben enseñarse asociadas a otra serie de contenidos, como pueden ser los léxicos, y que deben formar parte de la clase presencial, pues juegan un papel decisivo en cada una de esas etapas para la retroalimentación que recibe el alumnado, tanto por parte del profesor como por parte de sus compañeros, además de que sirven para identificar y compartir las estrategias que emplean entre ellos.

Esta misma autora destaca la taxonomía, de entre las muchas propuestas, de Schmitt (1997: 207-208), porque combina la clasificación de Oxford (1990) y la completa con la de Nation (1990). Schmitt agrupa las estrategias en dos grandes bloques: estrategias para descubrir el significado de las palabras y estrategias para consolidar su significado. En el caso de las primeras, explica que se puede determinar el significado haciendo listas de palabras, usando el diccionario (bilingüe o monolingüe en función de los casos), deduciendo por el contexto, etc. Además, se pueden desarrollar técnicas sociales, como preguntar la traducción al profesorado o bien a los compañeros, etc. En el segundo caso, se pueden consolidar los significados mediante estrategias sociales (interactuar con nativos, practicar en grupo, etc.), de memorización (crear mapas semánticos, conectar las palabras a una experiencia personal, agruparlas, usarlas en frases,

asociarlas a una imagen, usar palabras clave, etc.), cognitivas (repetir las palabras de forma oral o escrita, tener un cuaderno para el léxico, hacer listas, tomar notas en el aula, etc.) y metacognitivas (planificar, repasar, autoevaluar, etc.).

Higueras García (2004a) resalta también dos estrategias que destaca por su importancia a la hora de que el alumnado adquiera de forma autónoma el léxico: entrenarlo en el manejo del diccionario y enseñar a deducir por el contexto. Señala también que la segunda se debe intentar antes de recurrir a la primera. Se puede también intentar descubrir las palabras clave de un texto y luego deducir las palabras que se quieren aprender para, por último, recurrir al diccionario. Antes de usarlo, conviene explicarles cuándo conviene consultarlo y cómo (comentar sus partes y estructura, la información que ofrece, las abreviaturas y formas para buscar información gramatical, léxica y de registro). Respecto a la deducción por el contexto, ya Nation (2001: 233) indicaba que era necesaria cierta instrucción, pero que se puede aplicar a textos tanto orales como escritos. Y algo importante, que solo resulta efectiva con textos sencillos, en los que los aprendices conocen al menos el 95 % de las palabras que los integran. Volveremos en el siguiente capítulo a tratar estas cuestiones sobre el alcance en la comprensión.

Conviene también que el alumnado reflexione sobre la pertinencia de emplear cierta técnica en una situación concreta en función de los objetivos de la tarea, ya que enseñar estrategias no se limita a enseñar técnicas. Aplicado a la deducción por el contexto, Nation elaboró una actividad que muestra cinco pasos en los que un alumno ficticio reflexiona sobre esta estrategia tras la lectura de un texto sobre la Feria de Abril. La actividad se aplica a un grupo de nivel intermedio y consiste en la lectura de los pensamientos de un alumno irreal que intenta deducir el significado de las palabras. Una de ellas es “caseta”, y, para ello, busca los elementos que la acompañan (artículos, adjetivos, otros elementos de la frase, el cotexto, etc.) para intentar deducir la clase de palabra y su significado. Intenta ver a qué palabra de su lengua materna equivale e intenta descomponerla en partes, para luego confirmar o no sus hipótesis. Por último, si no se encuentra seguro de sus deducciones, recurrirá al diccionario (Nation 2001:233).

Gómez Molina (2004a: 4) es de la misma opinión, pues defiende el uso de una variedad de actividades que faciliten el desarrollo de estrategias de aprendizaje y comunicación. Apuesta por ello, en la medida en que las estrategias determinen el éxito en la adquisición del léxico. En general, no hay tareas buenas o malas, se puede partir de las que el alumnado ya conoce previamente por el bagaje de aprendizaje que acumula de su propia lengua o de otras lenguas extranjeras. Podemos también incitarlos a descubrir y probar otras nuevas y hacerlos conscientes de que en su aplicación debemos tener en cuenta el objetivo de la tarea, para evaluar cuáles son las estrategias más efectivas en cada caso⁶.

En conclusión, podemos explicar los beneficios de usar estas estrategias con las palabras de Higuera García:

Desde nuestro punto de vista, la enseñanza de estrategias está en consonancia con el principio del enfoque comunicativo, que otorga al alumno un papel protagonista y le hace responsable de su aprendizaje y responde a la necesidad y el deseo de conseguir que los alumnos sean autónomos en su aprendizaje. Permite que el alumno siga aprendiendo fuera de clase, que cree oportunidades de aprendizaje mediante la lectura o la audición de textos y que se produzca un aprendizaje incidental del léxico. Sólo de esta forma podrá el alumno aprender las palabras de baja frecuencia o el léxico especializado, al que es imposible dedicar mucho tiempo de enseñanza directa en las clases. [Higuera García 2004a: 22]

2.3. Características del vocabulario

En los epígrafes anteriores nos hemos detenido en dos aspectos importantes a la hora de hablar de la selección léxica: la complejidad del léxico y la importancia del vocabulario en el aprendizaje de una lengua.

La primera característica, su alta *complejidad*, deriva de su carácter abierto, variable e ilimitado, es decir, al hablar del léxico de una lengua estamos tratando con un número de unidades casi infinito, en las cuales se manifiesta la variación (diatópica, diastrática y diafásica) de forma más evidente. Además, las unidades léxicas están sujetas a un gran dinamismo a lo largo del tiempo (diacronía). Fernández Leyva lo explica en un trabajo reciente de esta manera:

⁶ Varios autores han estudiado la correspondencia entre tipos de estrategias y las actividades que las potencian, por ejemplo Oxford (1990) y Cervero & Pichardo (2000: 131-168).

[...] el léxico es uno de los contenidos que mayor dificultad tiene, especialmente en lo que respecta a su selección y presentación, de modo que posibilite que los estudiantes lo adquieran. Esto se debe a su complejidad, dada por su alto índice de variabilidad y por la gran cantidad de unidades que lo componen. [Fernández Leyva 2015: 2]

Y continúa su explicación con una cita de Müller al respecto: “pues sin ser infinito en el sentido matemático del término, no nos da nunca la impresión de ser estrictamente finito” (Müller 1973: 227). Este grado de dificultad disminuye al hablar de vocabulario, pues con este término no abarcamos todo el léxico de una lengua sino su actualización en el discurso, lo cual se plasma en lo que podemos llamar *vocablos*. El vocabulario, por tanto, se define como el conjunto de vocablos que aparecen en la lengua oral o escrita y representa una parte del léxico.

Se suele explicar que los hablantes nativos no utilizamos todo el caudal léxico para expresarnos de forma clara y precisa, ni tan siquiera usamos un gran número de unidades (Benítez Pérez 2009: 9-10; Justicia 1984: 7) (*vid. supra* § 1.5). Para tener una idea aproximada, podemos seleccionar las cifras que barajaba Verlée en este sentido. Este autor limitaba a 2.000 vocablos el número de unidades que un hablante nativo medio usa en su vida cotidiana, familiar, social y profesional. A esa cantidad se le debe sumar un número variable de términos técnicos específicos del trabajo que desempeñe. Para “un intelectual”, la cantidad sería 5.000 vocablos, lo que ya le posibilita para poder hablar de temas complejos (Verlée 1954: 5). Pero no existe un consenso definido sobre cuáles son las cifras exactas del número de palabras que manejan los hablantes nativos en función de su formación y profesión. Las diferencias en el vocabulario de los adultos pueden ser también notables y difíciles de determinar, así lo han manifestado autores como Watts, que indicaba que la dificultad de determinar el vocabulario de un adulto es mucho mayor que la de determinar el de un muchacho (Watts 1946: 57-8). Los datos que apuntó Max Müller (1862) a este respecto son conocidos. Este autor dijo a mediados del siglo XIX que el vocabulario de un aldeano inglés abarcaba alrededor de 300 palabras. En estas mismas fechas Farrar (1865: 59) consideraba que el vocabulario usado por tres

campesinos durante una conversación mientras estaban recogiendo manzanas no superaba las 300 palabras. Estas cifras contrastan con las que aportó Smedberg (1896 *apud* Watts 1946), quien valoró el vocabulario de un campesino sueco con unas 26.000. Según comenta Watts, Douglas Hyde, una autoridad en el estudio del irlandés, indicó que para él el vocabulario de un campesino medio estaba entre las 5.000 y las 6.000, y quizá esta sea una cifra más plausible, escribe Watts, para el tipo de persona que vive apartada de la vida urbana (Watts 1946: 57).

Con mejores oportunidades de educación y una vida más social, inevitablemente el vocabulario mejora, a nivel cuantitativo o cualitativo. Watts defendía que a los 21 años un adulto de una ciudad posee un vocabulario de 15.000 palabras, que se puede haber desarrollado desde la infancia al ritmo de 700 palabras por año desde los 14, edad en la cual la media son las 10.000 palabras. Ogden (1930) explicó que el vocabulario de una camarera de un salón de té en su época era de unas 7.000 u 8.000 palabras, el de un dirigente medio de un sindicato estaba entre las 10.000 y 15.000, el de Winston Churchill estaba entre las 25.000 y las 30.000 y el de un científico profesional entre las 60.000 y 80.000. El profesor García Hoz (1953) resuelve la cuestión manifestando que se puede utilizar como criterio la cifra de palabras usuales que conoce un adulto de tipo medio. Reconoce que determinar “el vocabulario del hombre medio” es bastante problemático, sin embargo, su solución sería estimar el “vocabulario usual”, constituido por palabras tomadas de la vida familiar, callejera, laboral y cultural. Este vocabulario abarca, para García Hoz, todos los términos que le aparecieron alguna vez en los documentos o fuentes de distinto tipo que utilizó él mismo en su investigación. Así, estos sumarían en su propuesta 12.655 palabras.

Más recientemente hay otras visiones, también relativas a las cifras sobre el conocimiento que un individuo tiene del caudal léxico de su lengua materna. Nation, por ejemplo, señaló que, según algunos estudios, los hablantes nativos de inglés con formación académica conocen alrededor de 20.000 familias de palabras (Nation 2001). Según Alvar Ezquerro (2003: 99) una persona culta puede llegar a tener una competencia de unas 30.000 palabras. Para Aitchison (1987: 7) esa cantidad es incluso mayor, ya que estima que el número de

palabras que un hablante medio conoce es superior a 50.000 y puede llegar hasta 250.000. Las cantidades que postulan las investigaciones psicolingüísticas más plausibles son considerables. Pérez Basanta (1999: 274) recoge la cifra de 17.500 como la media de vocablos que conoce un hablante de inglés de 18 años. Por su parte, Goulden, Nation y Read (1990) estiman que un estudiante universitario, hablante nativo de inglés, domina una media de 17.000 familias de palabras. Dado que, según los propios cálculos de Nation (1983), la relación en dicha lengua entre familias y unidades léxicas individuales mantiene una proporción de 1'6, eso nos da un volumen medio total de 27.200 vocablos diferentes. Estos datos son significativos, pues atañen a sujetos de la misma edad que la de los encuestados en el Proyecto Panhispánico de Estudio sobre la Disponibilidad Léxica (PPHDL)⁷, al que más adelante aludiremos (*vid. infra* § 6.2.2.).

En resumidas cuentas, el total de unidades que componen la competencia léxica de un hablante nativo se encuentra siempre muy por encima del número que se actualiza de modo efectivo en las situaciones habituales de uso. Si bien no disponemos de datos fiables a este respecto, sí sabemos que las unidades léxicas que se emplean habitualmente son muchas menos que las que componen la competencia léxica de los hablantes. Este es el léxico medio de una comunidad: el que suelen emplear todos sus miembros. Alvar Ezquerra apuesta por establecer esta cantidad entre 3.000 y 5.000 términos (Alvar Ezquerra 2005: 21). Para el francés, se cifra esa cantidad en unas 7.000 u 8.000 palabras (Picoche 1994: 48). Como se puede observar, estas cantidades sobre el conocimiento del léxico que tienen los hablantes serían muy difíciles de alcanzar en la enseñanza de una lengua extranjera, por lo que la frecuencia vendría a resolver esta dificultad, aunque solo sea en parte, ya que nos ayuda a determinar cuáles son las más usadas en la práctica. Así, partimos del conocimiento de que el número de palabras que utiliza un hablante nativo es reducido, así como el léxico medio de

⁷ Reúne a numerosos grupos de investigación de España e Hispanoamérica bajo la dirección del profesor Humberto López Morales e intenta elaborar un diccionario del léxico disponible de todos los países donde se habla español. Actualmente, ya contamos con gran número de aportaciones de diccionarios de disponibilidad de áreas concretas y ya se está trabajando en el diseño de la planta de lo que será el diccionario del léxico disponible de España.

una comunidad. Esas cifras son hacia las que debe apuntar el vocabulario de un estudiante de segundas lenguas. Así pues, resulta de gran utilidad saber cuáles son las voces más frecuentes, porque de ningún modo podemos pretender que alguien que aprende una segunda lengua, cuando la domine, conozca las cerca de cien mil voces que hay en los diccionarios generales, 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 aproximadamente 20.000 o 30.000 palabras que, según hemos explicado, se cree que constituyen la competencia de un hablante culto.

No obstante, continúa siendo una labor ardua seleccionar el vocabulario que consideramos que es el más usado por el “español de la calle” o “el español medio”, como estos autores lo denominan, y el que consideramos más útil para transmitir a nuestro alumnado en las diferentes fases de su proceso de aprendizaje. Sea cual sea el número de voces que un hablante medio usa en su día a día, debido a la variación léxica y a la propia evolución de un idioma, continúa siendo de gran dificultad hallar métodos científicos fiables que seleccionen esos vocablos más representativos y elementales. Otra propuesta de este tipo, a modo de ejemplo, es el *Vocabulario del español hablado* (Márquez Villegas 1975), que cifraba en 4.000 las palabras más útiles. Especificaba, además, que en un curso de una lengua extranjera se deben introducir en un nivel inicial 1.018, en un segundo nivel 1.198 y en un tercer nivel 1.905, lo cual supone que al alumnado que supere el curso completo se le habrá expuesto a 4.121 palabras diferentes (Márquez Villegas 1975). Revisitaremos estos números y otras propuestas más adelante (*vid. infra* § 4.1.).

Ya hemos comentado que al vocabulario se le ha otorgado una mayor primacía en la enseñanza de una lengua desde los años 90, después de una etapa en que el auge del enfoque comunicativo a principios de los 80 lo había colocado en un segundo plano frente a la gramática, aunque tampoco en ella se hacía demasiado hincapié. Existen obras teóricas “de cabecera” responsables de este giro⁸, así como algunas publicaciones que son fruto de esta nueva visión sobre el

⁸ Así, en inglés se puede mencionar a Lewis, impulsor del método *lexical approach*. Algunas obras tanto de él como de otros autores de interés al respecto son: Nattinger & DeCarrico, (1992), Lewis (1993, 1997a, 1997b), Cervero & Pichardo (2000), Nation (2001) y Thornbury (2002).

modelo de enseñanza de una lengua extranjera, como las publicaciones sobre técnicas, ideas y manuales⁹ para enseñar léxico o los diccionarios combinatorios.

El cambio de perspectiva ha tenido consecuencias claras en la organización de muchos cursos que toman como hilo conductor el vocabulario, a través del que se insertan las normas gramaticales. Y, si bien es cierto que el enfoque léxico apuesta también por el aprendizaje incidental del léxico, en un gran porcentaje a través de lectura y audición de textos reales, nos ha aportado otras nuevas visiones: resaltar los segmentos léxicos o *chunks*, la importancia de la organización sintagmática u horizontal y de las colocaciones, apostar por una enseñanza cualitativa del vocabulario y no solo cuantitativa, etc.

Fue en la década de los 90 del siglo pasado cuando se empezó a hablar de dos principios básicos para explicar el vocabulario y su organización: el principio de selección libre y el principio de idiomática (Sinclair 1991). El *principio de selección libre (open choice principle)* afirma que el léxico se entiende como un conjunto de palabras individuales que se pueden emplear con mucha libertad y que sólo están limitadas por reglas gramaticales. El de *idiomática (idiom principle)* defiende que un hablante dispone de un número amplio de bloques semiconstruidos que puede combinar al hablar. Gracias a esto sabemos que el lenguaje consiste sobre todo en segmentos léxicos prefabricados, formados por varias palabras que el hablante va ensamblando cuando habla o escribe, por lo que aprender vocabulario es como construir una gran pared en la que se van uniendo y ensamblando pequeños ladrillos, las palabras. Pensemos en el conocido juego de Lego, y en sus bloques de colores, y nos ayudará a entender visualmente estas afirmaciones.

Y ¿por qué *la importancia del vocabulario* en la enseñanza-aprendizaje de una lengua? La primera respuesta viene al comparar este proceso con la adquisición natural de una lengua, en la que las palabras constituyen el primer paso, sobre todo los sustantivos, que actúan como potenciadores del resto del aprendizaje lingüístico. Sin embargo, como docentes e investigadores sabemos que, a pesar del esfuerzo realizado para describir ambos procesos (adquisición y

⁹ *DidactiRed*, Instituto Cervantes, *Léxico fundamental de español (SGEL)*, *Vocabulario en movimiento* (Edinumen), *Uso interactivo del vocabulario* (Edelsa), *Modismos en su salsa* (Arco libros), etc.

aprendizaje), no existe acuerdo que permita hablar con seguridad de un paralelismo. No obstante, no se trata del mismo procedimiento, ya que existen dos diferencias básicas en el proceso de aprendizaje de una segunda lengua, esto es, la pre-existencia del sistema conceptual de la lengua materna y la mediación conceptual que este sistema establece con los referentes aludidos por la L2. Por una parte, en el contexto de aprendizaje de una L2, las palabras no funcionan como potenciadores del conocimiento del mundo o como un contextualizador. El alumnado no tiene que aprender el concepto de “mamá” y no se equivoca extendiendo su significado a otras señoras, pues la segunda lengua se instala sobre el conocimiento del mundo ya existente. Podemos pensar que existen diferencias culturales y esquemas que se reflejan en la lengua, y que debemos deconstruir cuando aprendemos una segunda lengua extranjera, pero esto no impide que este procedimiento funcione así en rasgos generales. A veces la L2 se instala sobre las estructuras o la red cognitiva de la lengua materna. Pero, aunque esta pueda ser una constatación bastante obvia, los métodos de enseñanza que hemos usado han ignorado en muchos casos las potencialidades de contar con ese conocimiento previo. Todo lo contrario, empezaban por las funciones y por los contenidos gramaticales, en lugar de tener como pilares las funciones y el vocabulario.

Un segundo argumento, centrado en el ámbito de la enseñanza-aprendizaje de una LE, para demostrar la importancia del vocabulario y la necesidad de partir de él para planificar nuestros cursos, es observar los errores propios de la interlengua de ELE y evaluar su gravedad. Por ejemplo, analizar qué tipos de errores se generan cuando no recordamos una palabra, o cuando no sabemos si es masculina o femenina, si un verbo es regular o no, si rige preposición, etc. La clave es ver si se producen errores que interrumpen la comunicación o no, y cuáles son los tipos de errores en los que esto ocurre. Varios autores han analizado errores para ver estas cuestiones, pero seguimos aquí dos trabajos de Ainciburu en los que ofrece muestras de errores de diferente tipo en la interlengua para demostrar que son los que tienen que ver con el vocabulario los que no permiten la comunicación, y no los de tipo gramatical (Ainciburu 2008: 1; 2014: 3). Aunque un enunciado sea correcto gramaticalmente, si no se hace una elección correcta de las palabras y estas no se

ajustan a lo que queremos decir, no nos entenderemos. Al contrario, los errores gramaticales no suelen impedir la comunicación, a pesar de que un enunciado sea incorrecto. Un ejemplo claro sería: “Nos encontramos”. Si queremos ver a alguien en la biblioteca de nuestro pueblo o facultad, probablemente nunca conseguiríamos encontrar a esa persona. Si, en cambio, presento a alguien diciendo que “Esto señor es mi profesores”, lo normal es que cualquier persona deduzca que se trata de un individuo que es mi profesor. Y esto es así no sólo si hablamos de niveles iniciales, pues parece ser, también, que las carencias léxicas son más notables que las gramaticales en el alumnado de ELE a la hora de determinar su nivel de lengua. Es decir, la diferencia entre un nivel intermedio y uno avanzado se detecta antes por la falta de unidades léxicas disponibles en el lexicón mental que por el control gramatical que demuestre el aprendiz, de ahí que sean aquellas y no el dominio gramatical el que haga progresar a los discentes en un nivel superior (Lewis 2000: 8)¹⁰. Una vez que nos familiarizamos con las estructuras básicas de un idioma, los problemas que se nos presentan cuando viajamos a un país extranjero no surgen de un conocimiento gramatical escaso, sino de nuestro limitado vocabulario. Por otra parte, cuando viajamos al extranjero, no nos llevamos una gramática, sino un diccionario, y esto es así porque, en una situación en la que es importante transmitir y entender la información (como por ejemplo para tomar un taxi, leer un menú, entender un horario o un mensaje electrónico), queremos tener el vocabulario necesario a nuestro alcance, aunque la gramática sea limitada.

Con respecto a la importancia del léxico y la que adquiere el vocabulario en el aprendizaje y evaluación de una lengua de la competencia lingüística, el propio *MCERL* lo refrenda e insiste en ello en varias ocasiones (Consejo de Europa 2001), aunque no con la contundencia y de la forma que algunos autores reclaman (Bartol Hernández 2010). Esta defensa ha ido manifestándose por parte de diferentes especialistas a lo largo de las últimas décadas desde diferentes áreas de la lingüística (el análisis del discurso, la lingüística del texto,

¹⁰ Trabajos que han llamado la atención sobre la importancia del léxico: Nattinger & DeCarrico (1992); Lewis (1993, 1997, 2000); Coady (1997); Nation (1990, 2001). En España han tratado esta cuestión, por ejemplo, Gómez Molina (1997, 2000, 2004), Martín Martín (1999) y López Pérez (2013).

la pragmática, la psicolingüística y la lingüística cognitiva, etc.). Bartol Hernández (2010: 86) subraya algunas consecuencias de este papel nuclear que se le ha empezado a otorgar al vocabulario en el proceso de aprendizaje:

[...] la aparición en el mercado de diccionarios de español como LE y la cada vez mayor importancia que se le está dando a su uso en el aula como fuente y método de aprendizaje; la aparición de materiales focalizados en el léxico específico de diferentes campos profesionales [...], y el surgimiento de métodos como el enfoque léxico, surgido en 1993 con la publicación de la obra de Michael Lewis, *The Lexical Approach*. [Bartol Hernández 2010: 86]

Así pues, si la riqueza, el alcance y el control del vocabulario son importantes en la adquisición de una lengua, no debemos olvidar que también lo serán en la planificación del aprendizaje y en la evaluación del dominio de lengua de nuestro alumnado.

Más controvertida ha sido la defensa de la *enseñanza explícita y planificada* del vocabulario. Por una parte, porque durante algún tiempo se ha asociado con las antiguas listas de vocabulario como método didáctico y, por otra, porque el enfoque comunicativo lo desechaba en favor de los *realia*, textos auténticos que aportan el vocabulario más adecuado a cada tema o función comunicativa trasladada al aula (*vid. infra* § 3.1.). Véanse a modo de ejemplo las palabras de Ávila Martín:

La llegada de los métodos comunicativos en los años 80 hizo que se abandonara el criterio de frecuencia y se diera más importancia a la intencionalidad comunicativa. El aprendizaje del léxico se circunscribe en estos métodos a los textos reales y a situaciones comunicativas que condicionan el aprendizaje. Desde el punto de vista de la frecuencia léxica, este método no interfiere en el conocimiento del léxico más frecuente de una lengua, pues la exposición a textos reales –*realia*– garantiza la repetición de ese léxico básico. Si se dominan las 5.000 palabras más frecuentes, se conoce el 90% de un texto y el resto, el léxico incidental, puede ser trabajado específicamente o con la consulta del diccionario. [Ávila Martín 2010: 172]

Estos fueron los argumentos que se vinieron defendiendo desde que el enfoque comunicativo irrumpió con fuerza en la docencia de las lenguas extranjeras. No obstante, la enseñanza explícita no se basa en que el alumnado elabore listas de palabras, ni parece lógico que si el vocabulario supone un elemento central y esencial en la docencia de una lengua extranjera tengamos que dejarla en manos del azar o en la confianza de que el vocabulario aparezca

en los textos que no dé tiempo a trabajar. Respecto a esto último, pensemos simplemente en la cantidad de textos de diferentes ámbitos que es necesario recopilar para elaborar corpus léxicos basados en la frecuencia (un vocabulario básico, por ejemplo), para que resulten realmente representativos y avalados por criterios científicos. La misma autora acaba su texto afirmando que:

No existe una planificación en la enseñanza y aprendizaje del léxico porque la utilización de textos reales garantiza el conocimiento del léxico más frecuente. A mayor número de textos trabajados, más garantía de que el léxico frecuente de la lengua se conozca mejor, en mayor diversidad de situaciones. [Ávila Martín 2010: 174]

Desde luego, no hay curso de ELE que tenga horas suficientes para cubrir tal cantidad de textos, ni tan siquiera una pequeña muestra que pueda resultar representativa. Hoy en día, sin embargo, son más los que defienden que la importancia del vocabulario tiene que verse reflejada en la docencia explícita y organizada del vocabulario, articulado en campos nocionales, aplicados en contextos reales y trabajando también las unidades léxicas complejas en función de los niveles. Y lo mejor de todo ello es que esto no supone contradecir los argumentos que desde el enfoque comunicativo han defendido el trabajo de textos reales para trabajar la competencia léxica, sino, más bien, un modo complementario de desarrollarla. De hecho, el enfoque léxico incide en la importancia del aprendizaje incidental de vocabulario mediante el *input* que aporta la lectura y audición de textos auténticos. Un gran porcentaje del vocabulario es aprendido de esta forma, y en mayor medida que mediante la enseñanza explícita.

Por otra parte, si la comprensión receptiva efectiva de textos orales y escritos cifra en aproximadamente un 95% o un 97% el número de palabras que el discente debe conocer para poder llevar a cabo un trabajo de inferencia de las palabras que no conoce, parece lógico que en un primer momento debamos presentar los ítems léxicos de otros modos, por lo menos en los niveles iniciales, pues, si no, el alumno no será capaz de aprehender aquellos que desconoce. No quiere esto decir que defendamos las famosas listas de palabras como método de enseñanza-aprendizaje, aunque el empleo de listas representaría un mal menor. En la docencia de ELE se han abandonado a favor de métodos comunicativos y

significativos, aunque hay estudios que recuerdan que existen resultados empíricos del avance del aprendizaje en principiantes usándolas. Aunque puedan resultar en ocasiones tediosas, en la experimentación con principiantes la lista de palabras triunfa frente al contexto solo, porque utiliza la estrategia más simple a nivel mnemotécnico (Ainciburu 2008: 9). Por lo tanto, nuestros discentes no tienen por qué desecharlas totalmente si es que las prefieren. Lo que sí debemos recordar los docentes es que el aprendizaje de las palabras no se produce si las unidades no se repiten un determinado número de veces, que se suele establecer en no menos de cinco. Por tanto, no debemos abandonar la repetición del vocabulario, aunque optemos por métodos no tan aburridos y mecánicos como las listas y recurramos a actividades más actuales y lúdicas, para lo cual nos pueden ser de gran ayuda las TIC. Para la repetición podemos emplear técnicas de mayor impacto cognitivo que ayuden a crear y fortalecer las redes léxicas. Los mapas mentales, por ejemplo, son útiles, como el enseñar a parafrasear o a definir, pero necesitan una masa crítica de palabras en la red léxica del aprendiz que no puede requerirse en un primer momento. De cara al aprovechamiento esta labor de repetición, sería altamente útil y eficaz potenciar la autonomía del alumnado y darle estrategias para que la lleve a cabo, también fuera del aula, para lo cual conocer los recursos que ofrecen las TIC es de gran ayuda.

Capítulo 3. Criterios y métodos de selección léxica. ¿Qué vocabulario enseñar?

Junto a los dos aspectos revisados en el capítulo precedente, que el vocabulario es importante en una lengua y que es diverso y amplio, una cuestión esencial en esta tesis es la de considerar qué vocabulario hay que enseñar y dónde podemos acudir para encontrar los vocablos más apropiados. La respuesta a la primera pregunta parece bastante evidente: debemos llevar a cabo una selección léxica. Esta necesidad ya la apunta el *MCERL*, pues explica que quienes se encargan de elaborar materiales didácticos y de evaluación deben seguir unos principios metodológicos al respecto y aplicarlos. No obstante, es necesario definir cómo, con qué criterios y de qué manera es conveniente llevar a cabo la selección léxica, para no sentirnos como un pescador aficionado en medio del océano.

Un factor importante para seleccionar las unidades léxicas es pensar en su funcionalidad. McCarthy (1999: 238-247) nos da algunas pistas sobre el tipo de unidades léxicas que debe contener un vocabulario esencial (lo que él llama *Core Vocabulary*), y señala que todas ellas tienen igual importancia, pues constituyen los componentes básicos de la comunicación. Su tipología abarca los nueve grupos siguientes (la traducción y la adaptación de los ejemplos al español es nuestra):

- a) Modalizadores discursivos referidos al grado de certeza, probabilidad y necesidad: *es necesario, es evidente, es aconsejable*.
- b) Verbos soporte (colocaciones) que incluyen tanto los verbos de régimen preposicional como la colocación del verbo con sustantivos. Los ejercicios basados en las colocaciones son muy útiles para este aprendizaje.

- c) Palabras interactivas que muestran las actitudes del hablante y permiten relaciones sociales apropiadas, al tiempo que respetan la interculturalidad.
- d) Marcadores del discurso, cuya importancia radica en el hecho de proveer al emisor de recursos para ejercer el control de la conversación y así poder ejercer su autoridad en la materia: *bien, así, pues, de cualquier forma, ya que, puesto que, lo que deseo demostrar, teniendo en cuenta, considerando, visto que*. Los marcadores utilizados se deben adaptar al nivel del *MCERL* del estudiante, permitiéndole progresar paulatinamente.
- e) Nombres básicos: se trata de una amplia relación de nombres de un significado muy general, tanto concreto como abstracto, del tipo *persona, problema, situación*, etc., junto con nombres de días, meses, colores, etc. y otro tipo de nombres sobre tiempo y lugar.
- f) Deícticos o términos relativos a la deixis personal, espacial y temporal: adverbios de lugar (*aquí, ahí, allí*), de tiempo (*en primer lugar, para empezar, en segundo lugar, antes, después*), locuciones adverbiales, verbos (*ir/venir, llevar/traer*), etc.
- g) Adjetivos básicos con connotaciones positivas o negativas: *inundación con daños leves/graves*. Los ejercicios utilizados con los deícticos son también útiles para utilizar y aprender los adjetivos.
- h) Adverbios básicos de modo, tiempo, locuciones adverbiales, combinaciones sintagmáticas: *llover torrencialmente*. Una relación a partir de columnas entre verbos y nombres con adverbios es una buena manera de aprender estas combinaciones.
- i) Verbos básicos de acción y sucesos para describir actividades cotidianas.

Sobre cómo se debe hacer esta selección léxica, el *MCERL* ofrece algunas opciones (Consejo de Europa 2001: 149):

- a) Elegir palabras y frases clave en áreas temáticas necesarias para la consecución de las tareas comunicativas adecuadas a las necesidades de los alumnos, que encarnan la diferencia cultural y, en su caso, los valores y creencias importantes compartidos por el grupo o los grupos sociales cuya lengua se está enseñando.

- b) Seguir unos principios léxico-estadísticos que seleccionen las palabras más frecuentes en recuentos generales y amplios o las palabras que se utilizan para áreas temáticas delimitadas
- c) Elegir textos (auténticos) hablados y escritos y aprender o enseñar todas las palabras que contienen.
- d) No realizar una planificación previa del desarrollo del vocabulario, pero permitir que se desarrolle orgánicamente en respuesta a la demanda del alumno cuando este se encuentre haciendo tareas comunicativas.

Ahora bien, más que propiamente criterios para quien afronte la tarea de selección, estas orientaciones son pistas a seguir para descubrir los mejores métodos. Unido a esto, está el hecho de que el último punto, el recogido en d), resulta controvertido para gran parte de la comunidad científica que ha revisado el tema más recientemente, que, como ya hemos visto, apuesta por una planificación, aunque sea una planificación tan flexible que permita *a posteriori* que el vocabulario se desarrolle en respuesta a la demanda del alumno cuando este se encuentre haciendo tareas comunicativas. No parece, por lo tanto, la planificación un aspecto contradictorio, sino más bien complementario. Más útil resulta al respecto el punto de vista más práctico del *Plan Curricular del Instituto Cervantes (PCIC, Instituto Cervantes 2006)*, ya que, aunque no ofrece un recuento exhaustivo de unidades léxicas que es preciso conocer, sí establece algunos ejemplos prototípicos que nos ilustran sobre lo que corresponde a cada nivel en sus nociones específicas. Por otra parte, la delimitación de estas nociones específicas, junto a las nociones generales, ya es de gran ayuda a la hora de guiar las tentativas de métodos de selección del vocabulario, en la medida en que debemos tomar como referencia estas nociones para buscar de forma fiable y objetiva aquellas unidades léxicas más útiles, eficaces y representativas de la lengua meta para cada nivel. Y es relevante recordar esto, porque los métodos objetivos de selección del vocabulario que se postulan como útiles para ser aplicados a la docencia de ELE no han tenido en cuenta hasta ahora esta clasificación.

3.1. Críticas a la selección léxica en los materiales de español como lengua extranjera (ELE)

En otro orden de cosas, y a la vista de las críticas que han recibido los materiales de ELE cuando se han estudiado sus contenidos léxicos, parece también evidente que son necesarios otros métodos y criterios¹¹.

Así, Benítez Pérez y Zebrowski (1993) y Benítez Pérez (1994a) pusieron de manifiesto la inadecuación del léxico de los manuales para la enseñanza del español a extranjeros a principios de los 90, al tiempo que señalaban la importancia de contar con una base científica sólida a la hora de seleccionar el vocabulario adecuado para la confección de los manuales de ELE.

El vocabulario utilizado en las clases debe ser un reflejo fiel del vocabulario empleado hoy en día por el hablante hispano. Sabemos de la dificultad de definir quién es el hablante, pues no hablamos exactamente igual en Venezuela y España, o en Madrid y Sevilla, pongamos por caso; pero creemos que los materiales publicados en nuestro país deben, sin olvidar nunca el español de América, ser reflejo de lo que conocemos como nuestra lengua estándar, aquella que nos sirve para comunicarnos todos los hablantes del mundo hispánico. [Benítez Pérez 1994a: 325]

En un primer momento, se centraron en ver cómo se distribuyen las formas léxicas en cuatro de los manuales polacos más utilizados en Polonia en el nivel básico, publicados en las décadas de los 70 y 80. Para ello, extrajeron las lexías y luego las compararon con las formas que aparecen en dos diccionarios de léxico frecuente de español (Juillard & Chang-Rodríguez 1964 y Márquez Villegas 1975). Los resultados indicaron que, por una parte, la cantidad de palabras que se introducen en los manuales no está lejos de las recomendadas por algunos autores para este nivel inicial, pues tres de ellos sobrepasan por poco las mil unidades y un cuarto llega a las 1.700. En cuanto a la repetición, en ninguno de los manuales el número de palabras que aparece por lo menos tres veces sobrepasa el 60% del vocabulario total, lo cual no es muy recomendado

¹¹ De la selección léxica que realizan los manuales y materiales de ELE se han ocupado hasta ahora Benítez Pérez & Zebrowski (1993), Benítez Pérez (1994a, 1994b, 1995, 2009), Sanz Álava (2000), García Marcos & Mateo García (2003), Izquierdo Gil (2005, 2016), Guerra Salas & Gómez Sánchez (2005) y Fernández Leyva (2015).

pues “la utilización reiterada de las mismas lexías para favorecer su aprendizaje” es necesaria (Benítez Pérez & Zebrowski 1993: 225). Por último, analizan la convergencia léxica entre el vocabulario de cada uno de los manuales y el de los dos repertorios de consulta citados. Su estudio indica gran discrepancia, tanto en la selección de las lexías como en el orden de aparición en las listas de frecuencia, pues las 50 palabras de mayor frecuencia coinciden en menos del 40%, aunque con diferencias entre manuales, y se incluyen muchas palabras que son consideradas unidades del léxico superfluo (“aquel que es utilizado por los autores pero, sin embargo, no está presente en ninguno de los diccionarios consultados”), como por ejemplo: *restorán*, *perdiz* (no se actualiza el genérico *pájaro*), *anfitrión*, *atómico* y *guardarropa* (Benítez Pérez & Zebrowski 1993: 227).

Benítez Pérez continúa un año más tarde este tipo de análisis, estudiando siete obras de enseñanza de ELE publicadas en España que gozaban de más popularidad entre el profesorado en ese momento. En este caso realizó un vaciado del léxico y comparó las listas resultantes con las listas que obtuvo en una investigación que, sobre disponibilidad léxica, se realizó en la zona metropolitana de Madrid. Escoge esta zona porque considera que en ese espacio geográfico “se produce una mezcla de hablantes que provienen de lugares muy diversos, por tanto, de hablas distintas, que llegan a configurar ese español estándar que defendemos” (Benítez Pérez 1994a: 325).

Así pues, en esta investigación ya no solo se parte de la frecuencia como referente objetivo para comparar el léxico de los manuales, sino que también tiene en cuenta la disponibilidad. En su estudio concluye que hay un gran número de voces que se incluyen que tienen un alto índice de disponibilidad pero, aun así, se usan otras de bajo índice y otras de índice cero, que son las que se deberían emplear en contextos precisos, pero son de poca productividad en un curso general de ELE (por ejemplo los distintos tipos de trenes, correo, tranvía, cercanías, expreso). Otra crítica que manifiesta incide en las pocas veces que se presentan en las obras las palabras más disponibles, lo contrario de lo que se recomienda, pues cuanto mayor sea la exposición, más probable será el aprendizaje. Además de que es recomendable una repetición gradual, mayor aparición para las palabras más frecuentes y más disponibles, que suelen ser las

más generales y que ofrecen mayores posibilidades de uso para un nivel inicial. La opinión de este autor es clara en este sentido y señala la importancia de contar con una base científica sólida a la hora de seleccionar el vocabulario adecuado en la confección de manuales:

Creemos que ahora que la enseñanza del español para extranjeros atraviesa su mejor momento, hay que conseguir que sea, también, el más fructífero. Para ello debemos crear materiales que no sean solo producto de la moda, sino de la reflexión seria, para lo cual debemos tener en cuenta las investigaciones que se están llevando a cabo, tanto en lingüística teórica como aplicada y que pueden ser aprovechadas por nosotros. [Benítez Pérez 1994a: 332]

Más recientemente, otros trabajos han revisado la selección de materiales léxicos en ELE. García Marcos y Mateo García comparan el vocabulario de algunos manuales de ELE de nivel inicial, intermedio y avanzado con los diccionarios que ofrecen los estudios de léxico disponible de Andalucía. Se centran en el campo “alimentación”, uno de los centros de interés característicos en este tipo de estudios. Trabajan con un solo tema a modo experimental, aunque apuestan por extrapolar los resultados al resto de centros de interés. Según concluyen de su revisión:

El léxico de los manuales reflejaría muy pálidamente la realidad empírica constatada y evaluada a partir de las investigaciones sincrónicas. Visto desde la frialdad y la inflexibilidad de los números, son 300 las apariciones registradas en los manuales, frente a las 520 de los materiales empíricos. [García Marcos & Mateo García 2003: 300]

No obstante, explican que cuando se aprende una lengua extranjera no se puede pensar en abarcar todo el vocabulario, por lo que los datos que aporta la disponibilidad léxica (DL) se deben tener en cuenta a nivel cualitativo y más pormenorizadamente, ya que lo realmente determinante en la confección de los materiales léxicos no es tanto el volumen como la selección por la que se opte. Centrados en el campo de la “alimentación”, estudian el tipo de palabras que se incorporan en cada nivel educativo y el porcentaje de léxico que se amplía, para concluir que no todos los vocablos que se ofrecen en el primer nivel son adecuados (de acuerdo con su frecuencia y disponibilidad) a pesar de que el porcentaje que se enseñe sea alto, a diferencia de lo que ocurre en los otros niveles. También en muchos casos la selección de sinónimos se hace sin tener en

cuenta su mayor disponibilidad ni el contexto de uso, y esto varía enormemente entre manuales, por lo que no cabe duda de que los criterios seguidos son aleatorios: García Marcos y Mateo García lo denominan “distribución irregular del léxico”. Atender a la disponibilidad léxica serviría, a su modo de ver, para unificar criterios y guiar a los autores de materiales con datos empíricos, para evitar que sean tan pocas las palabras que coinciden entre los manuales del mismo nivel analizados.

Como hemos explicado ya, la selección léxica se justifica a causa de la complejidad y variación del conjunto del vocabulario. La reflexión final de los investigadores incide también en este aspecto:

A la hora de diseñar propuestas didácticas, el léxico reclama un tratamiento más minucioso que el de otros planos del lenguaje. Los sistemas fónico y gramatical no pueden motivar excesivas dudas [...]. Por el contrario, sólo conocemos de manera muy aproximada cuáles serían las unidades léxicas en verdad relevantes, además, lo hacemos únicamente a partir de la experiencia como hablante que tiene cada investigador. Para solventar esa limitación será, más que aconsejable, imprescindible acudir a los repertorios léxicos de las investigaciones sincrónicas. De ellos obtendremos una variada y eficazísima información susceptible de ser aprovechada en diversos frentes. [García Marcos & Mateo García 2003: 299]

Apuestan, por tanto, por el uso de los repertorios léxico-estadísticos para elegir el vocabulario que se introduce en los materiales y la enseñanza de ELE.

En relación con la enseñanza-aprendizaje del vocabulario temático, Izquierdo Gil (2004) destaca algunas críticas hacia los manuales de español utilizados en las aulas francesas. Por una parte, los campos semánticos presentados varían de un manual a otro, por lo que el conjunto de alumnos francófonos de nivel elemental no llegarán a adquirir los mismos conocimientos léxicos. Por otra parte, algunos campos semánticos considerados básicos (por ejemplo, el cuerpo humano, la ropa, la familia, etc.) no aparecen en ciertos manuales. Por último, en numerosos casos, los documentos auténticos se explotan para practicar la técnica del comentario oral o para introducir aspectos culturales y gramaticales, y no para enseñar el vocabulario relacionado con un tema. A modo de ejemplo, la autora cita algunos manuales de español usados en Francia, en los cuales no se indican de forma más o menos explícita las principales unidades léxicas que el alumnado deberá aprender en cada campo

semántico presentado. Como consecuencia, los manuales se ven obligados a introducir numerosos documentos auténticos, con el fin de dar cuenta de las diferentes voces relacionadas con cada tema, y el alumnado debe localizar en ellos las unidades léxicas referentes a cada tema tratado. Esto puede dificultar la formación de redes semánticas. Del mismo modo, el vocabulario introducido por los *realia* sólo se practica de forma implícita a partir del comentario oral y/o escrito y mediante diversas preguntas de comprensión textual, ya que se considera que la realización de varios comentarios sobre diversos documentos auténticos que giren en torno a un mismo tema constituye una tarea suficiente para la asimilación del vocabulario. Los ejercicios específicos sobre campos semánticos se hallan, pues, ausentes en dicho enfoque pedagógico (Izquierdo Gil 2004: 399-400).

En su tesina de máster, Madison (2013) trata la cuestión de la selección léxica centrándose también en el análisis de dos manuales actuales y conocidos de nivel inicial¹², para compararlos con los datos ofrecidos por la listas de frecuencias léxicas, para lo cual toma como referencia las mil primeras palabras del diccionario de Davies (2006). Madison adelanta que, según los resultados aportados por Davies y Face (2006), el punto de partida es ya que en los manuales analizados no se presta mucha atención a las frecuencias generales de las palabras en español. Después de realizar su estudio, concluye que de las 1.252 palabras lematizadas que ofrece *Bitácora 1* en su glosario, 459 aparecen entre las mil más frecuentes de Davies y las otras 793 caen fuera. Por otro lado, de las mil más frecuentes de Davies, no se tienen en cuenta en el manual 541. En el caso del *Aula internacional 1* los resultados son que este libro ofrece un glosario de 2.319 palabras lematizadas, de las cuales 1.319 no se corresponden con las palabras más frecuentes en Davies. Resultados parecidos en cuanto a disparidad entre el léxico de los manuales y del diccionario de frecuencias encuentra cuando compara las nociones concretas, viendo qué palabras se incluyen en cada tema y cuáles no (Madison 2013: 41-42).

En un trabajo más reciente, Fernández Leyva lleva a cabo una comparación cuantitativa entre el vocabulario del manual *Español para Todos 1*,

¹² Se trata de los manuales *Aula internacional A1* (2005) y *Bitácora 1* (2011) de Difusión.

que se usará en la Universidad de Oriente, con la variedad de español que se habla en Santiago de Cuba, ya que este texto está destinado para ser empleado en contexto de inmersión y persigue que los estudiantes no hispanohablantes adquieran vocablos representativos de la cultura de esa ciudad. El procedimiento seguido fue el vaciado del libro, el posterior lematizado de las formas, agrupación de los vocablos resultantes en los 16 centros de interés clásicos y, por último, comparación con las listas del léxico disponible de estudiantes preuniversitarios santiagueros. Esta comparación se hizo para evaluar cuáles de los vocablos por centros de interés que propone el manual debían permanecer o excluirse, así como el número de apariciones que se recomienda para cada unidad léxica. La comparación cuantitativa indica que, en general, el número de vocablos por áreas temáticas que se incluye en el manual en relación al número de palabras de mayor disponibilidad es excesivo. Así, la autora, que lleva además a cabo el trabajo de revisión y edición del manual, recomienda una disminución de 228 vocablos, pues hay 9 centros de interés con un número mayor de vocablos de los que se recomiendan en este estudio y apuesta por ceder ese espacio “para regular las ocurrencias de los vocablos según lo indicado por Benítez Pérez, pues se notó una tendencia a emplear muchos vocablos pocas veces” (Fernández Leyva 2015: 10).

3.2. Criterios de selección léxica

Resulta un hecho comúnmente aceptado que el aprendizaje del léxico de una segunda lengua constituye un proceso limitado y esencialmente gradual. En este sentido, hay una notable coincidencia entre la adquisición de la lengua materna y el aprendizaje de un segundo idioma. Un hablante nativo aprende el léxico de su lengua de una manera paulatina y, si bien el progreso más espectacular se produce durante su infancia, en principio es un camino abierto que puede prolongarse a lo largo de toda la vida. Por muy culto que un individuo sea, podemos afirmar que no está capacitado para conocer por sí solo todo el acervo léxico de su idioma. Esto se cumple de manera paradigmática en lenguas de amplia presencia internacional, como es el caso del español, en las que la amplitud de su comunidad vernácula entraña una considerable diversidad

dialectal, sociolingüística y cultural. Para el caso de una segunda lengua, ese carácter limitado y gradual de la adquisición del vocabulario conlleva la necesidad de establecer un orden de prioridades de cara a la planificación curricular de la enseñanza de contenidos léxicos. Esto quiere decir que ha de determinarse, en la medida de lo posible, qué segmento del bagaje general del léxico constituye realmente el objeto de aprendizaje y, dentro de tal segmento, qué unidades han de recibir una atención focalizada en cada uno de los distintos niveles de dominio, desde el inicial hasta el superior (Alvar Ezquerro 2003: 244-5).

Debemos tener en cuenta que el objetivo de nuestra selección para la docencia de ELE será el desarrollo de la competencia léxica de nuestro alumnado y que, como hemos explicado, esta está formada por diferentes dimensiones, por lo que al llevar a cabo nuestra selección tendremos que pensar en los conocimientos gramaticales de las unidades para desarrollar la competencia lingüística, en los diferentes marcadores discursivos para la discursiva, en el vocabulario instrumental para la estratégica, y en las fórmulas convencionales para la sociolingüística. Al mismo tiempo no podemos olvidar que cuando hablamos de selección léxica, hablamos de escoger unidades léxicas y sus diferentes tipos, que también hemos explicado, por lo que tendremos que superar el concepto de palabra como “palabra gráfica” y tener en cuenta colocaciones y compuestos. Estamos de acuerdo con que existen segmentos prefabricados (*chunks*) que nos sirven desde un inicio y que es clave atender al principio de idiomatidad acuñado por Lewis. Lewis se suma a la línea de investigadores que reconocen la indiscutible importancia de las frases léxicas o los segmentos léxicos, y se decanta por el principio de idiomatidad (*idiom principle*), según la dicotomía establecida en 1991 por Sinclair, que también ha influido notablemente en nuestra concepción sobre el aprendizaje y la enseñanza del léxico. El principio de idiomatidad defiende que el léxico de una lengua está formado por un número amplio de bloques semiconstruidos que el hablante tiene memorizados como un todo y puede combinar al hablar. Esta visión implica superar el principio de selección libre (*the open choice principle*), que afirma que el léxico se entiende como un conjunto de palabras individuales que se pueden emplear con mucha libertad y que sólo están limitadas por reglas gramaticales

(es la visión tradicional, que supone entender la lengua como un sistema de reglas sintácticas en el que se encajan ciertas palabras). Esta versión fuerte, que casi destierra la enseñanza de la gramática de las clases, es difícil de aceptar, pero también es cierto que se trata de una cuestión de prioridades y que el enfoque léxico sí enseña gramática, pero a partir de las unidades léxicas. Tampoco podemos olvidar que para la lingüística cognitiva no hay separaciones tajantes entre la gramática y el léxico y es en esa línea en la que creemos que se deben interpretar las palabras de Lewis (*vid. supra* § 1.6 y 2.3.)

También cabe mencionar que, si el tipo de palabras que queremos enseñar es diverso, no vamos a conseguir escogerlas de la misma forma, ni podremos aplicar los mismos métodos de enseñanza-aprendizaje para todas ellas. Pensemos por ejemplo en la clasificación de las palabras en relación con su significado. Suele distinguirse entre palabras “léxicas”, que tienen sentido propio (como los sustantivos o los verbos), y palabras gramaticales o funcionales, que indican relación entre otras dos con contenido léxico (como las preposiciones o conjunciones). Las primeras deberían enseñarse también de forma explícita, mientras que las segundas son tan frecuentes que no es necesario enseñarlas como tales, dado que se enseñan conjuntamente con su función sintáctica. No obstante, es posible que haya aquí un prejuicio acerca de lo que es el significado de una palabra, dado que si tenemos que enseñar el artículo, una de las categorías gramaticales por excelencia, somos capaces de recurrir a instrucciones de significado como las propuestas por la gramática cognitiva. Aunque lo que nos interesa resaltar es que las palabras léxicas deberían enseñarse como lo que son, es decir, nudos en una red léxica, en la que cada nudo es diferente del anterior y del siguiente. Una clasificación que se puede hacer en este sentido es la siguiente:

Palabras contenido (“léxicas” o “llenas”)	Palabras función (“gramaticales” o “vacías”)
Verbos: <i>ir, salir, amar...</i>	Pronombres: <i>yo, tuyo, nos...</i>
Nombres: <i>José, libertad, rama...</i>	Artículos: <i>el, la, unos...</i>
Adjetivos: <i>bello, gracioso, estrecho...</i>	Demostrativos: <i>esta, aquel...</i>
Preposiciones: <i>sobre, contra...</i>	Preposiciones: <i>a, en, sobre...</i>
	Verbos auxiliares
	Conjunciones

CUADRO 4. Clasificación del tipo de unidades léxicas

Podemos pensar en las primeras como los ladrillos o piezas de lego, mientras que las otras serían la masa que las une, pero ambas son necesarias y útiles. Bogaards (1994: 116-9) llevó a cabo un experimento para comprobar qué ocurría quitando unas y otras de un texto. En un primer momento borró de un texto periodístico todas las palabras plenas (con mayor contenido semántico) y dejó solo las que entran en las 500 más frecuentes, con lo que el texto se convertía en incomprensible. Probó luego dejando las 5.000 más frecuentes, con lo que el texto se podía adivinar, aunque era muy pobre a nivel informativo. Realizó a continuación el mismo experimento a la inversa y comprobó que, tras eliminar las palabras que entran en la lista de las mil más frecuentes, lo esencial del mensaje se había transmitido. Y esto es así porque en esa lista de las mil palabras más frecuentes, solo se encuentran las palabras función o gramaticales, así como algunas palabras plenas de significado muy general. Por lo tanto, debemos tener recursos para seleccionar aquellas palabras que son útiles, teniendo en cuenta ambos tipos. Y en esto se estaba pensando cuando se empezó a recurrir a la disponibilidad léxica para seleccionar las unidades poco frecuentes, pero útiles en determinados contextos comunicativos. En cambio, muchas de las palabras frecuentes son muy poco precisas y con escaso valor informativo. A esta conclusión se llegó con los estudios de frecuencia, y de ahí que surgiera precisamente la investigación en disponibilidad léxica en el ámbito francés a mediados del siglo pasado, como una forma de intentar corregir los resultados de los estudios de frecuencia y complementarlos buscando las palabras plenas con otros métodos.

En definitiva, lo único que parece claro en este proceso de selección es su complejidad, lo que apunta a que, más allá de aplicar un solo criterio o método,

vamos a tener que conocer cuáles están en nuestras manos y aplicarlos en función de nuestros objetivos concretos, de forma complementaria. Por otro lado, como también afirmó Bogaards (1994), la utilidad de las palabras no se produce *per se*, sino que las unidades son útiles por su adecuación al objetivo que queremos conseguir, así que, por mucho que intentemos planificar de antemano, debemos adaptarnos a cada contexto y a las necesidades de nuestro alumnado. En conclusión, las unidades léxicas que son más útiles para nuestro alumnado serán aquellas que les ayuden en mayor medida a desarrollar las actividades de deseen hacer con la lengua que están aprendiendo, por lo que en esa línea debe situarse nuestra primera meta. El resto consistirá en aplicar nuestros conocimientos sobre selección léxica para conseguir hacer la mejor elección en cada caso y adecuarnos a los objetivos de nuestros cursos.

3.2.1. ¿Qué criterios deben guiar nuestra selección?

Toda selección es el resultado de aplicar una serie de criterios. La mayor parte de la bibliografía que trata la selección léxica toma el criterio de utilidad como el fundamental, es decir, debe seleccionarse para cada etapa de aprendizaje el vocabulario que sea útil al alumnado. Algunos otros criterios que también suelen nombrarse son la eficacia, la rentabilidad, la productividad o la facilidad en el aprendizaje. Sin embargo, todavía tenemos que resolver cómo medir la utilidad de una palabra o cuáles son las palabras más útiles de una lengua, pues también aquí ha habido y hay diversas interpretaciones. Podemos buscar criterios objetivos para seleccionar las palabras útiles y, aun así, la noción de utilidad todavía estaría por definir, ya que las palabras no son útiles de forma general e indeterminada, sino que tendrán mayor o menor utilidad dependiendo de su adecuación a nuestro objetivo: “est utile ce qui sert à atteindre un but qu’on s’est fixé” (Bogaards 1994: 111).

Como explica Alonso Ramos, para un médico profesional serán más útiles palabras como *anestesia* o *sangrar*, que no se encuentran entre las cinco mil más frecuentes. Y si bien es posible establecer frecuencias dentro de campos específicos, por ejemplo en el caso de las lenguas para fines específicos, sigue siendo igualmente dudoso que el criterio más importante para un médico sea

aprender los términos de su especialidad por su orden de frecuencias. En cambio, alguien que estudie español en un contexto de inmersión buscará antes las palabras de la actividad cotidiana, entre las que también hay palabras poco frecuentes, como *empadronamiento* o *tarjeta sanitaria* (Alonso Ramos 2012: 22).

Por nuestra parte, vamos a ofrecer un repaso de los diferentes criterios de ordenamiento y selección que se han venido utilizando en la enseñanza del léxico como segunda lengua o como lengua extranjera con la finalidad de comprobar su conveniencia para cada caso.

En su manual sobre enseñanza del vocabulario, Nation estudia la cuestión de las palabras más útiles (Nation 1990). Tras señalar las deficiencias de los recuentos basados solo en la frecuencia absoluta, este autor apuesta por tener en cuenta otros parámetros, como el *índice o rango de estabilidad —índice de dispersión* para Juilland y Chang-Rodríguez (1964) o lo que en otros autores es el *índice de uso*—, con el que no medimos únicamente la frecuencia sino también el uso. Así, Nation apostó por seguir en el aprendizaje de una lengua el siguiente orden:

1. Vocabulario frecuente (teniendo en cuenta su dispersión),
2. Vocabulario académico,
3. Vocabulario técnico (especializado),
4. Vocabulario poco frecuente.

Las cifras sobre la cantidad de palabras que se encuadran en cada uno de estos tipos de vocabulario parten de los cálculos léxico-métricos llevados a cabo por Nation (1990), que se han manejado frecuentemente en los estudios sobre la enseñanza del vocabulario de la lengua inglesa. El origen fue la realización de un corpus del inglés académico escrito. El primer estadio, el vocabulario frecuente, consta de 2.000 vocablos de aparición general (independientemente de las variables diafásicas o diastráticas), y acumula un 87% de cobertura textual, es decir, este conjunto de voces formaría parte del vocabulario fundamental de la lengua. El siguiente bloque está formado por 800 vocablos, que suponen un 8%, y en él figuran unidades léxicas que, aunque no constituyan un segmento nuclear en el acervo léxico de la lengua, sí gozan de una presencia significativa en un amplio espectro de obras académicas y de divulgación. Es, por lo tanto, un

vocabulario marcado diafásicamente como característico del registro formal escrito. En cuanto al técnico o especializado, implica a un total de 2.000 vocablos, con una incidencia del 3% en el total del corpus. Aquí se encuentran las unidades léxicas marcadas diatécnicamente como testimonio de las nomenclaturas y terminologías propias de cada especialidad académica. El último nivel engloba a más de 120.000 términos, que apenas si alcanzan una presencia del 2% en el corpus de referencia. Es un vocabulario marcado diastráticamente, que sólo estaría al alcance de la población con mayor formación cultural o académica. Como vemos, para este autor la frecuencia viene acompañada de la *cobertura*, otro criterio a tener en cuenta cuando hablamos de selección léxica. A estos dos criterios, este autor suma las *necesidades lingüísticas* de nuestro grupo de discentes, la *accesibilidad y familiaridad* de las palabras, así como su *regularidad y facilidad de aprendizaje*.

Bailly (1997), por su parte, se basa en dos criterios básicos para llevar a cabo la selección léxica, que son la utilidad y el grado de *facilidad del aprendizaje*. En cuanto al primero, para ella las palabras útiles son las que abarcan el vocabulario frecuente, el disponible, el instrumental y del aula.

Muy cerca de esta propuesta está la de Izquierdo Gil (2004). Para ella, los criterios básicos son el léxico frecuente, el léxico disponible, vocabulario instrumental y de aula y, en el caso de dos lenguas próximas como las lenguas romances, las cuestiones derivadas de un análisis contrastivo (diferencias, semejanzas, correspondencias, interferencias, cognados, parógrafos, homógrafos, etc.). Especifica la facilidad de aprendizaje en dos tipos de palabras: las palabras semejantes en la lengua materna (los cognados; esta autora le concede mucha importancia al análisis contrastivo) y las palabras con alto nivel de familiaridad.

Por estas mismas fechas, Salazar García (2004) en un artículo que trata directamente estas cuestiones analiza los que él cree que son los métodos y criterios fundamentales para la selección objetiva del vocabulario: la frecuencia léxica (léxico frecuente y léxico usual) y la disponibilidad léxica. Este autor expone las que considera ventajas y desventajas de ambos criterios, y apuesta por sumar a ellos otro método subjetivo que dé sentido a nuestros objetivos docentes: los intereses de nuestro alumnado. Según esto, no podemos partir de un listado cerrado, concebido como único e invariable para todos los contextos

de aprendizaje. Este debe ser, pues, el criterio que nos guía para seleccionar los contenidos léxicos de nuestras aulas, lo cual supone apostar por una programación abierta. Para este autor, será después de haber analizado los intereses del grupo cuando fijemos nuestros contenidos léxicos, y no antes, y es en este momento cuando aplicaremos los métodos objetivos para conseguirlo. Se parte, pues, del conocimiento de ellos por parte de los docentes. Dentro de los métodos objetivos, y a pesar de las limitaciones, el criterio de frecuencia para crear la secuencia de nuestro currículum es el que le parece más operativo.

Más recientemente, el profesor Bartol Hernández (2010) expone las virtudes de la disponibilidad léxica para seleccionar el vocabulario de las aulas de ELE. En su artículo analiza las dificultades a tener en cuenta en esta labor de selección, así como los diferentes criterios que se han esgrimido para llevarla a cabo. Entre ellos cita la cobertura —el proyecto BASIC por ejemplo— y otro criterio que está directamente relacionado con ella: la frecuencia de uso, a la que se accede a través de los léxicos de frecuencia y los léxicos básicos. No obstante, propone la disponibilidad léxica para corregir las deficiencias que la frecuencia pueda tener, ya que no todas las palabras frecuentes son las más útiles, sino que hay otras poco frecuentes pero muy útiles en determinados contextos. A estas sí se llega con los estudios de léxico disponible, como ya había expuesto López Morales (1999b) al explicar que hay términos muy comunes que no son frecuentes. A partir de ejemplos del *Léxico básico del español de Puerto Rico*, un estudio basado en la frecuencia, Bartol Hernández expone que no contempla palabras tan conocidas y útiles como “cuchara”, “bicicleta” o “lápiz”. Ofrece también el ejemplo del *Léxico de frecuencia del español hablado en la ciudad de Málaga*, en el que no aparecen “cuchillo”, ni “tenedor” (Bartol Hernández 2010: 88). Todas estas palabras las obtendremos en un listado de léxico disponible, y parece claro que también deben formar parte del vocabulario fundamental del español. Debido a esto, Bartol Hernández incide sobre todo en lo que la disponibilidad ofrece al proceso de selección léxica, observando lo que los documentos oficiales nos indican (el *Nivel Umbral*, el *Marco común europeo de referencia para las lenguas* y el *Plan Curricular del Instituto Cervantes*) y comparándolos con los datos que obtenemos mediante la disponibilidad. Para llevarlo a cabo, Bartol Hernández establece una comparación del vocabulario que

ofrece el Consejo de Europa para el *Nivel Umbral* y el que resulta de unas encuestas piloto realizadas por él mismo. Coteja, pues, la validez de sus encuestas, para ver si sus resultados se ajustan o no a los textos oficiales. Trabajó con estudiantes de segundo de bachillerato de dos institutos salmantinos y con seis nuevos centros de interés, ya que hay temas y subtemas del Consejo de Europa que no aparecen entre los 16 tradicionales que se suelen usar en estas investigaciones. En concreto, añadió los temas relativos a la vida cotidiana, las relaciones con otras personas o la salud, como los siguientes: *El mundo laboral, Relaciones familiares y sociales, La salud, La educación, El tiempo y Creencias*. Ofrece los resultados en una serie de cuadros en los que solo aparecen las 20 primeras palabras, pero aún así confirman la validez de la disponibilidad por la similitud entre los términos captados y los que aparecen en el *Nivel Umbral*. El problema aquí es que no conseguimos cubrir todas las nociones del *MCERL* ni las del *PCIC* con los centros de interés que se barajan hasta ahora, y si estos se amplían demasiado, las encuestas pueden dejar de ser viables.

Un trabajo centrado en establecer los criterios de selección de voces para la enseñanza del español como L2 en contextos escolares, tanto en el aula de ELE como en las otras disciplinas académicas, es el propuesto por López Pérez (2013: 160-165). Esta autora establece tres grupos diferenciados de criterios:

- a. dependientes de la situación de enseñanza,
- b. basados en la dificultad de las voces y
- c. relacionados con la descripción y el uso de la lengua.

Los del primer tipo, son criterios que están directamente relacionados con los objetivos del curso, los que Bogaards denominaba “criterios pedagógicos”. La elección de unas u otras palabras depende de cuestiones como los temas y las actuaciones comunicativas, el tipo de textos (la lengua hablada y los textos escritos presentan características lingüístico-discursivas diferentes), el registro de los textos usados, el tipo de tarea que el alumnado va a hacer, el tipo de tarea que va a llevar a cabo desde el punto de vista del uso del lenguaje (productiva o receptiva), el tiempo disponible en el curso y la metodología empleada, la necesidad de introducir el vocabulario instrumental y el

vocabulario metalingüístico¹³. Otro grupo de factores también relacionados con la situación de enseñanza son la oportunidad de lo inmediato y el componente cultural. Con la primera nos referimos a las “palabras que por el hecho de corresponder a objetos o acciones relacionados con la situación inmediata del aula pueden ser subjetivamente tanto o más necesarias que otras que figuran en las listas de frecuencia” (Becerra Hiraldo 1994: 313). En este sentido, se pueden incluir en la programación áreas temáticas, como por ejemplo las dependencias del centro u objetos de la clase. El componente cultural se refiere a la conveniencia de prestar atención al vocabulario como transmisor de los valores culturales de la comunidad de la lengua meta.

En el segundo grupo de criterios, los basados en la dificultad de las voces para los estudiantes de la L2 o LE, los factores a tener en cuenta derivan de la relación con la L1 de los estudiantes (*crosslinguistic factors*) y otros factores intrínsecos a las propias palabras (*intralexical factors*) y que pueden ser conceptuales, semánticos o formales. Laufer (1997a y 1997b) ofrece una clasificación en tres grupos de factores, según puedan facilitar o dificultar el aprendizaje de una palabra:

¹³ Al vocabulario metalingüístico pertenecen “todas las estructuras o palabras que utilizamos para describir y hablar sobre la lengua”, según Cervero & Pichardo (2000: 41).

Factores que facilitan el aprendizaje	Factores que dificultan el aprendizaje	Factores con un efecto no probado
<ul style="list-style-type: none"> • Fonemas semejantes • Combinaciones de letras similares • Acento de intensidad siempre en la misma sílaba • Coherencia en la relación entre sonido y grafía • Regularidad en la inflexión • Regularidad en la derivación • Transparencia de las partes de la palabra (<i>pre+ver => “ver con antelación”</i>) • Palabras usadas generalmente con registro neutro • Una forma y un significado 	<ul style="list-style-type: none"> • Presencia de fonemas extranjeros • Combinaciones de letras no existentes en la L1 • Acento de intensidad variable • Incongruencia entre sonidos y grafías • Complejidad en la inflexión • Complejidad en la derivación • Transparencia engañosa • Similitud entre las formas de las palabras (<i>casar/cazar</i>) • Palabras específicas con limitaciones de registro • Carácter idiomático • Una forma con muchos significados 	<ul style="list-style-type: none"> • Longitud de la palabra • Categoría gramatical • Palabras abstractas frente a palabras concretas

TABLA 2. Clasificación de los factores que pueden dificultar o facilitar el aprendizaje de una palabra (Laufer 1997a y 1997b *apud* López Pérez 2013)

El tercer tipo de criterios son los relacionados con la descripción y el uso de la lengua, denominados por Bogaards (1994) criterios “lingüísticos”, que derivan de las características lingüísticas de las palabras y de las funciones que estas realizan en la lengua. En este grupo está la frecuencia de las unidades léxicas en los textos, que junto con otros conceptos también estadísticos, como la dispersión y alcance, ha servido tradicionalmente para la planificación del léxico. Así lo afirma el *MCERL* al indicar que se pueden “seguir unos principios léxico-estadísticos que seleccionen las palabras más frecuentes en los recuentos generales y amplios o las palabras que se utilizan para áreas temáticas delimitadas” (Consejo de Europa 2001: 145). López Pérez (2013) se basa en trabajos anteriores¹⁴ sobre frecuencia para establecer una clasificación entre los

¹⁴ Alcaraz Varó (2002), Baker (1988), Dudley-Evans & St. John (1998), López Pérez (2005), Chung & Nation (2003).

tipos de vocabularios y sus características, que podemos obtener mediante la ordenación de palabras por su frecuencia. La tabla que ofrece es la siguiente:

Tipos de vocabulario	Rasgos característicos
Vocabulario técnico	<ul style="list-style-type: none"> • Son palabras relacionadas con áreas del conocimiento. • Se encuentran con una frecuencia alta en textos especializados. • Tienden a ser unívocas.
Vocabulario subtécnico o semitécnico	<ul style="list-style-type: none"> • Son palabras de la lengua general que adquieren sentidos especializados dentro de las llamadas lenguas de especialidad para designar conceptos propios de los campos del saber. • Aparecen en textos especializados. • Son polisémicas.
Vocabulario académico	<ul style="list-style-type: none"> • Expresan procesos mentales propios del aprendizaje y de las tareas con él relacionadas. • Pertenecen a la lengua general, en un registro formal. • Aparecen con una frecuencia alta en los textos generados en ámbitos académicos.
Vocabulario temático	<ul style="list-style-type: none"> • Son palabras clave que tienen relación directa con el tema tratado en el texto. • Son bastante frecuentes en el texto sobre un tema, pero poco frecuentes en recuentos de lengua general.
Vocabulario básico	<ul style="list-style-type: none"> • Son voces de uso común utilizadas en situaciones de comunicación, orales y escritas, sobre temas cotidianos. • No tienen vinculación a un área temática concreta. • Alta frecuencia en todo tipo de discursos.

TABLA 3. Tipos de vocabulario según Bogaards (1994 *apud* López Pérez 2013).

Otros criterios son subjetivos, como la familiaridad, la disponibilidad y lo que Bogaards (1994) denomina “estimaciones subjetivas de frecuencia”. Contamos también con el vocabulario disponible, que han aprovechado manuales como los de Candela y Benegas (2000), Muñoz (2005) o Villalba y Hernández (2004), dirigidos a la iniciación en el español como L2 en la enseñanza obligatoria. No obstante, no ofrecen información sobre los criterios con los que han seleccionado las voces que presentan. Por último, contamos con

el vocabulario fundamental, que une el léxico frecuente y el disponible, por lo que refleja el vocabulario activo de los hablantes y no el receptivo, que es el más patente en los recuentos basados exclusivamente en la frecuencia. Aunque de este vocabulario fundamental hablaremos en profundidad más adelante (*vid. infra* Capítulos 7 y 8).

3.3. Métodos de selección léxica

En el ámbito del español, la literatura que trata el tema de la selección léxica sigue la clasificación hecha por Benítez Pérez para explicar los diferentes métodos de selección del vocabulario. Este autor, que a su vez sigue a De Grève y Van Passel (1971), ha desarrollado su esquema de selección en diferentes publicaciones que venimos citando, pero nos centraremos ahora en la más reciente de ellas para exponer sus teorías al respecto. En ella establece tres formas de limitar las unidades llevadas al aula (Benítez Pérez 2009: 10-11):

1. limitación subjetiva,
2. limitación estrictamente objetiva y
3. limitación objetiva corregida.

En la limitación subjetiva es el docente o autor de materiales quien establece las unidades que introduce, basándose en su opinión, en la intuición o en el azar. Por ello, en muchas ocasiones se pueden encontrar palabras de dudosa utilidad y uso o poco adecuadas al nivel de lengua, como el propio Benítez ha manifestado en diferentes estudios al respecto (Benítez Pérez 1992, 1994a, 1994b). De ahí que se lamenta de que este haya sido el método más usado en el ámbito hispano hasta ahora, frente a los otros dos.

Frente al primero, el método objetivo recurre a los diccionarios de frecuencia para seleccionar aquellas voces que resulten más frecuentes, tanto del vocabulario activo como del pasivo. Esto posibilita que el docente conozca las palabras que puede emplear en un mayor número de circunstancias diferentes. Para Benítez Pérez, este sería, pues, el vocabulario más útil para la enseñanza de ELE. La limitación objetiva corregida, por su parte, no se basa sólo en la frecuencia, sino que también tiene en cuenta la eficacia y productividad de las palabras. Benítez Pérez explica que las primeras, las eficaces, son las que tienen

más posibilidad de utilizarse en muchos contextos y las segundas, las productivas, son las que pueden dar lugar a otras al añadirles formantes. Varela (2003) cita, además de los criterios anteriores, el de la universalidad, y comenta que en el caso del español supone incluir las voces que son utilizadas en todos los países hispanos. En sus explicaciones, Benítez Pérez incide en la necesidad de graduar el vocabulario que utilizamos en las clases y que ese vocabulario sea real. Para ello, defiende el uso de métodos léxico-estadísticos de los que nos nutre la lingüística aplicada, y selecciona los listados de frecuencia y de disponibilidad léxica como modo de escoger las palabras más usuales y más reales, sobre todo para los primeros niveles. Ambos listados permiten llevar a cabo el estudio ordenado del vocabulario en función de su frecuencia de aparición y su utilidad. Además de estos tres métodos básicos de selección (frecuencia, eficacia y productividad), apuesta por tener en cuenta otros factores, como reforzar más las formas léxicas que hacen alusión a realidades muy habituales en la cultura de la lengua meta y poco conocidas o totalmente desconocidas en la cultura de la lengua de partida, así como la motivación y finalidad por la que el alumnado estudia la lengua, meramente instrumental o integradora.

Esta clasificación de métodos fundamentada en tres pilares básicos será avalada por la bibliografía más reciente. Podemos verla reflejada, por ejemplo, en textos de Bartol Hernández (2002: 63), reducida a dos vías posteriormente (2010: 89-90): la subjetiva, la más utilizada durante mucho tiempo pero que puede derivar en diferencias notables entre los recuentos y la objetiva, basada en corpus a los que se aplican métodos estadísticos de análisis, favorecida por los avances informáticos de las últimas décadas. Sobre los métodos de selección objetiva de contenidos léxicos para la enseñanza de ELE ha hecho también su aportación Salazar García (2004). Este autor evalúa de forma crítica la efectividad de cada recurso estadístico y empírico, y llega a la conclusión de que la combinación de todos ellos será la mejor opción, pues todos tienen sus limitaciones. Entre los tres parámetros que establece están también la frecuencia y la disponibilidad, a los que suma el uso, es decir, establece las diferencias y semejanzas de lo que para él es el léxico frecuente, el léxico disponible y el léxico usual.

3.4. Métodos objetivos de selección léxica

Así pues, actualmente son los datos empíricos que nos aporta la léxico-estadística los que son apoyados como más fiables y eficaces por parte de los autores que tratan el tema. Tomemos como muestra las palabras que Bartol Hernández ofrece sobre dos de ellos:

Hoy día crece la idea de que la lingüística científica ha de ser una lingüística del corpus, es decir, que ha de incorporar un conjunto amplio y definido de materiales que le proporcione datos fiables. [Briz 2005: 7 *apud* Bartol Hernández 2010: 90]

Hoy día sería poco creíble pretender dar validez a una investigación, ya sea lexicográfica, ya sea gramatical, que no venga avalada por los datos reales e incontestables de un corpus. [Leech 1991 *apud* Bartol Hernández 2010: 90]

Tradicionalmente, los métodos objetivos para seleccionar datos léxicos a los que se ha recurrido, tanto para la elaboración de diccionarios como para el aprendizaje de lenguas dentro y fuera del ámbito hispano, se han basado en el análisis de frecuencias. Como veremos, cada uno de estos caminos que señalamos a continuación ofrece sus ventajas y sus inconvenientes. Para saber cuál es la frecuencia de las palabras, en la actualidad disponemos de tres vías: los diccionarios de frecuencia, los vocabularios básicos y la información que nos proporcionan los corpus.

Entendemos que el léxico frecuente es el conjunto de palabras que muestra un mayor grado de presencia en los textos, orales o escritos, emitidos en un idioma. Obviamente, la pretensión no es la de analizar la totalidad de enunciados de una lengua viva, sino una muestra que se estime suficientemente representativa, en forma de corpus textual. De ahí que, en la práctica, el concepto de frecuencia léxica remita esencialmente a la incidencia de una determinada palabra dentro de un corpus determinado (Richards, Platt & Platt 1992: 177). Estos datos quedan posteriormente reflejados en herramientas tales como los diccionarios de frecuencias, que ofrecen un cálculo cuantitativo de esta vertiente del vocabulario, sobre una base empírica y no meramente impresionista.

Ante la enormidad del léxico de una lengua, quienes se dedican a

confeccionar diccionarios, a elaborar materiales didácticos o a enseñar lenguas, se encuentran un dilema abordable gracias a que la frecuencia léxica ofrece objetividad y una base científica, que sirve para contrarrestar la subjetividad en la selección a la que nos hemos venido refiriendo. Diversos autores defienden que se deben, pues, enseñar antes las palabras más frecuentes porque aparecen en todos los textos y son de alto rendimiento. McEnery y Rayson, al prologar el diccionario de Davies (2006), argumentan que parece razonable priorizar el aprendizaje de palabras que tienen más probabilidades de ser oídas o usadas a menudo. Almela *et al.* (2005) apelan a la rentabilidad comunicativa de la frecuencia, ya que entienden que es mucho más rentable aprender una palabra muy usada y altamente polisémica que otra poco usada.

Por otro lado, el concepto de frecuencia léxica ha estado siempre vinculado a los diccionarios y a la enseñanza de lenguas. Los estudios de frecuencias léxicas se originaron en relación con el aprendizaje del vocabulario y siguen manteniendo esa relación hoy en día. Se puede observar cómo uno de los más recientes, el diccionario de frecuencias del español publicado por Davies (2006), deja clara esta orientación con el propio subtítulo: “core vocabulary for learners”. Muchos diccionarios de frecuencias léxicas son concebidos con una finalidad esencialmente didáctica: “comprobar la frecuencia de las palabras para poder usar en obras pedagógicas las palabras más frecuentes” (Haensch & Omeñaca 2004: 154). Con todo, la información sobre la frecuencia de las palabras puede ser útil para la lexicografía, puesto que los recuentos pueden servir de criterio para incluir o excluir material en su macroestructura o en su microestructura y, asimismo, la frecuencia de las acepciones puede servir para ordenarlas dentro de la entrada. La relevancia de la información de frecuencia en el ámbito del aprendizaje de lenguas ha sido señalada reiteradamente, aludiendo principalmente a conceptos como *utilidad, cobertura, rentabilidad comunicativa* (Alonso Ramos 2012: 21). En estos casos, la utilidad de las voces se mide por su frecuencia, ya que equivale a rentabilidad comunicativa, y esta se mide por el porcentaje de cobertura textual que ofrecen.

La aparición de los diccionarios de frecuencias impulsó los estudios para encontrar aplicación del vocabulario básico en la enseñanza y sobre todo en las clases de lengua extranjera. Entre las obras pioneras de este tipo está, para el

alemán, el diccionario de frecuencias de Kaeding, publicado en 1898, que sirvió de base para la elaboración del primer vocabulario básico alemán: el de Morgan, de 1928. En lo que al español respecta, los orígenes de los vocabularios básicos se remontan a la aparición en 1927 de la obra de Buchanan, publicada en Toronto, con la que el autor se propone prestar especial atención a la lengua hablada. Estos vocabularios deben mucho a los estudios de lexicometría y estadística lingüística y pretenden facilitar la difusión y el aprendizaje de una lengua en el menor tiempo posible, fijar el número mínimo de palabras que se necesitan para una comunicación básica. Por eso, entre 1950 y 1970 tuvo un gran auge la creación de glosarios mínimos, el más conocido de los cuales es el proyecto BASIC, que además fue el primer intento importante de establecer el vocabulario básico de la lengua inglesa. Lo constituye el llamado *Inglés básico*, que Ogden (www.BasicEnglish.org¹⁵) formuló a partir de 1930. Se trataba de seleccionar el mínimo número de palabras que permitiera expresarse al alumno en un espectro lo más amplio posible de situaciones. Ogden partió del hecho de que entre las palabras que hay en las definiciones de los diccionarios hay algunas que se repiten mucho. Pensó, así, que se podría formar un vocabulario muy reducido con una frecuencia de uso muy alta y reunió en una lista las 850 palabras que se consideraban más usuales. Con la Segunda Guerra Mundial este vocabulario se utilizó como método para la enseñanza rápida del inglés en más de 20 países de todo el mundo. En la introducción a esta obra, Ogden explica el motivo de optar por este método de selección léxica:

Al cabo de seis o más años de estudio del idioma inglés, los escolares hispanoamericanos llegan al desalentador resultado de no saber casi nada. Han asimilado un cúmulo de reglas, mucha gramática, literatura, etc. pero prácticamente son incapaces de sostener la más simple de las conversaciones en inglés [...]. [Ogden 1930]

Como vemos, la solución sería estructurar la enseñanza del inglés a partir del vocabulario, y no de la gramática, y escoger un número mínimo de unidades

¹⁵ Se puede consultar una bibliografía completa sobre el Inglés Básico en la página web creada para su difusión: <http://www.basic-english.org/learn/basicbibio.html>.

El sitio <http://www.basic-english.org/> pertenece al *Basic English Institute*, que se presenta como una institución “Dedicated to the expansion of Ogden’s Basic English”.

útiles y eficaces. Se esgrime como causa del fracaso en el aprendizaje del inglés los “innecesariamente complicados textos” que se usaban, que incluían palabras poco necesarias. En cambio, como otros muchos métodos que luego han salido al mercado e incluso recientemente vemos anunciados en la prensa (por ejemplo el conocido *Método Vaughan*), ellos prometían enseñar un número muy reducido de palabras “necesarias” y unas cuantas reglas gramaticales sencillas en pocas semanas, con tal eficacia que conseguirá el alumno un fácil y rápido dominio de esta lengua, que le permitirá viajar a cualquier país de habla inglesa y comunicarse con un nativo cuando las domine con un inglés corriente. Indican también que las 850 palabras han sido seleccionadas cuidadosa y científicamente durante más de 10 años de profundos estudios que fueron necesarios para que un grupo de catedráticos ingleses, encabezados por el famoso profesor Ogden, lograra compilar la selección de las palabras básicas. Y prometen, además, que equivalen en eficiencia a más de 20.000 palabras y que, aunque resulte difícil comprender cómo 850 palabras pueden equivaler a más de 20.000, tal afirmación quedaría prácticamente comprobada a medida que el estudiante fuera avanzando en el estudio de este libro. Otro de los problemas está en el número de palabras seleccionadas, pues la explicación no es muy satisfactoria ni extensa. Se limitan a exponer que se debe a que el Inglés Básico es absolutamente práctico, por lo que no hay nada en él que falte o sobre. Otra razón que esgrimen para demostrar que es el suyo un sistema inteligente y sencillo es la eliminación del complicado sistema verbal inglés y del enorme número de verbos regulares e irregulares, reemplazándolo solo por 16 formas verbales.

En realidad, el criterio que sostiene este método de selección ha recibido el nombre de *cobertura*, entendida esta como el espacio semántico y situacional que es capaz de cubrir una palabra o el porcentaje de palabras que componen un texto (oral o escrito). Es decir, las palabras más frecuentes y básicas son útiles en cuanto que son rentables a nivel comunicativo. La cobertura textual nos lleva a la pregunta de cuánto vocabulario seleccionar. Es una cuestión todavía abierta, pero lo que sí sabemos es que la cantidad de palabras que conoce una persona tiene implicaciones en su competencia comunicativa y en su capacidad de

comprensión de los textos. La respuesta a la pregunta de cuánto vocabulario hay que conocer para aprender una lengua no da como resultado una cifra demasiado elevada. Y es una cantidad que se puede graduar en su adquisición, por etapas educativas. Hay diferentes estudios al respecto¹⁶, en los que se trata la cuestión de qué porcentaje de comprensión alcanzamos con diferentes cantidades de palabras. Unido a esto, se ha tratado la existencia de un umbral léxico, una cantidad mínima de unidades léxicas conocidas, a partir de la cual somos capaces de comprender un texto oral o escrito, esto es, qué número o qué porcentaje de palabras debemos conocer para llegar a un nivel de comprensión aceptable u óptimo. Aquí la respuesta suele establecer que en un texto se deba alcanzar el 95% para un nivel aceptable y un 98% para un nivel óptimo (Cartaya Febres 2011). Estos serían, según los autores, los puntos a partir de los que un individuo es capaz de inferir las palabras desconocidas, lo que se conoce como *inferencias léxicas contextuales*.

Se ha establecido la existencia de este umbral léxico, o conocimiento mínimo de vocabulario que posibilita la comprensión de un texto auténtico, en las 3.000 familias¹⁷ de palabras más frecuentes. Otros, en cambio, hablan de 2.000. Pero sí sabemos que este umbral es el punto a partir del cual el alumnado podrá aplicar las estrategias de inferencia (Izquierdo Gil 2004: 64). Ya hemos expuesto valoraciones y datos sobre el conocimiento léxico de los hablantes y sobre el número de términos que compone el vocabulario usual de un individuo (*vid. supra* § 2.3.). Hemos comentado también algunas pinceladas sobre la cobertura textual (*vid. supra* § 2.2.). Nos detendremos ahora en este concepto,

¹⁶ Laufer 1989, 1992, 1997a, 1997b; Laufer & Sim 1985; Nation 1990, 2001, 2006; Hu & Nation 2000; Laufer & Ravenhorst-Kalovsky 2010; Cartaya Febres 2011.

¹⁷ Algunas investigaciones, como por ejemplo las de Nation, trabajan tomando como base las “familias de palabras”, un concepto muy similar al de familias léxicas. De ellas parten para ofrecer sus cálculos sobre el léxico frecuente, de ahí que lo reproduzcamos de esta manera. Aunque reconocemos la utilidad de su uso para el aprendizaje de una lengua, los léxicos básicos, los diccionarios disponibles y los documentos oficiales del español no hablan de “familias de palabras” sino de palabras. En el ámbito hispano el concepto de unidad léxica o de palabra es el concepto de referencia habitual, por lo que nosotros también seguimos esta tendencia. No obstante, lo ideal sería aprovechar la utilidad de las familias léxicas a nivel práctico, cuando van apareciendo las palabras base o las derivadas en el aula. De esta forma se ayuda a crear asociaciones léxicas y a solidificar contenidos. (*vid. infra* § 4.1.2.).

aunque lo retomaremos de nuevo más adelante para hacer otras consideraciones relacionadas con él (*vid. infra* § 4.1.1.).

Se han ofrecido diferentes datos sobre el número de palabras y sus porcentajes, como los indicados para el español por Patterson y Urrutibéheity (1975), quienes valoran que nuestras 100 palabras más frecuentes constituyen más del 30% del material léxico de cualquier texto, que las 1.000 primeras alcanzan el 50% de todo texto y que con las 5.000 se sobrepasa el 90%. Si bien estas cifras no coinciden exactamente con las que nos ofrecen otros autores. Difieren sustancialmente, por ejemplo, de lo expuesto por Juilland y Chang-Rodríguez, en cuyo diccionario de frecuencias sólo las 10 primeras formas son ya el 36.55% de la frecuencia total, las 100 primeras el 66.20%, y las 1.000 primeras el 85.50% del total (Juilland y Chang-Rodríguez 1964). Estas apreciaciones parecen estar corroboradas por estudios estadísticos más recientes. En las listas de frecuencia elaboradas por Ávila Muñoz, las primeras 5.000 palabras del corpus cubren el 99,7% de un texto dado. Pero debemos ser conscientes de que estos números están aplicados al léxico de frecuencia del español hablado en la ciudad de Málaga (Ávila Muñoz 1999). Evidentemente, el número de palabras que necesitamos varía en la comunicación oral, que suele manejar un vocabulario más reducido y simple. Sin embargo, son datos que corroboran que 5.000 palabras son suficientes para las necesidades básicas de un hablante y nos pueden ayudar a entender cualquier conversación. Pese a la diferencia de registro, los números hablan por sí solos y nos muestran aproximadamente qué cantidad de palabras habremos de enseñar según los fines perseguidos, pues entre estas 1.000/5.000 y las 100.000 de algunos diccionarios generales de la lengua hay un enorme espacio.

3.4.1. La frecuencia léxica

Los trabajos de frecuencias léxicas han ayudado a considerar si en el aprendizaje de una lengua extranjera los manuales y otros materiales presentan un vocabulario demasiado amplio y si el que incluyen es el más productivo y útil. Nos podemos preguntar si las palabras más frecuentes de una lengua son más

importantes a la hora del aprendizaje, si son la base adecuada para los principiantes y cuáles son los argumentos a favor y en contra.

El análisis de frecuencias se ha aplicado a diferentes publicaciones que suelen llevar en sus títulos las expresiones: “vocabulario”, “recuento”, “léxico frecuente”, “léxico básico”, “diccionario de frecuencias”, etc. Las primeras listas de frecuencia publicadas en inglés fueron creadas manualmente y, aunque el corpus era pequeño, lo importante es que era una guía léxica valiosa para la enseñanza de los idiomas ya en el siglo XX. La creación manual de las listas fue un trabajo duro, pero fue considerado como ejercicio de mérito. La idea principal era que a más frecuencia, más importancia para el aprendizaje (Leech 2011: 8). El desarrollo de la lingüística del corpus ha propiciado el análisis estadístico de la frecuencia de las palabras que aparecen en los diferentes textos que los forman. Una obra bastante reciente y difundida, es el diccionario publicado por Davies. Sin entrar ahora en detalles, antes de él hubo numerosas aportaciones en este campo y, con todo, no se puede resaltar ninguna como un auténtico listado de referencia que ofrezca datos actuales, representativos, fiables y rigurosos. Los listados son insuficientes debido a diversas causas: algunos, por ser demasiado antiguos en la selección del corpus, otros por su metodología de análisis poco actual o rigurosa, bien por circunscribirse solo a un área concreta o el tipo de lengua o registro, bien en una franja etaria reducida, etc. En este sentido, la lista de obras que se podría citar es numerosa¹⁸, aunque, como hemos dicho, las más recientes solo han solucionado en parte la cuestión de elaborar un verdadero “Léxico básico del español”.

De los diccionarios de frecuencias ya clásicos, gozaron de prestigio y difusión el de García Hoz (1953) y, sobre todo, el de Juilland y Chang-Rodríguez (1964) por su metodología y criterios, que sentaron las bases para otras obras, como la de Morales (1986), centrada en Puerto Rico. El diccionario de Davies no

¹⁸ Ordenadas por año de publicación, algunas de estas obras son: Keniston (1920), Thorndike (1921), Buchanan (1927), Eaton (1940), Rodríguez Bou (1952), García Hoz (1953), Gougenheim, Michéa, Rivenc & Sauvageot (1956, 1958), Müller (1964), Juilland & Chang-Rodríguez (1964), Márquez Villegas (1975), Díaz Castañón (1977), Morales (1986), Alvar Ezquerra & Villena Ponsoda (coords.) (1994), Alameda & Cuetos (1995), Benítez Pérez, Hernández Cabrera & Samper Padilla (1995), Justicia (1995), Molero Huertas (1999), Ávila Muñoz (1999), Sebastián, Cuetos, Martí & Carreiras (2000), Terrádez Gurrea (2001), Martínez & García (2004), Almela *et al.* (2005), Davies (2006), García Surrallés (2012).

goza de tan gran prestigio entre investigadores, aunque por ser el más reciente y difundido, gracias en parte a su disponibilidad en línea, resulta una obra muy útil y práctica para su consulta, y ha logrado cubrir el vacío que quizá debería haber llenado el proyecto LEBAES (*Léxicos Básicos de España*, Benítez Pérez, Hernández Cabrera & Samper Padilla 1995), que no ha sido llevado a cabo. En un artículo a modo de presentación, sientan las bases de lo que debe ser un léxico básico del español riguroso que podría servir como referente para aplicarlo a la docencia, así como para otros proyectos, como la elaboración del léxico fundamental del español. En otro capítulo comentaremos con más pormenores estos diccionarios de frecuencias (*vid. infra* § 5.2. y 5.3.); por ahora, nos centramos solo en valorar sus aportaciones a la cuestión de la selección léxica.

Si pensamos en las ventajas que tiene para la selección léxica el uso de las listas de frecuencia, se pueden esgrimir diferentes argumentos. Así, por ejemplo, según Alvar Ezquerro sirven para conocer la importancia de las palabras de acuerdo con su frecuencia y distribución, con lo que poseemos una información objetiva que como docentes nos puede ayudar a decidir qué palabras deben enseñarse y cuándo en función de cada nivel, en cuáles hay que insistir más y cuáles deben enseñarse para que el alumnado entienda los materiales que maneja. Para quienes confeccionan los materiales pedagógicos, les puede ayudar sin duda a seleccionar el léxico que debe aparecer en ellos. Nos permite, asimismo, extraer el inventario de palabras del ámbito lingüístico en que se mueven los alumnos, para facilitarles la comprensión de su entorno o determinar el grado de dificultad del léxico de los materiales de lectura y aprendizaje (Alvar Ezquerro 2005: 21). Se podría saber cuáles son las palabras de ámbitos concretos cuando se enseña la lengua con fines específicos, trabajando por ejemplo las frecuencias en corpus específicos. También podemos intentar seleccionar los grupos de palabras y formas que deben enseñarse conjuntamente, y en qué orden.

Sin embargo, y a pesar de sus ventajas, incluso hoy en día los diccionarios de frecuencia no son considerados como fuente principal para obtener información sobre qué palabras enseñar. Puede ser que haya otros factores que afectan a la credibilidad de las listas de frecuencias, ya que presuponen la existencia de un diccionario adecuado. Hay diversos autores que aseguran que la

falta de un diccionario y de un corpus fiable es probablemente la razón principal por la que no es extendido el uso de los diccionarios de frecuencia (Madison 2013: 27-32). Muchos de los diccionarios de frecuencias que tenemos para el español son antiguos, 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 uso. Uno de los argumentos en contra del empleo de las listas de frecuencia se basa en el hecho de que el corpus es considerado inadecuado. No obstante, las técnicas han mejorado y también el análisis del corpus. El primer corpus que contenía más de un millón de palabras en español, por ejemplo, fue publicado en 2001, así que estas críticas deben pensarse sobre todo para los publicados antes. Otro problema es que los diccionarios de frecuencias no sean verdaderamente representativos del español, lo que justificaría que no se tengan del todo en cuenta para la elección del vocabulario que se pretende incluir en un libro de texto (Davies & Face 2006: 2).

Con respecto a las voces que obtenemos gracias a este método, a partir de numerosos vocabularios de frecuencias 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. Además de estos elementos funcionales conseguimos extraer 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*). 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 (es decir, con diversos significados) y polivalentes (que se usan en diversas construcciones) tales como *ser, haber, tener, estar, ir, hacer, decir, ver, saber, venir, dar, poder*, entre otros. 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, por lo que tendremos que alcanzarlas mediante otros métodos. Y aquí reside uno de sus principales inconvenientes, ya que a estas unidades relativamente poco frecuentes no se consigue acceder ni siquiera con corpus muy amplios y representativos.

Salazar García hace una clasificación de las insuficiencias de los diccionarios de frecuencias (Salazar García 2004: 247), siguiendo a Bartol Hernández (2002: 63), de modo general, y a Ortega Olivares (1985) respecto al diccionario de Juilland y Chang-Rodríguez (1964). Indica que estas insuficiencias responden a motivaciones de distinta índole y se encarga de delimitarlas en tres grandes apartados:

1) Las cuestiones que derivan de la propia naturaleza del léxico que cada diccionario de frecuencias se limita a aportar. A pesar de que son fáciles de reconocer y de subsanar mediante mecanismos correctores no muy complejos, estarán siempre presentes, independientemente de la calidad del diccionario.

2) Las que tienen que ver con las decisiones metodológicas, tales como la representatividad del corpus de partida, su horquilla diacrónica y el etiquetado de las unidades discretas con las que se opera. Destaca que son las que mayor incidencia tienen en la calidad del producto final. Aunque pueden estar mediatizadas por limitaciones de tipo externo, recursos materiales, humanos, o hasta exigencias editoriales, de modo general es el diseño de la investigación lo que más influye.

3) Otras limitaciones técnicas que interfieren en la elaboración del diccionario. Durante los últimos tiempos se han mejorado enormemente las condiciones de viabilidad, gracias a los avances técnicos de la informática, así como su plasmación en el campo de la lingüística de corpus y de las llamadas industrias de la lengua.

No cabe duda que empezar con el aprendizaje de las palabras más frecuentes de una lengua puede ser útil, ya que teóricamente, esas son las palabras que los estudiantes van a encontrar más frecuentemente en el lenguaje de otras personas y las van a necesitar a menudo en su propio uso de la lengua extranjera. No obstante, la lista puede perder utilidad por diferentes motivos.

Algunos léxicos de frecuencias del español que han gozado de gran prestigio y han sido referentes de muchos otros por su cuidado en el trabajo de selección y tratamiento del corpus ofrecen hoy por hoy una selección de textos obsoleta. Un claro ejemplo es el de Juilland y Chang-Rodríguez (1964), que durante muchos años sirvió de referencia para léxicos de otras lenguas y perfeccionó algunas cuestiones en relación con las fórmulas de medir la frecuencia y la dispersión. Pero sus textos eran de las décadas de 1920 hasta 1940, por lo que el léxico que ofrecen no es actual.

Otra cuestión es la forma de organizar los textos, es decir, los diferentes mundos o dominios que se tienen en cuenta para seleccionar los textos. Por ejemplo, la muestra que Rodríguez Bou (1952) seleccionó en su día incluía cuadernos escolares, diarios y documentos religiosos. Según Leech, cuando el recuento de la frecuencia se basa en un corpus de lengua general, no tiene en cuenta las variedades (regionales, por ejemplo) o bien no introduce el lenguaje oral. También hay estudios que muestran que hay diferencias en la frecuencia entre inglés americano e inglés británico, entre el vocabulario que se usa en la escritura y en el lenguaje oral, y estos argumentos son aplicables también al español (Leech 2011: 12-14). El español escrito y hablado varían bastante, aunque la frecuencia es un criterio matemático y los resultados son objetivos, pero la lista de frecuencias resultante dependerá de la preparación de la muestra sobre la que se haya trabajado. Las primeras críticas en el campo de las lenguas modernas surgieron porque generalmente las muestras estaban tomadas de textos escritos, por lo que las diferencias entre lengua hablada y lengua escrita se manifestaban también en la selección. Gougenheim (1958) intenta paliar este problema utilizando además grabaciones, pero aun así no se libró de otras críticas que atañen más a la sociolingüística. Por eso, al usar un diccionario de frecuencias hay que prestar atención a los criterios que se han seguido para la elaboración del recuento. Por ejemplo, puede ocurrir que queramos enseñar una variedad concreta, o bien todo lo contrario, que el objetivo sea abarcar la lengua general. Así, podemos querer obtener datos de la lengua de una región concreta, o bien el objetivo puede ser dar una visión y conocimiento general. Siempre hay que tener en cuenta los objetivos del aprendizaje. Por ejemplo, en el corpus del diccionario de Davies (2006) aproximadamente el 43% de los textos procede de

España y el 57% proviene de Hispanoamérica (Davies 2006: 3), así que, en el diccionario van a aparecer palabras frecuentes en América pero quizás no tanto de España, por lo que puede ser una herramienta muy útil para quien quiera crear manuales y mostrar la variedad del español de América. Alguien que trabaje con la variedad del español de Puerto Rico, podría recurrir al *Léxico básico del español de Puerto Rico* de Morales (1986), o si lo que interesa es el de la variedad andaluza, existe el léxico básico de la modalidad andaluza del español (García Surrallés 2012), también contamos con un recuento del español hablado, el de Terrádez Gurrea (*Frecuencias léxicas del español coloquial*, 2001), y es posible seleccionar alguno de los que se centran en etapas infantiles, como el de Martínez y García (*Frecuencias del castellano escrito en niños de 6 a 12 años*, 2004). Debemos escoger un diccionario que coincida con las necesidades de nuestros estudiantes en cada caso: un corpus útil para estudiantes adultos, puede no serlo tanto para los más jóvenes (Leech 2011: 25-26).

Otros aspectos son los que tienen que ver con el tratamiento de los datos y de la muestra. Algunos trabajos, por ejemplo, no incluían la frecuencia relativa, como el de Rodríguez Bou. Después, el criterio de frecuencia empezó a ser complementado por otros, como el de la dispersión, que trata de medir la extensión de una palabra en los diversos ámbitos o dominios en que se había dividido previamente la muestra. Así, la importancia comunicativa de una palabra no se rige solo por el dato de frecuencia, sino que tenemos también en cuenta la *dispersión* y el *índice de uso*. En algunos trabajos, sobre todo del ámbito anglosajón, esto se conoce como *rango*. Dos unidades pueden tener la misma frecuencia pero tener una distribución diferente en los distintos tipos discursivos que forman la muestra de nuestro corpus. Debido a esto, una palabra que aparezca muchas veces en un mismo tipo de texto es menos importante que una que tenga la misma frecuencia pero que aparezca constantemente en diversos tipos de texto. En este caso estamos midiendo, además de la frecuencia, el rango, que consiste en medir la dispersión de un término en periodos temporales diversos y en la serie de fuentes, mundos o tipos de texto que constituyen el corpus. La dispersión nos señala el valor general de un término, al determinar su presencia en conjuntos o tipos de textos de la muestra total. Nos aporta, pues, la estabilidad de la frecuencia, con lo que obtenemos la información

de interés sobre la utilidad de una unidad léxica en la lengua (Lara 2006: 170). El diccionario de Davies, por ejemplo, otorga gran importancia a estos criterios y para ello establece una fórmula ideada por él mismo para hallar las palabras, que presenta ordenadas por su rango de uso, esto es, de frecuencia más dispersión. Alonso Ramos cita un ejemplo extraído del diccionario de Davies. La palabra *desconfianza* aparece en la posición 3.170 con 247 ocurrencias, frente a *pánico* que ocupa la posición 3.227, con 297 ocurrencias, debido a que la primera tiene más dispersión que la segunda en el corpus. La frecuencia y la dispersión son los criterios empleados para asignar el índice de uso, que se aproxima más al de utilidad.

Una cuestión importante en relación con las decisiones metodológicas es el trabajo de etiquetado, es decir, si se mide la frecuencia de las formas o ítems (*tokens*) o bien de lemas o vocablos (*types*). La frecuencia que etiqueta el corpus con *types* nos ofrece la frecuencia de patrones, mientras que si etiquetamos *tokens* llegamos a la de los ítems. Se puede ver esto en los verbos, por ejemplo. En el primer caso podríamos buscar la regularidad de las formas de pasado de todos los verbos, o bien la frecuencia de todas las formas del paradigma de un verbo concreto, como *hacer*, indistintamente de sus tiempos, personas, modos. Por el contrario, si un corpus no está etiquetado en lemas, solo podremos saber la frecuencia de formas exactas y no de su paradigma. Otra cuestión es cómo esté hecho este trabajo de etiquetado y el tipo de información que permita rescatar, con lo que podrá resultar más o menos útil en función de la calidad y afinación de este trabajo. En este sentido, las palabras que Davies dedica en la introducción de su diccionario a los anteriores diccionarios de frecuencias del español son muy ilustrativas, ya que resume de modo conciso y claro algunas de las críticas que venimos comentando:

There have been a number of other frequency dictionaries and lists for Spanish (Buchanan 1927, Eaton 1940, Rodríguez Bou 1952, García Hoz 1953, Juilland and Chang-Rodríguez 1964, Alameda and Cuetos 1995, Sebastián, Carreiras, and Cuetos 2000), but all of these suffer from significant limitations. First, all of these frequency dictionaries are based exclusively on written Spanish, and contain no data from the spoken register. Second, five of the dictionaries (Buchanan 1927, Eaton 1940, Rodríguez Bou 1952, García Hoz 1953, Juilland and Chang-Rodríguez 1964) are based on texts from the 1950s or earlier, and are now quite outdated. Third, the two dictionaries that have been produced in the last ten years both suffer from other important limitations. Alameda and Cuetos (1995)

only lists exact forms (e.g. digo, dices, dijeran) rather than lemma (e.g. decir), and very few of the written texts that it uses are from outside of Spain. The other recent dictionary –Sebastián, Carreiras, and Cuetos (2000)– exists only in electronic form and is extremely hard to acquire, especially outside of Spain. [Davies 2006: 2]

Los medios técnicos actuales ayudan cada vez más a obtener un etiquetado eficiente, aunque todavía no han resuelto todos los problemas. Por ejemplo, en la cuestión de qué es lo que entendemos por palabra se ha evolucionado, pero todavía hay retos por afrontar. Muchos autores se han detenido en sus introducciones en explicar cuál era el concepto de “palabra” con el que trabajaban pues, en definitiva, era esto lo que pretendían contar y ofrecer. También en cómo solucionaban las cuestiones que tienen que ver con los homógrafos o las polisemias. No obstante, en los listados clásicos de frecuencias del español, como el de Juilland y Chang-Rodríguez (1965) y el de Morales (1986), el recuento es únicamente de unidades monoverbales. Esta constatación nos indica que los autores partieron de una definición ortográfica y no semántica de la palabra. El problema es que, al proceder de tal manera, los listados ocultan la presencia de combinaciones fraseológicas y no dan información sobre su frecuencia; y ya hemos comentado la importancia de las colocaciones y de ciertas locuciones en la lengua. Sólo el reciente recuento de frecuencias léxicas de Ávila Muñoz (1999) ha superado la identificación tradicional de la unidad léxica al incluir, a diferencia de los anteriores listados, unidades fraseológicas (locuciones, fórmulas rutinarias, etc.). Ha supuesto también una excepción por prestar especial atención a los marcadores discursivos como *o sea, por ejemplo, a lo mejor, etc.*

Como hemos dicho, si bien los diccionarios que cuentan lemas son una ventaja sobre los que cuentan formas, no nos permiten obtener datos sobre el eje sintagmático, es decir, sobre colocaciones. Para eso tendríamos que medir la frecuencia de los diferentes tipos de “unidades léxicas” y no de palabras gráficas. Por ejemplo, en el ámbito del inglés como L2, varios autores llamaron la atención sobre el hecho de que los llamados *delexical verbs* no pueden ser enseñados sin referencia a las colocaciones en que entran. Estos *delexical verbs* son verbos comunes, como *have, take, make* o *give*, que cuando se usan con sustantivos

particulares tienen muy poco significado propio. En estas estructuras la mayor parte del significado se encuentra en el sustantivo, no en el verbo. En la mayoría de las ocasiones hay un verbo que tiene un significado similar o igual a la estructura "delexical". En muchos casos, hay verbos y estructuras equivalentes en español. Estas estructuras son muy comunes en inglés y comprenderlas y conocerlas enriquecerá la fluidez de la lengua (Sinclair & Renouf 1988: 152). En el ámbito francés, ya Galisson (1979: 58) había defendido la importancia del eje sintagmático, aunque para medir la disponibilidad de las palabras, por lo que propuso encuestas de disponibilidad sintagmática que consistían en preguntar por los verbos y adjetivos que vienen a la mente en coocurrencia con determinados nombres.

En el español, varias voces han comentado más recientemente la necesidad de este cambio de concepción de la palabra y la necesidad de un mayor esfuerzo para elaborar corpus anotados, ya que los problemas planteados por la mayoría de las listas de frecuencia apuntan en esta misma dirección. Bogaards (2008: 1234), entre otros, señaló que existe una fuerte correlación entre frecuencia y polisemia: cuantos más sentidos tenga una palabra, más ocurrencias tiene. Por ejemplo, en Davies (2006), en la palabra *carta* no se distingue entre el sentido "letter" y el de "playing card", aunque es más que probable que la frecuencia de los dos sentidos no sea la misma, pero, a pesar de ello, se le asigna la misma frecuencia y orden en la lista (Alonso Ramos 2012: 24). Sin embargo, quien estudia una lengua extranjera necesita aprender formas vinculadas a sentidos, al igual que el lexicógrafo necesita saber qué acepciones son más frecuentes que otras. Rojo (2009) indica que la meta es crear corpus desambiguados semánticamente, por muy arduo que sea el proceso, para distinguir la frecuencia de las acepciones.

Otro beneficio que tiene conocer las frecuencias de las unidades léxicas es su uso en los tests de vocabulario de los aprendices. Por ejemplo, Casso (2010: 82) elabora, en la línea de lo realizado por Nation anteriormente (2001), un test de vocabulario receptivo usando las 5.000 palabras más frecuentes del corpus *Cumbre*. En este test, el discente debe unir definiciones con palabras, pero, como la lista de *Cumbre* no refleja la frecuencia de las acepciones, quien recurra a ella se debe guiar por la intuición y no por criterios objetivos.

Y es que un corpus sin desambiguar puede ayudar al hablante nativo pero no tanto al aprendiz de la lengua. Por eso se hace necesario enriquecer el corpus con la desambiguación semántica, para mejorar el valor de las frecuencias léxicas y la identificación de locuciones o expresiones fraseológicas antes de hacer los recuentos. Si no, nos ocurrirá como en el trabajo de Davies (2006) o el de Almela *et al.* (2005), en los que *vez* está entre los tres nombres más frecuentes pero no se indica que esto ocurre por el hecho de aparecer recurrentemente en expresiones como *a la vez, a veces, de una vez, de vez en cuando, tal vez*, etc. Sería mucho más útil saber la frecuencia de todas estas locuciones con independencia de que contengan el nombre *vez*. Lo mismo nos ocurre si consultamos la palabra *embargo*, pues lo veremos en una posición de frecuencia alta, pero gracias a la locución *sin embargo* (Almela *et al.* 2005: 28). Existe todo un ámbito de investigación, conocido en inglés como *Word sense disambiguation*, que trata de la desambiguación automática de los sentidos de las palabras (Aguirre & Edmonds 2006). La frecuencia de las palabras aisladas no es directamente utilizable ni por el lexicógrafo ni por el autor de material didáctico que se interese por las colocaciones. Tomemos el ejemplo de verbos muy frecuentes, como *dar, tomar, hacer o tener*, que funcionan a menudo como verbos de apoyo. De nada sirve incluir estos cuatro verbos en un manual de ELE (o seleccionarlos para un vocabulario de ELE) sin el nombre que los selecciona léxicamente, porque un aprendiz tendrá que aprender que la combinatoria de estos verbos está controlada por el nombre de su base: los *paseos se dan*, los *viajes se hacen*, al igual que las *proposiciones*, pero no como las *decisiones*, que se *toman*. En otras palabras, incluir esos cuatro verbos como prioritarios en el aprendizaje del vocabulario sin los nombres asociados no parece adecuado, porque muchas de las palabras frecuentes lo son debido a que son escogidas frecuentemente como colocativos de bases que son menos frecuentes. Así, se puede dar el caso de dejar fuera de la lista de lo priorizable las bases que seleccionan los colocativos frecuentes. Sánchez (2000: 24) incide en la misma idea de que los aprendices comunicarán más fácilmente si conocen no solo las palabras más frecuentes sino las que las acompañan, que pueden ser no tan frecuentes. Menciona el caso de *teléfono*, que está entre las mil más frecuentes, pero no algunas de las que la acompañan usualmente. Podríamos dar un paso más: incluso aunque las

palabras que la acompañen sean frecuentes, como es el caso de *coger*, el significado de este verbo con *teléfono* nada tiene que ver con el que ofrece con *manzana*, pongamos por caso.

En los últimos años se ha promovido el recurso a listados de concordancias para que los aprendices puedan consultar directamente los corpus, pues tienen la ventaja de presentar las unidades léxicas de forma contextualizada. Con respecto a los listados de frecuencias clásicos, los actuales corpus electrónicos, consultables en línea, nos proporcionan las voces contextualizadas en las pantallas de concordancias. Por ejemplo, Izquierdo Gil (2002) propone recurrir al *CREA (Corpus de Referencia del Español Actual)* para ver las concordancias y acceder a los contextos de uso de las palabras, también para ver las colocaciones mediante una búsqueda por “agrupación” en 2 o 3 palabras. Este corpus puede evidenciar las combinaciones más destacables, ya que las concordancias suelen revelar la presencia de colocaciones, es decir, combinaciones frecuentes de dos o más palabras o unidades que tienen una tendencia muy fuerte a ser utilizadas juntas. Así pues, si buscamos en el *CREA* unidades léxicas con un elevado índice de uso en los léxicos básicos de Juilland y Chang-Rodríguez (1964), Morales (1986) y Ávila Muñoz (1999), tales como *cada, menos, mundo, ejemplo, lado, mejor, momento*, las encontraremos recurrentemente en colocaciones de producciones lingüísticas orales. Puede ser altamente útil este recurso también para explorar la frecuencia de las colocaciones gramaticales y de las construcciones con un verbo soporte. En este sentido, las concordancias de los corpus constituyen un complemento de los listados de frecuencias léxico-estadísticos.

No obstante, los corpus no hablan solos, y la dificultad en la consulta y sus resultados es, en ocasiones, excesiva para que un aprendiz de nivel inicial pueda distinguir los diferentes sentidos de una palabra en las concordancias, por lo que es necesario un filtrado y una descripción a la que solo se accede a través bien de un diccionario, bien de un corpus anotado (Alonso Ramos 2009). La léxico-estadística nos muestra, por consiguiente, que muchos enunciados fraseológicos tienen un peso considerable en el vocabulario de una lengua, así que, como apunta McCarthy, los más frecuentes de dichos enunciados deben hallarse presentes en un vocabulario básico. No podemos pasar por alto en el aprendizaje

de una lengua extranjera las existencia de unidades fraseológicas de elevada frecuencia. Actualmente se insiste en que un conocimiento de dichas unidades es esencial en la enseñanza del léxico L2 para desarrollar la competencia comunicativa, especialmente la habilidad léxico-discursiva y sociocomunicativa (McCarthy 1990: 67; Carter 1996: 176; Sinclair 1991; Willis 1999; Izquierdo Gil 2002; Higuera García 1996, 1997a, 1997b, 2004a, 2004b, 2005, 2006, 2009; Alonso Ramos 2012). Y, aun así, no basta con un corpus anotado, todavía más importante es que el corpus sea equilibrado. De no ser así, de nuevo, la objetividad de los números se ve reducida por la subjetividad en la selección de las muestras que constituyen el corpus.

A los problemas de índole conceptual o metodológico, a veces se añaden también las considerables oscilaciones terminológicas entre las diferentes obras. Por ejemplo, Nation (1990), tras señalar las deficiencias de los recuentos basados en la frecuencia, propone que hay que tener en cuenta otros parámetros, como el rango o índice de estabilidad, lo que Juilland y Chang-Rodríguez (1964) habían llamado *índice de dispersión*. García Hoz (1953) manejaba los términos “usual”, “común” y “fundamental” con un sentido muy diferente al que le damos nosotros o autores como Salazar García (2004) o Davies (2006), por ejemplo. Para García Hoz el “vocabulario usual” era el formado por el conjunto de vocablos documentados en su corpus de referencia, que sumaba un total de 12.913 unidades. En su definición de vocabulario usual explica que se identifica este tipo de léxico como el empleado en la práctica por lo que él entiende como el “hombre vulgar”. Entiende que es el que ha recibido una educación general y no tiene una especialización profesional, por lo que este tipo de individuos hacen un uso de la lengua sin mediar los conocimientos propios de la especialización. Con esto suponemos que se refiere a un individuo de educación media que emplea la lengua en contextos cotidianos, dando cabida al vocabulario más corriente y dejando fuera aquel más restringido, bien por limitaciones culturales, dialectales, técnicas, etc. A su vez, diferencia otros dos tipos de vocabularios, el usado en este tipo de contextos cotidianos y el que conoce cada individuo, pues en este último sí que puede haber una gran diferencia entre las distintas personas. También distingue entre el léxico activo y pasivo, aunque ambos forman parte del usual. Para determinar cuál es el “vocabulario usual” manifiesta que, ante la

imposibilidad de registrar el lenguaje oral como método ideal, recurre al lenguaje escrito como su transposición y como la expresión lingüística normal de la sociedad, entendida como el conjunto de individuos adultos. Para ello recurre como fuente a lo que denomina “distintas manifestaciones de la vida humana”: vida familiar, vida social regulada e indiferenciada y la vida cultural. El resultado es el vocabulario usual, a partir del que extrajo como subconjunto el vocabulario común (1.971 unidades) y, como listado esencial último, el vocabulario fundamental (208 unidades). Para establecer el vocabulario común selecciona las palabras que son comunes a los cuatro tipos de vocabularios (de los ámbitos referidos anteriormente), pues estas son las más generales e importantes, y tienen mayor probabilidad de aparecer, independientemente del ámbito en que nos encontremos. El uso de “fundamental” tampoco concuerda con el que se suele dar en la bibliografía francesa sobre el léxico disponible y fundamental (Gougenheim, Michéa, Rivenc & Sauvageot 1958) ni en la hispánica posterior (López Morales). En estas últimas tradiciones, el vocabulario fundamental abarca un volumen léxico más amplio, que surge de la fusión de los datos de frecuencia y los de disponibilidad. Tampoco el sentido que da Justicia (1995) a los términos de “vocabulario usual” y “vocabulario común” resulta coincidente con el que se documenta en otras investigaciones léxico-métricas. Para García Hoz, al “vocabulario fundamental” se llega después de realizar diferentes cálculos con los que alcanza a saber la correlación entre las palabras de los cuatro ámbitos delimitados en su corpus. Así, este tipo de vocabulario se caracteriza por estar constituido por las palabras que tienen sus frecuencias prácticamente distribuidas en partes iguales en uno y otro tipo de vocabulario.

Dejamos en este punto las limitaciones que tienen que ver con las decisiones metodológicas y observamos algunas otras de tipo técnico. Pensamos por ejemplo en el manejo y en la difusión de los recuentos de frecuencia. Es evidente que en esto el paso de los años ha tenido mucho que decir, pues en poco se parece el diccionario de Davies (2006) al *FDSW* de Juilland y Chang-Rodríguez (1964) y, sin embargo, este último gozó de gran difusión y seguimiento durante muchos años. Otro ejemplo es el *Recuento de vocabulario español* de Rodríguez Bou (1952), una obra de dos volúmenes, cada uno de los cuales comprende más de quinientas páginas. El primer volumen contiene las diez mil unidades léxicas

más frecuentes en orden de rango; las diez mil unidades léxicas más frecuentes en orden alfabético, con sus frecuencias y rangos; las veinte mil formas de flexión más frecuentes en orden alfabético (con sus frecuencias y rangos) y las unidades léxicas con frecuencias menores de 16. El segundo volumen (de dos tomos) contiene las frecuencias ponderables de las 20.542 unidades léxicas y de las 62.888 formas de inflexión que se encontraron en las diez fuentes estudiadas. Por si esto fuera poco, para emplear este *Recuento* es necesario comprender claramente la diferencia entre las diferentes listas que ofrece, las listas de unidades léxicas y las de formas de inflexión. Estas últimas contienen los elementos que se encontraron en las fuentes estudiadas exactamente en la forma en que allí aparecieron (verbos conjugados, sustantivos en singular y plural, adjetivos en masculino y femenino, etc.). En cambio, las listas de unidades léxicas contienen la forma base de cada palabra (infinitivos, sustantivos en el singular, adjetivos en el masculino singular y todas las formas que nunca se flexionan, como las conjunciones, adverbios, preposiciones, etc.). Esta diferenciación es importante, ya que el orden en que aparece, por ejemplo, un infinitivo, no indica que todas las formas conjugadas de ese verbo tengan la misma frecuencia. En resumen, el *Recuento* no contiene una lista práctica de palabras de alta frecuencia que combine unidades léxicas y formas de flexión ni tampoco puede usarse con facilidad en la docencia o en la investigación. Desde luego, no todos los léxicos frecuentes en papel son de esta envergadura y complejidad en su aplicación, pero hay diccionarios recientes que son también considerables en tamaño. En ese sentido, los léxicos básicos suelen ser mucho más concisos y selectivos, por lo que se muestran más adecuados para el manejo de alguien que no sea un investigador en el tema. Los medios electrónicos han ayudado mucho a mejorar el acceso. Uno de los primeros que supo sacar partido de ello fue el proyecto *LEXESP: Léxico informatizado del español* (Sebastián, Cuetos, Martí & Carreiras 2000), una base de datos de palabras del español que ofrece un libro y un CD. El uso del CD permite instalar un programa y usar su contenido en nuestro ordenador. Está compuesto además por dos partes, una que utiliza el corpus lematizado como base y permite hacer búsquedas de apariciones y coapariciones de palabras, lemas y categorías de palabras, dentro de un contexto, y otra que emplea listados de palabras, previamente indexados, para la creación

de estudios que permitan la búsqueda de características de palabras. Para quien prefiera, pues, el manejo de estas herramientas mediante el ordenador, puede resultar mucho más cómoda la consulta de esta obra, además de que ocupa menos espacio que un libro en papel. No obstante, su uso no resulta del todo simple e intuitivo sin tener que leer el manual de instrucciones, y el hecho de que no se pueda acceder a él en línea le hace un flaco favor a su difusión. En este sentido, el diccionario de Davies (2006) supone un adelanto claro: su distribución es en formato electrónico (extensión .pdf) y se puede acceder a cada capítulo con solo pinchar en el índice. Resulta una apuesta muy práctica y actual, y está extraído de un amplio corpus disponible en línea, por lo que ambas herramientas se pueden complementar.

3.4.1.1. Los vocabularios básicos y los corpus de frecuencias

El crecimiento y difusión de los recursos informáticos ha propiciado en nuestros días la aparición de una nueva disciplina lingüística, la denominada Lingüística de corpus (Ávila Muñoz 1999: 12). La Lingüística de corpus se basa en una técnica de trabajo que propicia el acopio masivo de datos gracias a las herramientas informáticas, permitiendo y favoreciendo la realización de investigaciones empíricas del lenguaje y las lenguas, más que exclusivamente el estudio introspectivo. Es el mismo principio que guiaba los estudios estadísticos más tradicionales sobre el lenguaje, pero las aplicaciones de la lingüística de corpus son mucho más ambiciosas, debido a los progresos que la informática ha experimentado desde los años 60 hasta nuestros días. Y ya que esta línea de trabajo va de la mano de las nuevas tecnologías, al igual que ellas es de esperar que en un futuro inmediato nos ofrezca unas posibilidades de consulta que se adecuen cada vez más a nuestras necesidades. Los corpus están formados por una gran cantidad de datos lingüísticos que se pretenden representativos de la lengua en alguna o varias de sus manifestaciones (lengua hablada o escrita), niveles (culto, literario, coloquial, etc.), medios por los que se transmite (novela, teatro, prensa diaria, radio, etc.), especialidades o ámbitos (en inglés llamados *domains*: economía, política, deporte, historia, aficiones y coleccionismo...), etc. Los datos 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 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. Para el español todavía no son muchos los corpus que existen, aunque cada vez hay más, de contenidos más ricos y de mayores posibilidades de acceso, si bien algunos de ellos no son de dominio público. En general, presentan las palabras en listados de concordancias, esto es, cada voz con un contexto más o menos largo, pero sin estadísticas globales, lo que no quiere decir que no existan corpus en los que se ofrezcan frecuencias, o que no se hayan desarrollado programas específicos para obtenerlas. El *CREA*, por ejemplo, así lo hace, y el *CORPES XXI* se orienta también hacia este fin. Sin embargo, lo que nos presenta la relación de concordancias son las formas que se encuentran en los textos, sin el menor análisis ni agrupación. Probablemente no pase mucho tiempo hasta que las consultas nos ofrezcan las formas agrupadas de acuerdo con filtros de búsqueda, con lo que serán un instrumento verdaderamente útil para la preparación de los materiales didácticos. Otra dificultad con que podemos tropezar en la consulta de los corpus es la limitación en el número de accesos o en el de elementos presentados.

El desarrollo de la lingüística del corpus proporcionó unos resultados que se manifestaban útiles, el desarrollo de la informática simplificaba su manejo y, finalmente, la difusión de internet ha contribuido a agilizar y facilitar su consulta, por lo que se continúa viendo como una fuente útil y empírica. Así, hoy en día existen numerosos corpus y de diferentes tipos, tanto orales como escritos, algunos incluso nacidos con la intención de ser aplicados a la docencia de ELE. No nos detendremos ahora en hacer un análisis pormenorizado de las diferentes bases de datos y de las publicaciones de este tipo que existen, pero sí en apuntar algunas características¹⁹. Según la definición de Rojo:

¹⁹ Para profundizar en el estado de la lingüística de corpus del español se puede obtener abundante información sobre el panorama actual y lo que ha supuesto en la lingüística hispánica, así como sobre su difusión en la red, en algunos trabajos de Rojo (2005, 2005-2006, 2008, 2010, 2016) y Rojo & Sánchez (2010). También se encuentra una extensa lista de referencias de los corpus existentes en español en la actualidad en Rojo 2016.

Un corpus es un conjunto de textos (o fragmentos de textos) naturales, almacenados en formato electrónico, que resultan conjuntamente representativos de una variedad lingüística en su totalidad o en alguno de sus componentes, reunidos con el fin de que puedan ser estudiados científicamente. [Rojo 2015]

Y los aspectos que los caracterizan son:

- Los textos (o fragmentos de textos) tienen que haber sido producidos en condiciones reales y naturales.
- Los textos deben estar en formato electrónico.
- Deben ser representativos de la variedad de la que han sido extraídos. Además, deben estar equilibrados.
- El corpus tiene que estar construido de modo que sea posible el análisis científico (no solo lingüístico).
- Además, su organización debe permitir el enriquecimiento del corpus mediante la adición de codificación y anotación morfosintáctica, sintáctica, semántica y pragmática.

Desde el desarrollo de los corpus, después de mediado el siglo XX, uno de sus objetivos fue determinar la frecuencia de las palabras. Se pensaba que usando métodos objetivos de cuantificar aquellas voces que eran usadas un mayor número de veces, daríamos con las más rentables. La frecuencia, así pues, era lo que determinaba cuáles eran las palabras más útiles. Este método servía para seleccionar aquellos elementos que debían formar parte de los diccionarios, por ejemplo. Así, en lugar de confiar la elección a los conocimientos del lexicógrafo, aunque fueran enciclopédicos y vastos, se pasaba a buscar las palabras más usadas en los textos reales producidos de forma natural. El uso de los ordenadores permitía llevar a cabo estos métodos que pretendían otorgar un halo científico y empírico a la investigación lingüística. En la década de los 60 y la primera mitad de los 70 se construyen corpus electrónicos con objetivos orientados a este trabajo lexicográfico. Los primeros nacieron en el ámbito anglosajón y en EE. UU., aunque luego se tomó el relevo en Inglaterra, Noruega y, en menor medida, Alemania, Francia e Italia. En inglés son conocidos los proyectos LOB (Lancaster-Oslo/Bergen Corpus, de 1960 para el inglés británico), el corpus de Longman o el COBUILD. Aplicados a la enseñanza-aprendizaje de lenguas, la posibilidad de cuantificar la frecuencia de las palabras permite hoy en día la creación de listas de frecuencia y diccionarios básicos más precisos. El pensamiento que estaba por detrás de esta búsqueda era que, a pesar de que el vocabulario que consigue reconocer un hablante puede ser muy numeroso

(entre léxico activo y pasivo), el que realmente usan los nativos de una lengua en contextos habituales, e incluso académicos, es mucho menor. De ahí que conseguir saber cuáles son esas palabras que son usadas habitualmente por los hablantes de un idioma sea altamente útil, para tomarlas como referencia en el aprendizaje y la enseñanza a extranjeros.

De todos los que existen en el ámbito hispánico, por nuestra parte resaltamos solo tres corpus que pueden ser útiles para la selección léxica: el *Corpus Cumbre*, el *Corpus del Español* de Davies y el *CORPES* de la Real Academia Española. A estos añadimos, por estar construido específicamente con muestras obtenidas de estudiantes de ELE, el *Corpus de Aprendices de Español, CAES*).

El *Corpus Cumbre* (2001) fue un proyecto financiado por la editorial *SGEL* y dirigido por el profesor Aquilino Sánchez. Se trata de un corpus de 20 millones de palabras, representativo del español contemporáneo oral y escrito de todos los países de habla hispana. Ha sido utilizado para la elaboración de un diccionario²⁰ y una gramática práctica.

Los otros corpus están disponibles en la red. El primero es el de Davies, (<http://www.corpusdelespanol.org>), realizado en 2002 pero con sucesivas nuevas versiones en 2007 y 2016. Se basa en unos 100 millones de palabras procedentes de más de 20.000 textos del español, desde 1200 a 1900. Para el siglo XX ha realizado cuatro divisiones de igual tamaño para los distintos registros —5 millones de palabras cada uno—: oral, ficción, periodismo y textos académicos. A raíz de él, Davies ha publicado *A frequency dictionary of Spanish: core vocabulary for learners* (2006), que recoge las 5.000 palabras más frecuentes del español. El índice principal del diccionario recoge los lemas, ordenados por frecuencia, con la indicación de la categoría gramatical, la traducción en inglés, un ejemplo extraído del corpus, el rango total de la frecuencia —*range count*— para indicar su presencia en los distintos bloques y el número total de veces que aparece —*raw count*—. Incorpora una indicación sobre el tipo de registro en que cada unidad aparece con más frecuencia. Algo

²⁰ Se trata del *Gran diccionario de uso del español* (2001), en el que se indica, por primera vez en la lexicografía española, la frecuencia de las palabras mediante la marcación con un índice numérico de frecuencia (del 1 al 5).

que llama la atención de este volumen es el intento, aunque sin base científica que explique la selección, de una organización temática del vocabulario, que el autor denomina *Thematic vocabulary* (“*call-out boxes*”) y que consiste en intentar agrupar en 30 temas (*body, food, family, weather, professions, nationalities, colors, emotions, verbs of movement and communication, etc.*) aquellas palabras que están dentro de las más frecuentes, aunque muchas de ellas caigan fuera de las 5.000. Así, a lo largo del texto tenemos una serie de cuadros con esta información ordenada por temas.

Recientemente contamos con el *Corpus del Español del Siglo XXI* (CORPES XXI)²¹, que al igual que el *Corpus de Referencia del Español Actual* (CREA) es un corpus de referencia construido por la Real Academia Española con el apoyo de las diferentes academias de la lengua española. Está formado por textos escritos y orales procedentes de España, América, Filipinas y Guinea Ecuatorial. La versión de abril de 2015 recoge algo más de 200 millones de formas y permite la posibilidad de recuperar el sonido alineado con textos orales y la consulta por categoría gramatical. La RAE y el Banco Santander han anunciado un convenio de colaboración para desarrollar este proyecto hasta 2018 con el objetivo de reunir un conjunto textual constituido por 400 millones de formas y palabras de la lengua común de casi 500 millones de hispanohablantes. Del mismo modo que hacía el CREA, se pretende ofrecer también las listas de las formas más frecuentes del español, con indicación de su frecuencia absoluta y normalizada. Por estar disponible para todo el público y realizado por una institución de prestigio, se le augura la misma difusión y reconocimiento que ha logrado el CREA.

El CAES (*Corpus de Aprendices de Español*) fue un encargo del Instituto Cervantes a un equipo de la Universidad de Santiago de Compostela dirigido por Palacios y Rojo²². Se trata de un corpus elaborado únicamente con muestras de estudiantes de ELE de diferentes países. El resultado, que está disponible para su consulta desde octubre de 2014, es un conjunto de unos 575.000 elementos

²¹ <http://www.rae.es/recursos/banco-de-datos/corpes-xxi>

²² CAES se puede consultar desde la web del Instituto Cervantes en <http://galvan.usc.es/caes>.

lingüísticos, caracterizados desde el punto de vista léxico y gramatical y clasificados teniendo en cuenta la L1, el país de origen, la edad y el sexo de los informantes. Los aprendices trabajaron directamente en un formulario preparado previamente en el que incluían sus datos personales y el texto, utilizando los ordenadores existentes en alguna de las salas del centro participante. El conjunto de textos producidos era enviado por la persona responsable al final de cada sesión, acompañado de un informe de incidencias. Así, cada texto producido se convertía automáticamente en un fichero en formato XML. Se aplicó a los ficheros el etiquetador *FreeLing*, de uso libre, que atribuye a cada forma ortográfica de un texto las etiquetas gramaticales que podrían corresponderle y trata de seleccionar la adecuada en cada caso, lo que se conoce como desambiguación. Dadas las características de los textos, el etiquetado automático tenía un porcentaje de errores muy elevado, de modo que la desambiguación manual fue laboriosa.

La aplicación de la lingüística del corpus a la docencia de ELE ha sido tratada y defendida en diversos trabajos²³. Una autora que se ha detenido en analizar sus virtudes ha sido Izquierdo Gil (2002), quien asegura que los trabajos estadísticos son una forma útil de selección de vocabulario, al margen de ciertas limitaciones pedagógicas que surgen al usar la frecuencia de aparición sin tener en cuenta la dispersión. Cuando hacemos esto no medimos el índice de uso, cuya virtud es reflejar si las palabras son usadas en diferentes campos o ámbitos. Y es en los niveles iniciales donde conviene afinar más en esta selección, por lo que el uso puede pasar a ser un criterio importante para ella. Propone usar los corpus para corregir las deficiencias de los léxicos de frecuencias o léxicos básicos.

De los corpus de frecuencias podemos extraer los léxicos básicos, seleccionando aquellas palabras más usuales y aplicando una criba que pondere su aparición en los diferentes dominios que conforman el corpus así como su pluralidad de contenidos (textos orales, escritos, diferentes registros, por tipología textual, etc.). El léxico básico es el conjunto de palabras más frecuentes de una lengua, y se configura mediante la aplicación de herramientas estadísticas a una muestra representativa de textos. La mayor parte de sus unidades se

²³ Benítez Pérez & Zebrowski 1993: 229; Carter 1996: 140, 181; Nation 1990: 16-20, 2001: 9; McCarthy 1990: 66-67; Sinclair & Renouf 1996: 148; Carter & McCarthy 1996: 210.

caracteriza por su condición de atemática, es decir, no especialmente vinculada a un campo conceptual concreto (Mas Álvarez & Santos Palmou 2014a: 2). Este tipo de léxico es el que encontramos en las obras que denominamos vocabularios básicos, que nacen de la necesidad de realizar la selección de la enorme cantidad de léxico con el que cuenta una lengua con vistas a su aplicación para la enseñanza a extranjeros. Estos vocabularios aspiran a ser una herramienta útil para facilitar el aprendizaje de una lengua extranjera en el menor tiempo posible y, en teoría, se encuentra en ellos el número mínimo de palabras que se necesitan para la comunicación básica. El primer intento importante fue el llamado *Basic English* (Ogden 1930).

Los diccionarios y las listas de frecuencias son fundamentales para el conocimiento cuantitativo del léxico, sin embargo, no siempre tienen una inmediata aplicación práctica en la enseñanza de la lengua, “pues son simples listas de palabras acompañadas de índices”, por lo que su posible utilidad didáctica “puede derivarse de su uso en forma de vocabularios básicos” (Fajardo Aguirre 1991: 217). Fajardo Aguirre consideraba que estos vocabularios eran aprovechables tal como estaban, con sus limitaciones, pero solo como medio de estudio memorístico. Afirmaba también que tenían otras posibilidades de aplicación en la enseñanza de español para extranjeros, que estaban por desarrollar. En concreto, orientaba sus críticas hacia su limitada posibilidad de utilización en el aula: “pueden ser material para el estudio en casa o instrumento de control a medio o largo plazo para comprobar el dominio del vocabulario, pero no son utilizables sin dificultades como instrumento de uso activo” (Fajardo Aguirre 1991: 221-2).

Comentaba, además, que su uso como diccionario elemental para traducciones de principiantes no es posible desde el momento en que se impuso la ordenación temática frente a la alfabética, aunque hoy en día, y teniendo en cuenta los medios electrónicos, los diccionarios básicos pueden y deben ofrecer las dos posibilidades de ordenación de las palabras, pues ambas son de utilidad. Pensemos por ejemplo en el de Davies, que brinda las dos opciones. Apuntaba, asimismo, un uso posible para trabajar el vocabulario en el aula, que era su “combinación con un libro de ejercicios adecuados sobre ese léxico fundamental, quizá orientados a trabajar teniendo en cuenta los campos léxicos”; tales

ejercicios podrían convertir, en su opinión, los vocabularios básicos en un material realmente aprovechable en la clase de lengua extranjera. Otra posibilidad de carácter didáctico para la que opinaba que todavía no se habían aprovechado eran para controlar el vocabulario básico que podría ser usado en la elaboración de definiciones de diccionarios. De esta manera serían más comprensibles y se facilitaría su manejo (Fajardo Aguirre 1991).

3.4.1.2. Puntos débiles de la frecuencia y los corpus textuales para la selección léxica

Hemos ido trazando hasta aquí cuáles son los puntos fuertes y los puntos débiles que ofrece el recurso a la frecuencia léxica para la selección del vocabulario. Centrémonos ahora en analizar las deficiencias que puede conllevar si la consideramos de forma exclusiva en la selección de los contenidos léxicos para la enseñanza de segundas lenguas (Izquierdo Gil 2003: 357-373; Salazar García 2004: 247-251; Bartol Hernández 2010: 90-91).

En primer lugar, hay diferencias importantes en los resultados entre los distintos recuentos, debidas a los tipos de textos que sirven para formar los corpus. Así, por ejemplo, en el *Léxico básico de Puerto Rico* no se encuentra *cuchara*, pero en el *Léxico del español hablado de Málaga* sí está. Puede parecer que se debe a la distancia geográfica, a la temporal y al tipo de criterios usados en uno y otro, pero los motivos van más allá y tienen que ver con la metodología y con las muestras del corpus sobre el que se ha trabajado.

En segundo lugar, como ya sabemos, las palabras más frecuentes son las palabras gramaticales. Los sustantivos ocupan un lugar mucho más bajo, y eso que son en su mayoría palabras con un gran poder designativo, lo que los hace imprescindibles en la comunicación. Entre las 100 palabras de mayor frecuencia y uso del *Léxico de frecuencia del español hablado en Málaga* los únicos sustantivos que aparecen son *año, cosa, casa, día, niño, vez, gente y tiempo*, y lo hacen en los lugares 45, 51, 62, 64, 79, 81, 90 y 94, respectivamente. Se constata, además, que algunas palabras muy útiles aparecen muy abajo en listados de frecuencia y no siempre estos listados proporcionan un orden adecuado en términos didácticos. No es válida la creencia de que el máximo de eficacia

pedagógica se consigue enseñando primeramente los términos más frecuentes. En algunos casos el orden en el que aparecen las palabras en los listados entra en conflicto con el orden lógico en el que se han de enseñar, por lo que además de seleccionarlas debemos considerar cómo las vamos a presentar (López-Mezquita 2007: 351). Además, mediante los recuentos de frecuencias priorizaremos la obtención de un tipo de unidades léxicas, mientras que otras van a estar fuera de esas 5 o 7 mil más frecuentes, que son las que se suelen considerar para un recuento de léxico básico. Así, las palabras frecuentes que se obtienen mayoritariamente suelen ser las consideradas vacías, las gramaticales, los nombres y adjetivos de significado muy general (como *cosa, casa, tiempo, lugar, persona, sitio...*). Esto es así independientemente del idioma en el que se elaboren. En inglés, por ejemplo, Nation (1990) sostiene que las listas de frecuencias a veces nos ofrecen una información incompleta o inapropiada debido a problemas como la ausencia entre las primeras mil o dos mil palabras de determinadas palabras importantes y necesarias en el léxico de un principiante. En concreto ejemplifica esto con los términos *soap, bath, chalk o stomach*, que no se encuentran entre las dos mil palabras más frecuentes del listado de Thorndike y Lorge (1944). El problema también sucede al contrario, es decir, entre las primeras palabras aparecen algunas que no son apropiadas para el vocabulario de un principiante, como *bill, company, deal, issue, labor, stock, supply, thee y thou*, que pueden parecer irrelevantes para una persona que está empezando a aprender inglés.

Quizás esta sea una de las grandes debilidades que presenta recurrir solo a la frecuencia como método de selección léxica, algo que se pretendía corregir con el uso de la metodología de la disponibilidad léxica. Los ejemplos de palabras que se consideran cotidianas y muy conocidas por cualquier hablante pero no aparecen entre las más frecuentes son múltiples. Un ejemplo altamente significativo es el que Gougenheim, Michéa, Rivenc y Sauvageot (1964: 138-139) usaron para ilustrar este tipo de vocablos que son muy conocidos por los hablantes pero pueden ser poco usados fuera de un contexto concreto: la palabra *fourchette* (“tenedor”). También lo percibió Paredes García (2015), quien rescata el ejemplo y apunta que en el *CREA* la palabra *tenedor* ocupa la posición 17.942, lo que evidencia que estaría lejos de introducirse en la enseñanza del español si

este fuera el único principio para seleccionar vocablos. Por consiguiente, el porcentaje de uso de las unidades léxicas en los corpus no siempre es sinónimo de utilidad didáctica, ya que hay otras palabras que no alcanzamos con este criterio que tendremos que perseguir con ayuda de otros, entre ellos la disponibilidad léxica.

Al mismo tiempo, existen determinadas cuestiones metodológicas que la lingüística del corpus debe actualizar y perfeccionar para conseguir responder a algunas necesidades para las que todavía no ofrece demasiadas soluciones. En concreto, sería útil etiquetar todo tipo de unidades léxicas, no solo las palabras gráficas, para poder obtener también las pluriverbales y rescatar las colocaciones. Como hemos comentado, estas resultan muy importantes a la hora de aprender una lengua extranjera, ya que es una cuestión que presenta especiales dificultades. Si los corpus nos permitieran saber cuáles son las colocaciones más frecuentes, podríamos incorporarlas con mayor eficacia desde un inicio y graduar así también su aprendizaje.

3.4.2. La disponibilidad léxica

Como vemos, la utilidad de las unidades léxicas no siempre viene determinada por la frecuencia o el índice de uso. La ausencia de palabras de uso común en los listados de frecuencia fue lo que llevó a proponer los estudios de disponibilidad léxica (DL), que han tenido gran seguimiento en España e Hispanoamérica. El origen de estos estudios lo constituyó un grupo de lingüistas franceses (Gougenheim, Michéa, Rivenc & Sauvageot 1964) al diseñar un método básico para la enseñanza del francés a extranjeros, que en aquel momento eran tanto los habitantes de antiguas colonias como los inmigrantes que llegaban a Francia. Recurrieron en principio a recuentos léxicos basados en la frecuencia para determinar el léxico elemental del francés. Sin embargo, durante este proceso se dieron cuenta de que los resultados eran parciales, ya que en los recuentos no aparecían muchas palabras muy normales y necesarias en el uso diario y que se referían a realidades concretas. Estas no aparecían en los listados basados en la frecuencia porque su uso estaba muy condicionado por el tema de la conversación. Palabras que refieren realidades cotidianas como *ducha* y

cepillar(se) empleadas habitualmente por cualquier hablante nativo no son frecuentes, pero son disponibles si se activa un centro de interés como “aseo personal” o “acciones cotidianas”, por ejemplo. Estas últimas fueron llamadas más tarde por Michéa “palabras temáticas”, frente a las anteriores que eran “atemáticas” (Michéa 1953). Recordemos que tras esta etiqueta de “atemáticas” se encuentran las palabras gramaticales, un gran número de adjetivos, de verbos comunes, algunos nombres muy generales y la mayor parte de los adverbios. El mérito de la puesta a punto de la noción de disponibilidad se debe a Michéa, quien en su artículo de 1953 “Mots fréquents et mots disponibles” expone los principios que regirán las investigaciones sobre disponibilidad. Este investigador notó que la inestabilidad de dicha parcela de vocabulario, especialmente formada por sustantivos concretos en las listas de frecuencia, no se podía corregir mediante el aumento de las muestras introducidas, ya que estos términos escapaban a la estadística general. Michéa llegó a la conclusión de que las palabras estadísticamente más inestables estaban vinculadas a áreas temáticas o situaciones específicas y que encontraban su estabilidad si se utilizaba como punto de partida las situaciones mismas.

Al dar este primer paso, se descubrió que para hallar las palabras temáticas era necesario implementar su método con pruebas asociativas, pues era la forma de llegar a estimular el lexicón mental de los informantes y obtener de ellos las palabras, ya no solo más frecuentes, sino también más disponibles cuando las circunstancias comunicativas lo exigen. Así, se deslindaron los conceptos de “palabras frecuentes” y “palabras disponibles” y surgieron los cimientos de la distinción hoy establecida entre léxico básico (LB), léxico disponible (LD) y léxico fundamental (LF)²⁴, tres conceptos bien diferenciados que constituyen a su vez otras tres líneas temáticas profundamente imbricadas, pero con objetivos y aplicaciones diferentes (Mas Álvarez & Santos Palmou 2014a).

A partir de entonces se empezaron a deslindar conceptos tales como palabras frecuentes, palabras comunes y palabras usuales, que hasta la fecha se

²⁴ A los que también podemos denominar vocabulario básico, vocabulario disponible y vocabulario fundamental, respectivamente.

habían manejado como sinónimos (López Morales 1999b: 10). Los estudios de léxico disponible o de disponibilidad léxica nacieron, pues, como complemento de los léxicos básicos o de frecuencia, cuando los investigadores intentaron trascender las limitaciones de éstos últimos. Las palabras temáticas y atemáticas se relacionarán respectivamente con lo que Michéa denominará vocabulario de interés (disponible) y vocabulario general (frecuente).

En el mundo hispánico, tras el estudio pionero de López Morales (1973), las recopilaciones de léxico disponible han conocido a partir de los años 90 un gran desarrollo, tanto desde un punto de vista cuantitativo como cualitativo. Los diferentes equipos están coordinados en torno al denominado Proyecto Panhispánico de Estudio sobre la Disponibilidad Léxica (PPHDL), dirigido por el profesor López Morales²⁵. Ofrecemos una información más amplia y detallada sobre la disponibilidad léxica y sobre este proyecto en el Capítulo 6 (*vid. infra* § 6.1., 6.2. y 6.3.)

3.4.2.1. Ventajas de la disponibilidad léxica en la selección del vocabulario

Hay, por tanto, palabras muy útiles pero poco frecuentes, mientras que muchas de las palabras más frecuentes son muy poco precisas y con escaso valor informativo, lo que necesariamente les resta rentabilidad. Ya hemos comentado una experiencia que Bogaards desarrolló para medir el peso informativo que estos dos tipos de palabras tienen en los textos al borrar unas y otras. Con ello demostró que las palabras plenas hacen que el texto sea comprensible en gran medida, mientras que las más frecuentes son las que más fácilmente se pueden rescatar (1994: 116-119). También Izquierdo Gil (2004: 360) subraya que la frecuencia no puede ser el único criterio en el que basar la selección léxica del material destinado a aprendices, puesto que estos necesitan conocer palabras

²⁵ Puede obtenerse información completa y detallada sobre los objetivos del PPHDL, los equipos de investigación que participan, el estado en que se encuentran las investigaciones en los diferentes territorios y los resultados de las mismas en DISPOLEX, el sitio web creado y mantenido desde el 2003 hasta la actualidad por Borrego Nieto, Bartol Hernández, Fernández Juncal y Hernández Muñoz (<http://www.disporex.com/info/el-proyecto-panhispanico>). La página recoge también la bibliografía sobre el proyecto y la disponibilidad léxica en general, así como información acerca de otros proyectos próximos a esta orientación disciplinar. Pueden verse más detalles en Bartol Hernández (2007) y Hernández Muñoz & Bartol Hernández (2006).

con alto contenido informativo para poder comunicarse. Y los estudios sobre disponibilidad sirvieron precisamente para poner de relevancia estas palabras útiles, aunque poco frecuentes. En contraposición con lo que sucede con el léxico frecuente, el léxico disponible nos ofrece principalmente palabras con fuerte carga semántica: sustantivos, en mucha menor medida verbos y adjetivos.

Al margen de otras diferentes aplicaciones de la disponibilidad léxica (a la sociolingüística, dialectología, psicolingüística y etnolingüística) diversos investigadores han resaltado la proyección pedagógica de estos trabajos tanto en la enseñanza del español como lengua materna como del español como lengua extranjera²⁶. Sostienen que los léxicos disponibles constituyen valiosos instrumentos objetivos que, junto a los léxicos básicos o de frecuencias, facilitan la planificación o programación de la enseñanza del vocabulario del español como lengua extranjera, ya que la información que recogen los trabajos sincrónicos muestran cuáles son las palabras más usuales en cada centro de interés temático, al tiempo que marcan el entorno comunicativo al que corresponden (Mateo García 1998: 55). Sirven, también, para caracterizar el vocabulario de una determinada comunidad, para conocer los rasgos que en los niveles diatópico y diastrático los diferencian de otros grupos de hablantes de la misma lengua, así como para la aplicación a la enseñanza de idiomas (Carcedo González 1997-1998: 76). Benítez Pérez (1992; 1994a: 12) ya había defendido claramente la utilidad pedagógica de los materiales extraídos de dichas investigaciones. Según este investigador, después de concluir la recogida y organización del léxico disponible de los diferentes países del ámbito hispano, dispondremos de una base de datos donde acudir a la hora de programar cursos, preparar ejercicios o elaborar manuales y materiales de apoyo. En la actualidad algunos de los estudios de disponibilidad llevados a cabo en diversas provincias españolas han dado sus frutos:

[...] podemos saber cuál es el léxico que los estudiantes deben aprender y en qué

²⁶ Entre otros, López Morales (1990: 53; 1999b: 19-20); Benítez Pérez (1992, 1994a, 1994b, 1995, 2009); López Morales & García Marcos (1995); Mateo García (1998: 55); Carcedo González (1997-1998, 1998a, 1998b, 1999a, 1999b, 2000, 2001, 2003, 2004); García Marcos & Mateo García (2003); Salazar García (2004); Bombarelli (2005); Medina Arejita (2009); Bartol Hernández (2010); Madison (2013), Fernández Leyva (2015); Paredes García (2015); Ávila Muñoz (2016); Jiménez Catalán (2017).

prioridad deben hacerlo y, en consecuencia, en qué formas léxicas deben hacer hincapié los maestros y profesores y cuáles deben aparecer en los manuales. [Benítez Pérez 1992: 102]

En este sentido, se puede afirmar que los resultados de los estudios de disponibilidad léxica nos proporcionan una base estadística sólida para efectuar la selección y graduación de las unidades léxicas que debemos enseñar y se podrá reducir así la arbitrariedad del vocabulario introducido en los manuales o en el aula que este mismo autor había detectado.

Tener en cuenta los listados panhispánicos puede suponer la eliminación en los manuales de todo el léxico superfluo, de lujo, a favor de aquel que vive en nuestra comunidad idiomática, como demuestran con claridad estos estudios de léxicoestadística. [Samper Padilla, Bellón Hernández & Samper Hernández 2003]

Bartol Hernández indicaba en un artículo reciente que el léxico disponible supone un nuevo enfoque, por no tratarse de las palabras utilizadas al producir determinados textos, sino de las palabras que están en nuestro lexicón mental organizadas y que están “disponibles” para ser utilizadas cuando las circunstancias comunicativas lo requieran, cuando un estímulo active el nodo cerebral adecuado (Bartol Hernández 2010: 91). Y continúa enumerando los beneficios que la disponibilidad léxica aporta a la selección del vocabulario. El primero de ellos es compartido con la frecuencia, ya que ambas vías aspiran a ofrecer una metodología empírica y objetiva basada en corpus y fórmulas matemáticas que reduzcan la arbitrariedad en el proceso de selección, como Benítez Pérez había indicado (1994a y 1994b).

Estas pruebas de producción léxica también han resultado muy útiles para medir el dominio léxico activo de los estudiantes de ELE. Esta línea de investigación ha sido iniciada dentro de los estudios de disponibilidad léxica por Carcedo González (1999a, 2000) y Samper Hernández (2001, 2002) y seguida luego por otros muchos. Algunos han realizado encuestas de disponibilidad a alumnos de ELE y, posteriormente, han comparado los resultados con los listados producidos por nativos y, por ejemplo, se han detectado errores en el proceso de aprendizaje, como los derivados de la influencia de la lengua materna.

El hecho de usar fórmulas matemáticas —se conjugan la frecuencia y el

orden de activación de la palabra para hallar el índice de disponibilidad²⁷, la aparición, frecuencia absoluta, frecuencia relativa, frecuencia acumulada, entre otros— permite establecer cortes según el grado de disponibilidad, que se pueden emplear como criterio para escoger el léxico de los diferentes niveles educativos.

Otra de las ventajas de la disponibilidad léxica es que nos da una información muy valiosa sobre la organización de la memoria semántica y el lexicón mental, en la línea de las asociaciones mentales de la psicología cognitiva (Izquierdo Gil 2004). En este sentido, la disponibilidad léxica agrupa el léxico que se recoge en campos nocionales, lo que en disponibilidad léxica se llama centros de interés (CI)²⁸. En *L'élaboration du français fondamental*, Gougenheim *et al.* (1964: 152-153) pidieron a 904 adolescentes (de 9 a 12 años) de 4 provincias francesas que escribieran las primeras 20 palabras que se les ocurrieran de forma inmediata para cada uno de los 16 centros de interés. Suponían que estos representaban los sectores semánticos más universales, es decir, los más comúnmente compartidos por todos los hablantes, cualquiera que fuera su procedencia social, geográfica, etc. Si los comparamos con los ámbitos léxicos enunciados en el *MCERL*, observamos que los centros de interés de los léxicos disponibles se asemejan a los temas y subtemas contemplados, con sus proximidades y diferencias, pues hay temas y subtemas del Consejo de Europa que no aparecen entre los centros de interés. Organizar el vocabulario en campos nocionales también ayuda a favorecer su aprendizaje, en la medida en que tiene

²⁷ El pilar fundamental de la disponibilidad léxica lo constituye el “índice de disponibilidad” léxica de cada vocablo (ID). Se interpreta como el grado de accesibilidad de un vocablo en el lexicón cuando el tema de la conversación lo requiere, y se observa en los listados como un valor que se expresa entre 1 y 0 (donde 1 es disponibilidad máxima y 0 es disponibilidad mínima). Mientras que la frecuencia solo tiene en cuenta el número de veces que una palabra aparece en los listados, el ID pondera los datos a partir no solo del número de veces que aparece un término, sino también de la posición que ocupa en las listas.

²⁸ Los centros de interés clásicos que fueron usados en las investigaciones sobre la disponibilidad léxica del francés por Gougenheim *et al.* (1956, 1964) y Gougenheim (1958) son:

01, Partes del cuerpo; 02, La Ropa; 03, Partes de la casa (sin muebles); 04, Los muebles de la casa; 05, Alimentos y bebidas; 06, Objetos colocados en la mesa para la comida; 07, La Cocina y sus utensilios; 08, La escuela: muebles y materiales; 09, Iluminación, calefacción y medios de airear un recinto; 10, La ciudad; 11, El Campo; 12, Medios de transporte; 13, Trabajos de campo y del jardín; 14, Los Animales; 15, Juegos y distracciones; 16, Profesiones y oficios.

más que ver con la organización de nuestro lexicón mental. Además, se corrigen considerablemente los problemas planteados en los recuentos de léxicos básicos por la homonimia y la polisemia, pues con las palabras ubicadas dentro de un campo de sentido queda más claro su significado en ese contexto. Los listados obtenidos son como una fotografía de la vitalidad de las palabras en un determinado lugar y tiempo.

Sabemos, además, que las palabras más disponibles son palabras muy estables de la lengua. Y esto se debe a que la disponibilidad de una palabra tiene que ver fundamentalmente con los siguientes parámetros (Hernández Muñoz 2005): la tipicidad —las encuestas de disponibilidad léxica ilustran claramente la categorización humana desde el punto de vista de los prototipos, la edad de adquisición y la familiaridad.

No se ha determinado por ahora cuál es el peso de cada parámetro en la disponibilidad de cada una de las palabras, pero sí se puede afirmar que la tipicidad y la edad de adquisición son las responsables de la unidad del léxico, así como de las semejanzas que se establecen entre el léxico de diferentes zonas, lo que se observa en las coincidencias en las respuestas de sujetos muy alejados geográficamente. Estos parámetros inciden en lo que es común y compartido por los miembros de la comunidad. De hecho, hay psicolingüistas que afirman que los miembros de una misma comunidad comparten muchos de los rasgos de su organización conceptual y que la realidad psicológica que posibilita los niveles de categorización es compartida por todos los seres humanos y se manifiesta por igual en todas las culturas. Este hecho tendría una gran repercusión en la comparación de léxicos disponibles de diferentes lenguas y, en especial, en las zonas bilingües. El grado de familiaridad de las palabras puede variar, de hecho varía en función de las zonas, de las épocas y de los individuos. A tal variación se deben respuestas como *t-shirt* en el léxico frecuente de Málaga (Ávila Muñoz 1999) o en el de Castilla-La Mancha en lugar de *camiseta*, como seguramente obtenemos si preguntamos a informantes de más edad, de una zona rural o de hace varias décadas.

El uso de una misma metodología, en especial en los trabajos que integran el Proyecto Panhispánico de Estudio sobre la Disponibilidad Léxica, permite la comparación diatópica, lo que nos ofrece datos sobre la unidad y variación léxica

del español. Otros datos que muestran son las variaciones de carácter social, sociolingüístico o las del eje diacrónico. Los listados aportan también interesantes datos sobre la influencia de las variables sociales en el dominio léxico de los informantes. La variable que más influencia parece tener sobre el dominio léxico es el nivel cultural de la familia, más que la variable sexo, o la pertenencia a un ambiente urbano o rural. La comparación entre las distintas variantes de una variable nos informa de la diferente integración de algunas palabras en la sociedad.

En general, los hechos que avalan la fiabilidad de las pruebas de asociación léxica son su semejanza con las pruebas usadas en psicolingüística, así como los modelos de producción léxica que las sustentan (Caramazza 1997; Levelt 1999, 2001; Levelt, Roelofs & Meyer 1999) y, en concreto, los ejercicios de fluencia verbal semántica utilizados en psicología experimental clínica para diagnosticar problemas en la memoria semántica de los individuos (Tomé Cornejo 2010).

Si bien los estudios de disponibilidad léxica sirven de base para la selección del vocabulario de la L1, principalmente en el nivel elemental, también se pueden utilizar para evaluar el dominio del léxico activo en estudiantes de una lengua extranjera, para detectar errores en el proceso de aprendizaje y para verificar la influencia de la lengua materna. En esta línea, Echeverría (2002; Echeverría, Unzúa y Figueroa 2005) diseñó un programa informático llamado *Vocabulario Disponible* basado en su investigación con 2.100 estudiantes chilenos sobre 18 centros de interés. Con él somete a los alumnos a una encuesta léxica para elaborar un diagnóstico sobre su rendimiento en cada centro de interés por separado. Luego, el programa compara los datos ofrecidos por el alumno con los del promedio nacional y emite un diagnóstico. Pero no solamente le informa de sus aciertos, sino que también le señala los vocablos más disponibles que él ha omitido²⁹. Detrás de este programa están las ideas de las

²⁹ Se puede consultar una descripción de este proyecto en el sitio <http://www.c5.cl/ieinvestiga/actas/tise99/html/papers/metacognicion/>, donde se indica que el “software se encuentra en etapa de evaluación, así como también lo está el efecto psicolingüístico del mismo en los sujetos estudiados”.

teorías metacognitivas, en las que se considera muy positivo que el sujeto tenga plena conciencia del proceso de enriquecimiento (del qué, del cómo y del cuánto está aprendiendo). En el sitio web diseñado por el propio Echeverría se indica la utilización pedagógica del *Vocabulario Disponible*:

una herramienta pedagógica flexible. Instalado en estaciones autónomas, los estudiantes pueden trabajar en parejas. El profesor de una asignatura cualquiera puede motivarlos con respecto a un tema de tipo ecológico, por ejemplo, y luego enviarlos a trabajar con el programa en el centro *Problemas del ambiente*. Los juegos de léxico correspondientes no solamente harán consciente al estudiante de su léxico disponible en el área sino también despertarán en él variadas interrogantes gatilladas por el vocabulario allí encontrado. Si el maestro está motivado y tiene tiempo, podrá ingresar nuevos ítemes y nuevos centros de interés que le sean de particular importancia pedagógica. [Echeverría 2002]

El programa se pensó para ser usado con adolescentes que tienen el español como lengua materna, pero perfectamente podría aplicarse a alumnos de español como lengua extranjera (nivel intermedio o avanzado). Este mismo autor y otros colegas elaboraron también otro tipo de herramienta computacional para analizar las relaciones semánticas entre el vocabulario disponible, denominada *Dispografo* (Echeverría, Vargas, Urzúa & Ferreira 2008).

Hay otros trabajos en los que se comparan las encuestas de disponibilidad pasadas a los alumnos de español con los resultados obtenidos en hablantes nativos españoles. Por ejemplo, uno de los primeros fue Carcedo González (1998a, 1998b, 1999a), quien determinó con estudiantes finófonos de español los vacíos léxicos y las necesidades léxicas específicas de su alumnado extranjero:

[...] mediante la comparación de aquél [el vocabulario disponible del alumnado extranjero] con los léxicos disponibles de diferentes partes del Mundo Hispánico se detectarían las deficiencias: unidades poco productivas, ausencia de términos importantes, etc., datos todos ellos relevantes a la hora de realizar una programación eficaz de la enseñanza del léxico. [Carcedo González 1999a: 76]

Por su parte, Izquierdo Gil (2004) indica que la aplicación de la disponibilidad léxica como criterio de selección del vocabulario de los campos semánticos adquiere importancia para el desarrollo de la competencia referencial. Este componente de la competencia léxica hace alusión tanto al estudio de los campos semánticos (Galisson 1991: 160-166), como a las cadenas léxicas que describen un *script* o comportamiento cotidiano estereotipado.

Son evidentes, pues, las utilidades de los estudios dedicados al léxico disponible: su utilidad para la medición de la competencia léxica de los individuos, la reparación de carencias si estas fueran detectadas y la planificación de la enseñanza del vocabulario. Sin embargo, siempre se han encontrado dos deficiencias. La primera, el hecho de que el léxico disponible es, por definición, *temático*, esto es, de aparición ligada a un tema determinado. Por este motivo, si al comentar el tipo de palabras que se obtenían con los recuentos de frecuencia revelábamos su parcialidad, lo mismo ocurre con las del léxico disponible. En el primero, obteníamos las palabras vacías, gramaticales o sustantivos y adjetivos muy generales y con poca carga significativa. En este sentido, el léxico disponible es mucho más concreto, específico y con mayor valor comunicativo, pero es igualmente un resultado parcial de nuestro bagaje léxico: no conseguimos obtener mediante esta metodología una fotografía completa y totalmente eficaz de nuestra competencia léxica. Una de las deficiencias, que se observó ya desde el inicio de esta línea de trabajo, es que se obtienen sobre todo sustantivos, algunos menos adjetivos y pocos verbos, pero no otras unidades léxicas. Las primeras encuestas de léxico disponible dejaban espacio para estas tres categorías gramaticales en cada centro de interés, por lo que era de esperar que se cubrieran las tres. Se consigue paliar así el efecto de recolectar casi exclusivamente sustantivos. No obstante, pedían a los encuestados un número máximo de elementos en cada columna, en lugar de optar por una limitación temporal, como se prefiere en la actualidad, por lo que se restaba objetividad a la prueba. Otra cuestión es el hecho de que, efectivamente, se pueden observar las relaciones semánticas que se establecen entre respuestas, su proximidad en la posición en la que se nombran, etc., pero no se obtiene información sobre las relaciones sintagmáticas de las palabras ni sobre colocaciones (léxicas y gramaticales), como tampoco sobre fraseología.

El segundo problema es la selección de los temas utilizados como estímulo en las pruebas, lo que se conoce técnicamente como centros de interés (CI). No vamos a entrar en los problemas y debates que los centros de interés han generado aplicados a las investigaciones con estudiantes de ELE, pues para llevar a cabo la selección léxica del vocabulario que nos sirva como referente en este proceso de enseñanza-aprendizaje, es más lógico aplicar de partida los

resultados obtenidos con hablantes nativos de español³⁰, como ha mostrado en su tesis doctoral Tomé Cornejo (2015). La mayor parte de los estudios de disponibilidad léxica han usado los 16 centros de interés clásicos que establecieron Gougenheim *et al.* (1964) con la intención de abarcar la parcelas más importantes del universo léxico de los hablantes. Sobre todo han seguido estas pautas los estudios integrados en el Proyecto Panhispánico de Estudio sobre la Disponibilidad Léxica (PPHDL), pues esto permite las comparaciones entre los resultados de diferentes áreas geográficas. También en los estudios de disponibilidad léxica aplicados a estudiantes de ELE se empezó a trabajar con estos centros de interés, como en los dos trabajos pioneros en aplicar esta línea, los de Carcedo González 1998a, 2000 y Samper Hernández 2002, aunque fueran realizados en diferentes contextos, ya que el primero lleva a cabo su estudio en un país extranjero (Finlandia) y la segunda en un contexto de inmersión de una ciudad española (Salamanca)³¹.

El número de centros de interés establecido de inicio pretendía abarcar todas las áreas semánticas en las que se divide nuestro mundo, es decir, intentaba encasillar en los 16 centros de interés todos los intereses humanos. Para ello se buscaron enunciados muy universalistas y semánticamente muy

³⁰ Para ver estas cuestiones con mayor detenimiento pueden consultarse las siguientes referencias: Medina Arejita (2009), Mas Álvarez & Santos Palmou (2014a y 2014b), el artículo sobre este tema de González Fernández (2014) en el n.º16 de la *Revista Nebrija de Lingüística aplicada a la enseñanza de lenguas*, así como dos repuestas en comentarios de Bartol Hernández y Paredes García en el mismo número.

³¹ Como hemos expuesto en otro trabajo (Mas Álvarez & Santos Palmou 2014b), este tipo de estudios —muchos de ellos trabajos académicos, como tesis o memorias de máster— se están produciendo profusamente, siguiendo la línea de estas dos primeras propuestas. Sirvan como muestra los puntos geográficos que han sido analizados hasta ahora en el extranjero —Finlandia, Islandia, Polonia, Turquía, China, Marruecos, Eslovenia, Rumanía, Alemania, Portugal— y los lugares de la geografía española en los que se han hecho encuestas a estudiantes extranjeros en un contexto de inmersión lingüística —Valladolid (trabajando con personas inmigrantes), Salamanca (centro de idiomas de la Universidad), Alcalá de Henares (centro de idiomas de la universidad), Sevilla y Cádiz—.

En un primer momento, se toma como referente en cuanto a la metodología la establecida en el PPHDL, pues es la seguida por el amplio número de investigadores que se encuadran en él y por numerosas publicaciones. Esta continuidad sirve, además, para poder hacer la comparación entre sintopías, muy útil para el cotejo de resultados de diferentes lugares, tanto teniendo en cuenta a hablantes solamente extranjeros o entre extranjeros y nativos. Aun así, ya desde un primer momento se incluyeron nuevas variables, necesarias por el nuevo tipo de informantes con el que se trabajaba. Por otra parte, una línea interesante, aunque todavía poco explotada, en la aplicación de la metodología de la disponibilidad léxica a la docencia de ELE es la que se lleva a cabo en el ámbito de la enseñanza del español con fines específicos, teniendo en cuenta la relevancia del vocabulario técnico.

amplios, lo cual puede ya de por sí constituir también un problema, pues limita lo que cada cultura ofrece en concreto como diferencial (y que es interesante para los estudios de la etnolingüística y la didáctica de lenguas). También acaban por ser muy generales, lo cual hace que el número de respuestas diferentes sea elevado. Esto se refleja en otro de los índices matemáticos que miden y analizan las respuestas: el índice de dispersión. Este nos informa del grado de coincidencia en las respuestas entre informantes y entre centros de interés, ya que se determina dividiendo la frecuencia total de respuestas entre el número de ejemplos producidos. Así estudiamos la semejanza o no en el número de palabras producidas y conseguimos saber si son cerrados o difusos. Los centros compactos o cerrados, que son los poco difusos, tienen sus ventajas y desventajas, y con los que son más abiertos ocurre lo mismo, por lo que hay que definir qué tipo de centros de interés nos aportan lo que queremos en cada caso.

Esto nos lleva a la cuestión de que los centros de interés son muy diferentes entre sí. Algunos tienen una relación de inclusión con respecto a sus respuestas y otros establecen lazos relacionales de diferente índole. Por lo tanto, el tipo de centros de interés y su forma de enunciarlos también deben ser revisados para ajustarlos a nuestros objetivos. En primer lugar, sus nombres no siguen el mismo esquema; en segundo, los hay que son binarios, por lo que se ha demostrado que los resultados varían en función del orden de sus dos elementos al enunciarlos en las encuestas (denominado efecto *priming*). Para sortear estos problemas, se ha propuesto la utilización de guiones de comportamiento (*scripts*) como estímulos temáticos (Tomé Cornejo 2012).

Otra cuestión a resolver sería la del número de centros, pues ya desde un inicio se vio que no cumplían con el objetivo de abarcar todos los ámbitos de la realidad humana. Así pues, algunos trabajos han optado por mantener los 16 clásicos y añadir algunos más, bien para obtener más léxico que no se había considerado en estos centros de interés iniciales, bien para contemplar ámbitos recientes de nuestra realidad, también para intentar reflejar otras categorías gramaticales diferentes al sustantivo, o bien para trabajar el vocabulario específico. Con todo, un problema que puede aparecer al ampliar esta nómina de centros de interés es que sean demasiados, hasta el punto de alargar las pruebas y que los informantes no consigan realizarlas con éxito o tarden demasiado

tiempo. Una propuesta en este sentido ha sido la de cambiar de informantes y hacer encuestas diferentes a unos y otros (Tomé Cornejo 2011).

Si se estudian los recuentos basados en la frecuencia de las palabras y se comparan con los listados del léxico disponible, se observa que faltan palabras muy disponibles en los primeros, que por ser de uso común cuando se activan determinados temas (por ejemplo, al hablar de la ropa, *calzoncillo*, *calcetín* o *braga*) deberían estar como palabras objeto de estudio en los niveles iniciales o intermedio. Así lo resaltan algunas comparaciones que se han hecho (Benítez Pérez 1994a, 1994b, 1995, 2009; Benítez Pérez, P. & Zebrowski 2003; Bartol Hernández 2010), que ponen en evidencia la utilidad de la disponibilidad léxica para corregir estas deficiencias en la selección léxica. No obstante, si comparamos las palabras más disponibles, y aunque comparemos listados de diferentes lugares del ámbito hispánico para intentar que se complementen, observamos que no aparecen palabras que en el *Plan Curricular del Instituto Cervantes (PCIC)* se indica que deben ser enseñadas en los niveles iniciales. Por una parte esto se debe a que los 16 centros de interés no consiguen cubrir los temas del Consejo de Europa y las nociones específicas del *PCIC*, y los que sí coinciden no lo hacen al cien por cien. Bartol Hernández (2010: 95) muestra que es posible recurrir a la disponibilidad para buscar las palabras de la mayor parte de los temas del Consejo de Europa, pues hay una gran equivalencia. Partimos de su propuesta para elaborar esta tabla comparativa a la que hemos añadido las nociones generales del *PCIC*. Como se puede observar, efectivamente se pueden encontrar equivalencias al cotejar las tres columnas. No obstante, si partimos del *PCIC* como el referente para elaborar la planificación de un curso de ELE general, se observa que los centros de interés solo cubren parte de los aspectos que entran en cada noción.

Centros de interés clásicos (disponibilidad léxica)	Temas y subtemas del Consejo de Europa	Nociones específicas del <i>Plan Curricular del Instituto Cervantes</i>
01. Partes del Cuerpo	7.1. Partes del cuerpo	1. Individuo: dimensión física (Partes del cuerpo, Características físicas, Acciones y posiciones que se realizan con el cuerpo, Ciclo de la vida y reproducción).
02. La ropa	9.3. Ropa y moda	16. Compras, tiendas y establecimientos (Lugares, personas y actividades, Ropa, calzado y complementos, Alimentación, Pagos).
03. Partes de la casa	2. Vivienda	10. Vivienda (Acciones relacionadas con la vivienda, Características de la vivienda, Actividades domésticas, Objetos domésticos).
04. Los muebles de la casa	2. Vivienda	10. Vivienda (Acciones relacionadas con la vivienda, Características de la vivienda, Actividades domésticas, Objetos domésticos).
05. Comidas y bebidas	10. Comidas y bebidas	9. Alimentación (Dieta y nutrición, Bebida, alimentos, recetas, platos, utensilios de cocina y mesa, restaurante).
06. Objetos colocados en la mesa para la comida		9. Alimentación (Dieta y nutrición, Bebida, alimentos, recetas, platos, utensilios de cocina y mesa, restaurante).
07. La cocina y sus utensilios	2. Vivienda	10. Vivienda (Acciones relacionadas con la vivienda, Características de la vivienda, Actividades domésticas, Objetos domésticos).
08. La escuela	8. Educación	6. Educación (Centros e instituciones educativas, Profesorado y alumnado, sistema educativo, aprendizaje y enseñanza, exámenes y calificaciones, estudios y titulaciones, lengua de aula, material educativo y mobiliario de aula).
09. Iluminación, calefacción y medios de airear un recinto.		10. Vivienda (Acciones relacionadas con la vivienda, Características de la vivienda, Actividades domésticas, Objetos domésticos).
10. La ciudad	9. Compras 11. Servicios públicos 12. Lugares	20. Geografía y naturaleza (Universo y espacio, Geografía, Espacios urbanos o rústicos, Clima y tiempo atmosférico, Fauna, Flora).
11. El campo	12. Lugares 2.8. Flora y fauna	20. Geografía y naturaleza (Universo y espacio, Geografía, Espacios urbanos o rústicos, Clima y tiempo atmosférico, Fauna, Flora).
12. Medios de transporte	5.1. Transporte público 5.2. Transporte privado	14. Viajes, alojamiento y transporte (Viajes, Alojamiento, Sistema de transporte).

Centros de interés clásicos (disponibilidad léxica)	Temas y subtemas del Consejo de Europa	Nociones específicas del <i>Plan Curricular del Instituto Cervantes</i>
13. Trabajos del campo y el jardín		20.Geografía y naturaleza (Universo y espacio, Geografía, Espacios urbanos o rústicos, Clima y tiempo atmosférico, Fauna, Flora).
14. Animales	2.8. Flora y fauna	20.Geografía y naturaleza (Universo y espacio, Geografía, Espacios urbanos o rústicos, Clima y tiempo atmosférico, Fauna, Flora).
15. Juegos y diversiones	4. Tiempo libre, entretenimiento	8.Ocio (Tiempo libre y entretenimiento, Espectáculos y exposiciones, Deportes, Juegos).
16. Profesiones y oficios.	1.10. Profesión, trabajo	7.Trabajo (Profesiones y cargos, Lugares, herramientas y ropa de trabajo, Actividad laboral, Desempleo y búsqueda de trabajo, Derechos y obligaciones laborales, Características de un trabajador).

TABLA 4. Comparación entre los centros de interés de la disponibilidad léxica, los temas y subtemas del Consejo de Europa y las nociones específicas del *PCIC* (Bartol Hernández 2010: 95)

Por otra parte, si observamos la tabla de los 20 puntos que establece el *PCIC* en sus nociones específicas, se advierte que muchos de ellos no encuentran equivalente con ningún centro de interés de los 16 clásicos y que, en otros, solo lo encuentran parcialmente. A continuación ofrecemos la lista de las 20 nociones específicas del *PCIC* para poder hacer esta comparación. En ella resaltamos en versalitas lo que no hemos podido cubrir con ningún centro de interés y en letras minúsculas subrayadas lo que creemos que sí se podría cubrir con los datos de la disponibilidad léxica.

1. Individuo: dimensión física (Partes del cuerpo, CARACTERÍSTICAS FÍSICAS, ACCIONES Y POSICIONES QUE SE REALIZAN CON EL CUERPO, CICLO DE LA VIDA Y REPRODUCCIÓN).
2. INDIVIDUO: DIMENSIÓN PERCEPTIVA Y ANÍMICA (CARÁCTER Y PERSONALIDAD, SENTIMIENTOS Y ESTADOS DE ÁNIMO, SENSACIONES Y PERCEPCIONES FÍSICAS, VALORES PERSONALES).
3. IDENTIDAD PERSONAL (DATOS PERSONALES, DOCUMENTACIÓN, OBJETIVOS PERSONALES).
4. RELACIONES PERSONALES (RELACIONES FAMILIARES, RELACIONES SOCIALES, CELEBRACIONES Y ACTOS FAMILIARES, SOCIALES Y RELIGIOSOS).
5. Alimentación (DIETA Y NUTRICIÓN, Bebidas, Alimentos, RECETAS, PLATOS, Utensilios de cocina y mesa, RESTAURANTE).
6. Educación (CENTROS E INSTITUCIONES EDUCATIVAS, PROFESORADO Y ALUMNADO, SISTEMA EDUCATIVO, APRENDIZAJE Y ENSEÑANZA, EXÁMENES Y CALIFICACIONES, ESTUDIOS Y TITULACIONES, Lenguaje de aula, Material educativo y mobiliario de aula).
7. Trabajo (Profesiones y cargos, LUGARES, HERRAMIENTAS Y ROPA DE TRABAJO, ACTIVIDAD LABORAL, DESEMPLEO Y BÚSQUEDA DE TRABAJO, DERECHOS Y OBLIGACIONES LABORALES, CARACTERÍSTICAS DE UN TRABAJADOR).

8. Ocio (Tiempo libre y entretenimiento, ESPECTÁCULOS Y EXPOSICIONES, Deportes, Juegos).
9. INFORMACIÓN Y MEDIOS DE COMUNICACIÓN (INFORMACIÓN Y COMUNICACIÓN, CORRESPONDENCIA ESCRITA, TELÉFONO, PRENSA ESCRITA, TELEVISIÓN Y RADIO, INTERNET).
10. Vivienda (Acciones relacionadas con la vivienda, Características de la vivienda, Actividades domésticas, Objetos domésticos).
11. SERVICIOS (SERVICIO POSTAL, SERVICIO DE TRANSPORTE, SERVICIOS FINANCIEROS, SERVICIOS SANITARIOS, SERVICIOS EDUCATIVOS, SERVICIOS DE PROTECCIÓN Y SEGURIDAD, SERVICIOS SOCIALES, SERVICIOS DE ABASTECIMIENTO PÚBLICO).
12. COMPRAS, TIENDAS Y ESTABLECIMIENTOS (LUGARES, PERSONAS Y ACTIVIDADES, ROPA, CALZADO Y COMPLEMENTOS, ALIMENTACIÓN, PAGOS).
13. SALUD E HIGIENE (SALUD Y ENFERMEDADES, SÍNTOMAS, CENTROS DE ASISTENCIA SANITARIA, MEDICINA Y MEDICAMENTOS, HIGIENE, ESTÉTICA).
14. VIAJES, ALOJAMIENTO Y TRANSPORTE (VIAJES, ALOJAMIENTO, SISTEMA DE TRANSPORTE).
15. ECONOMÍA E INDUSTRIA (FINANZAS Y BOLSA, COMERCIO, ENTIDADES Y EMPRESA, INDUSTRIA Y ENERGÍA).
16. CIENCIA Y TECNOLOGÍA (CUESTIONES GENERALES, MATEMÁTICAS, INFORMÁTICA Y NUEVAS TECNOLOGÍAS).
17. GOBIERNO, POLÍTICA Y SOCIEDAD (SOCIEDAD, POLÍTICA Y GOBIERNO, LEY Y JUSTICIA, EJÉRCITO).
18. ACTIVIDADES ARTÍSTICAS (DISCIPLINAS Y CUALIDADES ARTÍSTICAS, MÚSICAS Y DANZA, ARQUITECTURA, ESCULTURA Y PINTURAS, LITERATURA, FOTOGRAFÍA, CINE Y TEATRO).
19. RELIGIÓN Y FILOSOFÍA (RELIGIÓN).
20. Geografía y naturaleza (UNIVERSO Y ESPACIO, GEOGRAFÍA, Espacios urbanos o rústicos, CLIMA Y TIEMPO ATMOSFÉRICO, Fauna, Flora).

Se puede observar claramente la primacía de las versalitas en esta lista, por lo que las lagunas que dejan los 16 centros de interés tradicionales de la disponibilidad léxica son notables. Con todo, sí hay bastantes centros que se podrían relacionar con las nociones específicas. No obstante, cuando lo hacemos, las comparaciones no siempre muestran resultados idénticos. En realidad, si cotejamos más profundamente las unidades léxicas que se ofrecen de ejemplo en el *PCIC* para los niveles A1 y A2 con las que obtenemos en los centros de interés, los resultados evidencian diferencias notables entre unos y otros. A veces incluso se puede deducir que las palabras más disponibles de determinados centros de interés muy poco tienen que ver con las unidades léxicas que interesan en un curso de ELE. Pensemos, por ejemplo, en las voces que ofrece el *PCIC* para *Ocio*, que se puede poner en relación con el centro de interés 15. *Juegos y diversiones*. Así pues, al fijarnos en el contenido y no solo los enunciados, vemos que las palabras de muestra que ofrece el *PCIC* y las que se aprecian entre las 50 primeras del *Léxico disponible en el español de Galicia* son las siguientes:

<i>Léxico disponible en el español de Galicia</i>	<i>Plan Curricular del Instituto Cervantes: Ocio</i>
<p>CI 15. Juegos y diversiones parchís (jugar a las) cartas fútbol ajedrez (juego de la) oca (jugar al) escondite (jugar con el) ordenador tele(visión) (jugar a las) damas (jugar a la) (video)consola baloncesto videojuego Monopoly* (jugar al) Trivial* (jugar a la) PlayStation* tute leer dominó cine escoba bailar brisca tenis (pilla) pilla juego de mesa póquer música billar (jugar a la) petanca pasear correr deporte ver la tele(visión) fútbol (juego de) tres en raya comba chinchón (navegar por) internet cuerda escuchar música (jugar a las) canicas brilé cantar mus (jugar a la) pita solitario saltar pelí(cula) volei(bol) pelota</p>	<p>Juegos juego, compañero (de juego), jugar, jugar a las cartas/al parchís/ala oca/al ajedrez</p> <hr/> <p>Deportes correr, nadar, esquiar, hacer deporte/ejercicio/gimnasia, jugar al fútbol/al baloncesto/al tenis, gimnasio, piscina, practicar un deporte, montar en bicicleta, hacer senderismo</p> <hr/> <p>Espectáculos y exposiciones película, concierto, ir al teatro/al cine/al circo, ir a un concierto, entrada, invitación, programa, exposición, obra de teatro, ópera, musical, concierto de música clásica/de rock, comprar/reservar una entrada (por teléfono, por Internet).</p> <hr/> <p>Tiempo libre y entretenimiento vacaciones, turista, oficina de información/turismo, viajar, leer el periódico/un libro/una revista, ver la televisión/una película, escuchar música/la radio/una canción, ir a un bar/a una discoteca/a un restaurante/a un museo/ ir al cine/al teatro/al parque, cantar, bailar, tiempo libre, afición, hobby, parque de atracciones/zoológico, zoo, invitar, quedar, ir de paseo/de compras/de excursión, pasear, caminar, tener una cita/planes, salir con amigos, tomar algo/un café/una caña/una copa</p>

TABLA 5. Comparación entre la noción *Ocio* (PCIC) y las primeras 50 palabras del CI 15. *Juegos y diversiones* del *Léxico disponible en el español de Galicia* (López Meirama 2008)

Como se puede notar, los términos más disponibles no consiguen cubrir la variedad de unidades léxicas que se muestran en el *PCIC* y hay bastante divergencias entre ellos. Las semejanzas se concentran en los tipos de juegos y algunas actividades. No obstante, el enunciado del centro de interés número 15 es mucho más concreto que el campo *Ocio* del *PCIC*, por lo que es de esperar que este contenga otras palabras, expresiones y frases que no se encuentran en el primero. De todas formas, “si hasta el momento no se ha avanzado demasiado en esta línea, esto no significa incapacidad o inadecuación teóricas. En todo caso, inadecuación práctica” (Bartol Hernández 2010: 92).

Se está, pues, trabajando en la línea de conseguir cubrir más nociones específicas del *PCIC* y mejorar los enunciados de los centros de interés, con el objetivo de obtener realmente aquel tipo de léxico que se debe ofrecer en un curso de ELE, y no tanto otro que no se muestra tan importante en un nivel inicial (como *comba, chinchón, brilé, mus, solitario, tute, escoba, brisca...*), por mucho que esté constituido por unidades que representan un exponente cultural de las tradiciones españolas de una época. Un gran paso adelante en esta línea lo ha supuesto la tesis presentada por Tomé Cornejo (2015), quien, desde los presupuestos de las ciencias cognitivas, hace una revisión de la metodología llevada a cabo hasta hoy en los diferentes estudios, cuestionándose en concreto por la mejor fórmula para adaptarla al ámbito de la enseñanza de ELE. Como esta investigadora explica, la disponibilidad léxica surgió con un objetivo claro, que era la selección del vocabulario francés para la enseñanza a extranjeros a mediados del siglo pasado. Desde la década de los 50 hasta ahora se han multiplicado las contribuciones de este tipo de obras, pero se ha perdido la intención de aplicarlas al ámbito de la enseñanza de lenguas, difuminándose en favor de objetivos más generales y puramente científicos. El Proyecto Panhispánico de Estudio sobre la Disponibilidad léxica es un claro ejemplo de esta pérdida de especialización para poder contribuir a enriquecer otras disciplinas como la dialectología, la sociolingüística, la etnolingüística, la psicolingüística. Las pautas metodológicas claras y definidas que este Proyecto marca contribuyen a permitir comparar los resultados de los múltiples estudios, en detrimento de una readaptación clara en función de los objetivos. Por una parte, se llevaron a cabo avances en el tratamiento de los datos y mejoras en los

análisis (como son las fórmulas para establecer el índice de disponibilidad o el de cohesión), pero en relación a los centros de interés las modificaciones fueron mínimas, y casi siempre supusieron una pequeña ampliación de los 16 clásicos.

Incluso más recientemente cuando esta línea de trabajo empezó a ser reorientada hacia el ámbito de la enseñanza de ELE, se mantuvo casi intacto su proceder metodológico. Los primeros trabajos usaron los mismos centros de interés y modificaron apenas los factores sociales a tener en cuenta. Posteriormente se empezó a debatir esta cuestión y a tener en mente las referencias de las nociones específicas del *PCIC*. No obstante, la aplicación de la disponibilidad léxica al español como lengua extranjera se llevó a cabo sin un objetivo claro. Se empezaron a hacer estudios en los que se pasaban las encuestas a estudiantes extranjeros, en contextos de inmersión o no y en grupos multilingües o monolingües. Por otra parte, se reivindicaba esta metodología como un método válido para seleccionar el vocabulario que enseñar en las aulas de español. Es en este punto en el que las aportaciones de Tomé Cornejo suponen un gran avance a la hora de distinguir procedimientos en función de los objetivos. Esta investigadora estableció dos distinciones claras: por un lado, si queremos usar la disponibilidad léxica para seleccionar el vocabulario que trasladar a las aulas de ELE, debemos tomar como referencia los estudios hechos a hablantes nativos; por otra, si lo que pretendemos es evaluar el conocimiento léxico de nuestros estudiantes, entonces deben ser estos los que sean objeto de análisis. Con todo, en ambos casos se deben adaptar las pautas a seguir, entendida esta adaptación como una necesaria especialización.

Para nuestro caso, de la propuesta de Tomé Cornejo nos interesa sobre todo la transformación profunda de los centros de interés para seleccionar el léxico disponible y aplicarlo a la docencia del español. Se parte de los temas propuestos en el inventario de nociones específicas del *PCIC* para reelaborar la nómina de centros de interés. Por nuestra parte concordamos plenamente con esta medida, teniendo en cuenta que estos son los ámbitos temáticos que se deben cubrir en un curso de ELE. Lo que esta investigación procura es aumentar el número de centros, dotarlos de mayor amplitud semántica y buscar otros de naturaleza heterogénea, para promover la evocación de diferentes clases de palabras y de diversas agrupaciones sobre las que llevar a cabo la selección

posterior. Su nómina final de centros de interés es de 47. Para pobrar sus hipótesis realiza encuestas a 180 alumnos de primer curso de Filología de la Universidad de Salamanca, con lo que la edad de los encuestados ya varía levemente, pues se pasa del último curso del Bachillerato al primero de universidad. Analiza luego los resultados cuantitativamente y se hace una propuesta de selección léxica, que posteriormente coteja con los ejemplos que en el *PCIC* se ofrecen en el nivel B1 para cada uno de estos temas. La viabilidad de esta prueba se avala con un estudio piloto, cuyos resultados compara con las nociones específicas del *PCIC* y los ejemplos que se ofrecen en cada una de ellas en el nivel B1. La comparación muestra porcentajes de similitud considerables.

Resumimos a continuación las modificaciones que se llevaron a cabo en las pruebas para intentar corregir las limitaciones observadas en los estudios anteriores. A la ampliación y reformulación de los centros de interés se añade un cambio en el tiempo que se toma para aplicar las pruebas. Así, si en la mayor parte de los estudios que se unen al Proyecto Panhispánico se usaban pruebas escritas con dos minutos para cada centro de interés, mientras que Tomé Cornejo propone 3 minutos. Por otra parte, el aumento más que considerable en la nómina de centros de interés (de 16 a 47) supone trabajar con un mayor número de informantes, para evitar el cansancio que supondría una prueba de estas dimensiones. El resultado es preguntar a más informantes, elaborando así 4 cuestionarios diferentes con dos modelos cada uno. En cada uno de estos cuestionarios se reparten 11 o 12 centros de interés distribuidos equitativamente, teniendo en cuenta las diferentes clases que se incluyen. Los dos modelos, A y B, en los que se invierte el orden de los centros, sirven para evitar las cuestiones que tienen que ver con el efecto *priming*, que consiste en que un estímulo anterior influya sobre los siguientes. Los diferentes patrones de centros de interés elaborados parten de que hay múltiples realidades y diferentes procesos de categorización. Así, por ejemplo, los que se corresponden con una escena se enuncian añadiendo la preposición “En”, para promover variadas asociaciones.

Los resultados de estas pruebas piloto muestran que, a pesar de que hay centros de interés que se deben trabajar todavía más, la mayor parte de ellos consiguen un promedio de respuestas mayor (si bien tenemos que tener en

cuenta que se daban 3 minutos y no 2) y que el grado de cohesión es muy similar a las cifras que se habían obtenido anteriormente. Algunas ventajas cualitativas que muestra el uso de estos 47 centros de interés propuestos son que se obtienen una mayor variedad de clases de palabras, no tanto sustantivos, y que se consiguen hacer aflorar más unidades pluriverbales que con los anteriores, con lo que se da una mejor respuesta a los objetivos perseguidos.

Copiamos a continuación la tabla que recoge la propuesta de esta investigadora con los 47 centros de interés en la columna de la izquierda y la correspondencia que se prevé con las nociones específicas del *PCIC*. La cursiva en algunos centros de interés indica que se tuvieron que realizar adaptaciones con respecto al plan inicial (Tomé Cornejo 2015: 441-4):

Centros de interés	Temas del <i>Plan curricular del Instituto Cervantes</i>
1. El cuerpo humano: partes, acciones y posturas.	11. Partes del cuerpo
	1.3. Acciones y posiciones que se realizan con el cuerpo.
2. El aspecto físico y la personalidad.	1.2. Características físicas.
	2.1. Carácter y personalidad.
	2.5. Modales y comportamiento.
	2.6. Valores personales.
	4.4. Actitudes y formas de comportarse
3. El ciclo de la vida: nacimiento, reproducción y muerte.	1.4. Ciclo de la vida y reproducción
	3.1.6. Edad
4. Sentimientos y sensaciones.	3.1.7. Sexo
	3.1.8. Estado civil
	2.2. Sentimientos y estados de ánimo
	2.3. Sensaciones y percepciones físicas
5. Para identificarse (documentación, datos y objetos personales).	3.1. Datos personales
	3.2. Documentación
	3.3. Objetos personales

6. Las relaciones personales (familiares y no familiares)	4.1. Relaciones familiares
	4.2. Relaciones sociales
	4.3. Celebraciones y actos familiares, sociales y religiosos.
7. Actos sociales y celebraciones.	
8. Alimentación y nutrición	5.1. Dieta y nutrición
	5.2. Bebida
	5.3. Alimentos
9. Preparar una comida	5.4. Recetas
10. En el restaurante	5.5. Platos
11. En el bar	5.6. Utensilios de cocina y mesa
	5.7. Restaurante
	5.2. Bebida
12. El sistema educativo: etapas, personal y centros de enseñanza	6.1. Centros e instituciones educativas
	6.3. Sistema educativo
	6.6. Estudios y titulaciones
	11.5. Servicios educativos
13. En clase	6.2. Profesorado y alumnado
	6.4. Aprendizaje y enseñanza
	6.5. Exámenes y clasificaciones
	6.7. Lenguaje de aula
	6.8. Material educativo y mobiliario de aula
14. El mundo laboral	7.1. Profesiones y cargos
	7.2. Lugares, herramientas y ropa de trabajo.
	7.3. Actividad laboral
	7.4. Desempleo y búsqueda de trabajo
	7.5. Derechos y obligaciones laborales
	7.6. Características de un trabajador
	15.4. Entidades y empresas

Centros de interés	Temas del <i>Plan curricular del Instituto Cervantes</i>
15. Actividades de ocio y tiempo libre	8.1. Tiempo libre y entretenimiento
	8.2. Espectáculos y exposiciones
	8.3. Deportes
	8.4. Juegos
16. Actividades artísticas	18.1. Disciplinas y cualidades artísticas
	18.2. Música y danza
	18.3. Arquitectura, escultura y pintura
	18.4. Literatura
	18.5. Fotografía
	18.6. Cine y teatro
17. La información y la publicidad	9.1. Información y comunicación
	9.4. Prensa escrita
	9.5. Televisión y radio
	15.3.3. Publicidad, mercadotecnia
18. Comunicarse por teléfono o carta	9.2. Correspondencia escrita
	9.3. Teléfono
	11.1. Servicio postal
19. Informática y nuevas tecnologías	9.6. Internet
	16.4. Informática y nuevas tecnologías
20. Buscar una vivienda	10.1. Acciones relacionadas con la vivienda
	10.2. Características de la vivienda
	10.3.2. Decoración de la casa
21. La casa y las tareas domésticas	10.2. Características de la vivienda
	10.3.1. Limpieza de la casa
	10.4. Objetos domésticos
	11.2. Servicios de transporte
22. Ir de viaje	14.1. Viajes
	14.2. Alojamiento
	14.3. Sistema de transporte
23. En el banco	11.3. Servicios financieros
	15.1. Finanzas y bolsa
24. El sistema sanitario: personal y centros	11.4. Servicios sanitarios
25. En comisaría	11.6. Servicios de protección y seguridad
26. Servicios sociales y solidaridad	11.7. Servicios sociales
27. La ciudad y el campo	11.8. Servicios de abastecimiento público
	20.3. Espacios urbanos o rústicos
	12.1. Lugares

28. En la tienda de ropa	12.2. Ropa, calzado y complementos
<i>29. Hacer la compra</i>	12.3. Alimentación 12.4. Pagos
<i>30. En el médico</i>	13.1. Salud y enfermedades 13.2. Heridas y traumatismos 13.3. Síntomas 13.4. Centros de asistencia sanitaria 13.5. Medicina y medicamentos
<i>31. En el baño o aseo</i>	13.6. Higiene 13.7. Estética
<i>32. Conducir un coche</i>	14.3.3. La conducción
33. Economía y hacienda	15.1. Finanzas y bolsa 15.2. Renta 15.3.2. Comercio exterior
34. Sectores de producción (primario y secundario)	15.3.2. Comercio exterior 15.5. Industria y energía
<i>35. La ciencia</i>	16.1. Cuestiones generales 16.2. Biología 16.3. Matemáticas 16.5. Física y química
<i>36. Organización política y social</i>	17.1. Sociedad 17.2. Política y gobierno
<i>37. En un juicio</i>	17.3. Ley y justicia
<i>38. El ejército</i>	17.4. Ejército
39. En el cine o en el teatro	8.2. Espectáculos y exposiciones
<i>40. Creencias religiosas y espiritualidad</i>	19.1. Religión 19.2. Filosofía
<i>41. El universo y los viajes espaciales</i>	20.1. Universo y espacio
<i>42. La geografía y los accidentes geográficos</i>	20.2. Geografía
<i>43. El tiempo atmosférico</i>	20.4. Clima y tiempo atmosférico
<i>44. Animales y plantas</i>	20.5. Fauna 20.6. Flora
<i>45. Medioambiente y desastres naturales</i>	20.7. Problemas medioambientales y desastres naturales

TABLA 6. Los centros de interés propuestos por Tomé Cornejo (2015: 441-4)

A partir de los listados obtenidos en las pruebas, Tomé Cornejo propone una selección léxica hasta el nivel B1. En cuanto al número de palabras que se escogen considera las que el *PCIC* ofrece en los niveles A1, A2 y B1 en la categoría equivalente, es decir, si en la noción Partes del cuerpo el *PCIC* presenta 29 palabras, el número de vocablos de los listados de disponibilidad léxica que se seleccionan es 29. Este umbral de corte en el número de palabras seleccionado se justifica por ser una prueba piloto y con la mera intención de cotejar los datos, no como un criterio que se deba aplicar en investigaciones que pretendan emplear la disponibilidad léxica para obtener el número de palabras que enseñar en el nivel B1, para lo cual creemos más conveniente partir de otro tipo de criterios, como más adelante expondremos.

En líneas generales, la propuesta que resume el Capítulo 4 de la tesis de Tomé Cornejo supone un gran avance para poder usar la disponibilidad léxica como método de selección léxica, pues hace que esta sea capaz de dar respuesta de forma mucho más eficaz a este propósito. Esperamos, así mismo, que se continúe con esta línea de investigación y que se intente aplicar y perfeccionar en posteriores trabajos.

3.4.2.2. Algunas propuestas de aplicación de la disponibilidad léxica

Como hemos venido subrayando, consideramos que las virtudes de la disponibilidad léxica todavía no se han sabido explotar en su aplicación a la enseñanza y aprendizaje de ELE. No obstante, sí existen propuestas puntuales para ello, algunas de carácter más teórico y otras un poco más prácticas, aunque la mayor parte de ellas más parciales que completas, en la medida en que no ofrecen un vocabulario final para llevar al aula. Haremos un breve repaso a aquellas que consideramos más interesantes.

Destacan las recientes aportaciones de Tomé Cornejo (2015) y Ávila Muñoz (2014, 2016) por su utilidad para la selección del léxico disponible. La primera, como ya hemos comentado, ofrece una actualización y ampliación de la nómina de centros de interés, que resulta mucho más adecuada para seleccionar el vocabulario teniendo en cuenta los temas del Consejo de Europa y el *PCIC* (*vid. supra* § 3.4.2.1. y *vid. infra* § 6.3.3.1 y 6.3.3.2.). El segundo, no trata la metodología

que se aplica en las pruebas de disponibilidad léxica sino el tratamiento de sus resultados a la hora de escoger aquellas unidades de forma más objetiva y adecuada para su planificación (*vid. infra* § 6.3.3.2 y 8.3.3.2.).

Anteriormente, López Pérez (2013: 171) aportó una serie de criterios para la selección de voces en el ámbito de la enseñanza del español como L2 en contextos escolares. Distinguía los tipos de vocabularios (básico, disponible, técnico y académico y sus correspondientes repertorios). Estos criterios le posibilitan la planificación de la enseñanza del léxico en una L2 y le permiten una gradación de los contenidos léxicos en las distintas etapas del aprendizaje de la nueva lengua. Esta autora afirma que los vocabularios básicos y disponibles son adecuados para el inicio del aprendizaje de una L2, mientras que el vocabulario técnico, subtécnico y académico es útil para el período en el que el alumnado se enfrenta a los contenidos de las asignaturas, que a su vez se extiende en los distintos niveles escolares.

Existen también las aportaciones de Izquierdo Gil (2004, 2005) sobre la selección del léxico en la enseñanza de ELE. Este trabajo doctoral, a pesar de centrarse solo en un nivel elemental en estudiantes francófonos, resulta igualmente interesante por tratarse de un estudio muy completo, cuyos resultados se podrían extrapolar a otros contextos y países, sobre todo los que atañen al análisis contrastivo (diferencias, semejanzas, correspondencias, interferencias, cognados, parógrafos, homógrafos), pues pueden resultar muy prácticos también con otras lenguas romances.

Sanz Álava (2000) y Palapanidi (2012) llevan a cabo dos trabajos más parciales sobre la aplicación del léxico disponible a la selección del vocabulario. En el primer caso se pretende mostrar una breve aplicación práctica para el nivel intermedio. En el segundo, va dirigida en general a la didáctica del léxico de una lengua extranjera.

Fernández Leyva (2015) escribe recientemente un artículo sobre la cuestión puntual en el proceso de elección de las palabras disponibles. Se trata de la aplicación del índice de disponibilidad léxica a la selección del vocabulario de manuales de ELE. Para ello contempla otros trabajos anteriores en los que se ha usado este criterio, se fija en cuál era en concreto el índice de disponibilidad que se usaba como corte y compara las palabras de las listas de disponibilidad

con las que se incluyen en un manual de ELE concreto.

Dos artículos que pueden resultar interesantes a nivel teórico para un acercamiento al tema de la selección del léxico disponible para la didáctica de una L2 son los de Rodríguez Muñoz y Muñoz Hernández (2009) y Bartol Hernández (2010). En ellos hacen un repaso a las cuestiones que justifican la inclusión de este tipo de unidades en la planificación del léxico:

A) La disponibilidad léxica para seleccionar el léxico específico

Además de la lengua general, existen también algunas propuestas para el Español para Fines Específicos. Dos de ellas se desarrollaron en el ámbito de los medios de comunicación (González Sánchez 2003) y más en concreto sobre el léxico de la radio (Guerra Salas & Gómez Sánchez 2005). Más recientemente surgió una aportación al Español de los Negocios, una rama con mucho peso dentro de los Fines Específicos. Más concretamente trataba las unidades léxicas que se usan en la banca española con un trabajo en el que también se consideraban las más disponibles (Mesa Betancor 2011).

B) La disponibilidad léxica para medir el desarrollo de la competencia léxica en una L2 o en una lengua extranjera

Otra aplicación que se puede dar a los datos de las listas de disponibilidad es su uso para medir el desarrollo de la competencia léxica en el alumnado de una segunda lengua. Ha tratado estas cuestiones Frey Pereira (2007), al utilizar la disponibilidad léxica como método de comparación con los textos escritos por estudiantes de ELE, y López González (2010), partiendo de la disponibilidad léxica para evaluar el desarrollo de la competencia léxica.

3.4.3. El vocabulario fundamental

A lo largo de este capítulo nos hemos detenido en explicar la características de las palabras frecuentes y de las disponibles, así como la utilidad y los puntos débiles de estas dos líneas de investigación en relación a la selección del vocabulario. Debido a la parcialidad y los problemas que presentan las dos líneas de selección léxica anteriores (la frecuencia y la disponibilidad),

desde muy pronto se vio la necesidad de conjugar ambas para sumar sus beneficios, ya que con una se obtiene un tipo de unidades y con otra justo el opuesto.

Por este motivo, a la hora de escoger el vocabulario no sólo se deben tener en cuenta las unidades léxicas más frecuentes, sino también una serie de palabras temáticas, pues cuando hablamos necesitamos expresar conceptos vinculados a temas y situaciones concretas. En este sentido, tanto Michéa como los investigadores actuales en disponibilidad léxica han conseguido confirmar, pues, que el léxico básico y el léxico disponible constituyen dos parcelas altamente complementarias y que ambas forman el llamado *vocabulario fundamental* de una lengua. El hecho de que ambos léxicos sean diferentes pero compatibles, revierte en que al fusionarlos estamos sumando sus virtudes y minimizando sus limitaciones. Son muchas las voces que se muestran de acuerdo con las ventajas de su unión: “Su identificación es un ejercicio imprescindible para emprender cualquier tipo de planificación relacionada con el léxico”, indica, por ejemplo, López Morales (1999: 2).

Se parte de la hipótesis de que el vocabulario fundamental refleja de forma más completa el lexicón mental y ayuda a desarrollar mejor la competencia léxica en una lengua, de ahí que pueda resultar altamente útil como propuesta de selección léxica para la docencia de un idioma.

En la Tabla 7 se presenta de forma resumida y esquemática cuáles son las ventajas y los inconvenientes asociados al empleo de la frecuencia, por una parte, y de la disponibilidad léxica, por otra, como criterios para la selección del vocabulario. Recurrir a la fusión de ambos permitirá sumar los puntos fuertes.

Léxico frecuente		Léxico disponible	
Puntos fuertes	Puntos débiles	Puntos fuertes	Puntos débiles
<ul style="list-style-type: none"> •Reducir arbitrariedad: lexicometría. •Establecer cortes: índices frecuencia y uso. •Obtener las palabras más usuales (índice de uso). •Permitir gradación del vocabulario por niveles 	<ul style="list-style-type: none"> •Tipo de unidades léxicas: llegamos a palabras vacías, gramaticales, nombres y adjetivos de significado muy general (<i>cosa, casa, tiempo, lugar, persona, sitio</i>) y con débil valor informativo. •Frecuencia no siempre es sinónimo de utilidad didáctica: Existen palabras infrecuentes pero útiles. •Palabra gráfica como unidad de recuento: no reconocen palabras pluriverbales. •Orden didáctico inadecuado: no establecen una secuenciación precisa del vocabulario para la docencia. •Falta de correspondencia entre el lema y la unidad léxica de aprendizaje (se ofrece la frecuencia de <i>ir</i>, pero no de su paradigma). •Divergencias entre los listados de frecuencia •Falta establecer el número de unidades para cada nivel y establecer parámetros de selección. 	<ul style="list-style-type: none"> •Reducir arbitrariedad: lexicometría. •Palabras temáticas de alto valor significativo: sustantivos, adjetivos, verbos. •Establecer cortes: índice disponibilidad. •Los centros de interés son muy semejantes a las nociones o campos léxicos (<i>Marco y PCIC</i>). •Seleccionar léxico concreto (Fines Específicos). •Conocer la organización del lexicón mental. •Integrar palabras reales y actuales (variación léxica). •Conocer la tipicidad de las palabras (prototipos). 	<ul style="list-style-type: none"> •Tipo de unidades léxicas: llegamos sobre todo a sustantivos, pero muy pocos adjetivos y verbos. •No obtenemos casi unidades pluriverbales. •Todavía es necesario mejorar los CI •Debemos acercarnos más a las áreas temáticas del <i>PCIC</i>. •Debemos establecer criterios de selección.

TABLA 7. Resumen de los puntos fuertes y débiles de la frecuencia y la disponibilidad léxicas para la selección del vocabulario

No nos detendremos aquí en comentar las diferentes propuestas de vocabularios fundamentales que ya ha habido, ni tampoco profundizaremos en su definición, metodología y organización final, pues es algo que llevamos a cabo de forma más extensa en los Capítulos 7 y 8. No obstante, ya podemos definir el

concepto de vocabulario fundamental en pocas líneas como “aquel vocabulario necesario para que un hablante pueda desenvolverse con eficacia en las situaciones comunicativas cotidianas tratando temas habituales”. En el Capítulo 8 desarrollamos también la metodología para llegar al vocabulario fundamental de una lengua, en este caso la española, con el fin de aplicarlo a la didáctica de este idioma. Su organización final tiene en cuenta las nociones específicas del *PCIC*, con el objetivo de que resulte práctico para el profesorado y estudiantado de ELE.

3.4.4. El vocabulario instrumental

Hasta aquí hemos presentado algunos métodos de selección de diferentes tipos de vocabulario existentes en la lengua. Aun así, si nos situamos en un contexto de enseñanza-aprendizaje, podemos adoptar otros criterios para decidir qué unidades enseñamos en el aula, pues debido a otros motivos, como el entorno o el tipo de alumnado, nos pueden ser igualmente ventajosas.

Si tenemos en cuenta el entorno educativo, nos parece útil presentar en niveles iniciales una serie de unidades léxicas mediante las cuales el alumno pueda llevar a cabo ciertas estrategias de comunicación o compensación con el fin de resolver sus problemas léxicos como el olvido o desconocimiento de una palabra. Este tipo de vocabulario recibe el nombre de “instrumental” y está compuesto fundamentalmente por hiperónimos, verbos muy generales, términos vagos o aproximativos, adjetivos referidos al color, la forma y el tamaño, etc. Mediante el vocabulario instrumental los alumnos podría llegar a comprender definiciones ofrecidas por el docente para explicar el significado de unidades específicas, por lo que este léxico resultaría rentable para la comunicación, tanto a nivel receptivo como expresivo, para compensar los vacíos léxicos. Así, el control de un vocabulario instrumental mínimo es bueno para el desarrollo de la competencia estratégica (Izquierdo Gil 2004: 275-294).

Desde una perspectiva léxica, dicha competencia implica la habilidad para compensar o reparar el olvido momentáneo o el desconocimiento de una unidad léxica necesaria en un momento determinado por medio de diferentes recursos comunicativos, entre los cuales cabe destacar el uso de formulaciones

aproximativas, circunlocuciones o definiciones. Gracias a estas tácticas o estrategias, cualquier hablante puede evitar una ruptura comunicativa y llevar a cabo una comunicación efectiva (Berry-Bravo 1993: 371). Diversas investigaciones han relacionado esta rentabilidad comunicativa de las unidades léxicas instrumentales con su gran cobertura semántica, lo cual justificaría su consideración en un programa de español LE en el nivel elemental (Nation & Waring 1997: 19; Izquierdo Gil 2004: 275).

Las dos características más notables de las unidades léxicas instrumentales son: su alta capacidad señalizadora o riqueza referencial (*high indexical potential*) y su escaso contenido léxico o escasa especificidad semántica (*high indexicality and low lexicality*). La relación entre ellas es que la primera es inversamente proporcional a la segunda (Robinson 1989: 543; Widdowson 1984).

3.4.5. La lingüística contrastiva

Cuando nos situamos en un entorno educativo de enseñanzas de segundas lenguas, otra de las estrategias a la que podemos recurrir para escoger aquellos contenidos léxicos que llevar al aula es el análisis contrastivo. Si bien los errores pueden tener lugar tanto en áreas de gran semejanza o convergencia, como de gran divergencia, es entre lenguas próximas cuando las interferencias lingüísticas adquieren un mayor peso. En concreto, esta táctica comparativa puede resultar altamente práctica entre lenguas románicas.

La lingüística contrastiva se puede entender como la comparación lingüística sincrónica para describir las semejanzas y diferencias parciales o totales entre dos o más lenguas. Aunque la lingüística contrastiva surgió como gramática contrastiva, este cotejo puede aplicarse a todos los ámbitos de la lengua, ya que se pueden establecer diferencias y semejanzas tanto gramaticales, como fónicas, léxicas, textuales o pragmáticas. En consecuencia, tanto las transferencias positivas como las interferencias pueden desarrollarse en cualquier nivel de la lengua: fónico, gramatical, léxico, semántico, sociolingüístico, sociocultural, etc. En la década de los 70 se demostró que la interferencia se produce con más frecuencia entre aquellas lenguas que el

discente siente próximas y que las grandes semejanzas pueden llegar a convertirse en fuente de confusión. El análisis contrastivo fue fuertemente criticado, sobre todo desde los partidarios del generativismo. Recientemente diversos autores defienden los efectos positivos de llevar al aula comparaciones interlingüísticas, basadas en las analogías y las divergencias.

Las comparaciones hechas en el aula gracias a un análisis contrastivo de la lengua materna y la lengua meta potenciará el desarrollo de la conciencia metalingüística del grupo. Estas actividades de comparación y reflexión sobre analogías y diferencias entre los idiomas ayudarán además a remediar la incapacidad del alumnado para establecer vínculos entre los conocimientos que posee de su lengua materna y los que va adquiriendo de la extranjera.

Seguimos el capítulo de Izquierdo Gil dedicado a la enseñanza-aprendizaje del léxico en dos lenguas románicas, en su caso el español en un contexto francófono, para resumir algunas de las explotaciones didácticas que se pueden llevar a cabo de las semejanzas interlingüísticas (2004:189-269). Nos referimos por ejemplo al aprovechamiento de conocimientos léxicos previos, gracias a la existencia de cognados entre ambas lenguas. Los cognados son palabras que comparten significado, ortografía y pronunciación similares en dos idiomas, cuyo porcentaje aumenta entre lenguas próximas. Además de los cognados verdaderos, también existen los parógrafos cognados, que son palabras que en dos lenguas comparten el mismo significado pero entre las que existe algún tipo de variación gráfica. Muchas veces pueden ser generalizables en reglas de correspondencias. Por ejemplo, observando los sufijos entre el español y portugués se comprueba que hay claras identificaciones entre algunas parejas, como *-ción* y *-ção*, *-aje* y *-agem*, *-ble* y *-vel*, *-dad/tad* y *-dade*, etc. Aunque también existen cognados con variaciones gráficas no sistematizables.

También se puede realizar una selección de palabras en función de sus correspondencias interlingüísticas. Por ejemplo, volviendo al caso del español y el portugués, entre las unidades léxicas con diptongación que presenta el español y la ausencia de ella en portugués, donde muchas veces se observa una abertura de las vocales *o* y *e*. El portugués tiene un sistema fonético de 7 vocales, donde estas son de grado medio (abierto o cerrado). Los ejemplos de este tipo son muchos, y si somos conscientes de esta equivalencia podemos extrapolarla a

otros casos, deduciendo su significado sin antes haber estudiado la palabra concreta. Así por ejemplo se podría trabajar con pares como estos, en los que se presenta claramente esta correspondencia entre las vocales de grado medio portuguesas y los diptongos españoles:

O --> UE

E --> IE

morte/muerte

ferro/hierro

porto/puerto

serpente/serpiente

porta/puerta

vento/viento

ponte/puente

ventre/vientre

La enseñanza explícita de las correspondencias más productivas mediante actividades para ver analogías pueden resultar en estos casos muy productivas para hacer visibles las similitudes y las divergencias. Por ejemplo la correspondencia de sufijos *-aje* y *-agem* se podría aprovechar también para mostrar las diferencias de género, ya que en portugués las palabras que acaban así son femeninas, al contrario que en español en su gran mayoría.

Las divergencias léxicas formales se pueden clasificar en interferencias ortográficas, divergencias fonético-fonológicas, divergencias de género y número y, por último, las divergencias léxico-semánticas. Entre estas últimas están los conocidos como falsos amigos (totales o parciales), que son abundantes entre algunas lenguas. Volviendo al ejemplo del portugués con respecto al español, este tipo de casos son realmente numerosos, por lo que es conveniente tratarlos debidamente por ámbitos temáticos, según se vaya estudiando cada campo semántico. En ocasiones, incluso se observa lo que se puede considerar una cadena de falsos amigos. Nos referimos a ejemplos como *taller-oficina*, *oficina-escritório*, *cubiertos-talheres*, donde el primera palabra es la española y la segunda la portuguesa.

Capítulo 4. El número de unidades léxicas: ¿Cuánto vocabulario seleccionar para cada nivel?

4.1. ¿Cuánto vocabulario seleccionar para cada nivel?

Hemos debatido con anterioridad cuáles son los criterios y métodos de selección del vocabulario para aplicarlos a la docencia de ELE y, solo de forma sucinta, hemos considerado la cuestión del número de unidades léxicas que enseñar en cada nivel. Nos centraremos ahora en ello, intentando vislumbrar cuál es el número de palabras necesarias para cada uno, sin entrar por el momento en cómo llevar a cabo el corte o gradación en los procesos de selección, una cuestión sobre la que volvemos más adelante.

A la pregunta de cuánto vocabulario debemos enseñar se le podría dar la vuelta, planteando cuánto vocabulario necesitaría en realidad nuestro alumnado para desarrollar aquellas funciones que se definen en el *MCERL* en cada nivel. Para ofrecer una respuesta satisfactoria a esta tarea, podemos pensar en diferentes planos. Partimos, por supuesto, de que debemos analizar nuestro grupo de estudiantes concreto, para estudiar las necesidades léxicas e intereses. Así, en un primer plano, podemos ir viendo las funciones, nociones generales y específicas que se especifican en el *PCIC* para los diferentes niveles y podemos intentar seleccionar, con los métodos que hemos comentado que tenemos a nuestro alcance, las unidades léxicas que necesitamos para cada una de las nociones. En otro plano, es posible recurrir a las cuatro destrezas que el alumno debe trabajar para desarrollar su competencia comunicativa, es decir, sus

capacidades productivas y receptoras. Para las destrezas de producción es el hablante quien escoge las palabras que usa, adecuándose, lógicamente, al contexto. Sin embargo, para comprender lo que alguien dice o escribe, el aprendiz depende de las palabras que los demás escogen, por lo que en la selección del vocabulario para la competencia receptiva, el factor frecuencia resulta importante, de cara a elegir las palabras que tienen mayor probabilidad de aparecer en textos orales y escritos. Y a esto sumamos que tendemos a utilizar un número reducido de voces (*actuación*) aunque conozcamos en realidad un número elevado (*competencia*). Recordemos que según los conceptos de vocabulario receptivo y productivo (Bogaards 1994: 115), el léxico pasivo que solemos dominar es mayor que el activo. Parece razonable, pues, priorizar el aprendizaje de palabras que tienen más probabilidades de ser oídas o usadas a menudo. Es este sentido en el que se defiende la rentabilidad comunicativa de las palabras con alta frecuencia. Por ejemplo, Almela *et al.* (2005) señalan que es mucho más rentable aprender una palabra muy usada y altamente polisémica que otra poco usada. Bogaards (2008: 1234), entre otros, explica que existe una fuerte correlación entre frecuencia y polisemia: cuantos más sentidos tenga una palabra, más ocurrencias tiene.

Pues bien, dicho esto, debemos pensar cómo determinar el tipo y el número de palabras que resultan necesarias para que nuestra capacidad receptiva pueda llevarse a cabo con éxito en los diferentes intercambios comunicativos orales y escritos. Para este fin, se han abordado diversos estudios que toman los conceptos de *cobertura léxica* y de *nivel umbral de comprensión óptima y aceptable* como uno de los indicadores clave para saber cuántas unidades léxicas son necesarias para entender de modo efectivo un texto o una conversación. La cobertura textual debemos entenderla como el espacio semántico y situacional que es capaz de cubrir una palabra o el porcentaje de unidades que componen un texto (oral o escrito). Así pues, la cobertura de un corpus de frecuencias léxicas hace referencia al número de palabras que es necesario utilizar para dar cuenta de la mayoría del léxico incluido en él, tal y como expone Terrádez Gurrea (2001: 66).

Como ya hemos explicado, las voces más frecuentes o básicas son útiles en cuanto que son rentables a nivel comunicativo, ya que la cantidad de palabras

que conoce el alumnado tiene implicaciones en su competencia comunicativa y en su capacidad de comprensión de los textos. Si nos centramos ahora en la comprensión, debemos saber entonces qué cantidad o porcentaje de vocabulario debemos dominar para comprender los textos y situaciones comunicativas que se describen para cada nivel del *MCERL*. Las investigaciones hechas en este sentido no han sido tan concretas, pero sí disponemos de bibliografía científica suficiente que ha abordado la cuestión en el ámbito de alumnos de inglés como lengua extranjera. Lo ideal, para tener datos más fiables y concretos, sería aplicar estas investigaciones no solo al español, sino a un corpus (suficientemente variado y representativo) que incluyese lengua oral y escrita en español y se centrase en analizar aquellos textos y situaciones comunicativas que describe el *MCERL* para los niveles de referencia. Así, podríamos realmente definir mucho mejor los porcentajes y aplicarlos a cada nivel por separado. De todas formas, según Izquierdo (2004: 337), los resultados sobre cobertura textual parecen coincidir independientemente de la lengua y del corpus, por lo que los estudios hechos hasta ahora para el inglés nos ofrecen datos muy útiles y extrapolables igualmente al español y a los diferentes niveles de lengua, aunque los cálculos tengan que ser más aproximativos.

López Morales (1984), al explicar la aparición y desarrollo de este tipo de estudios, señala que contamos con cuatro procedimientos para la medición estadística del léxico existente en una comunidad lingüística (y de habla): diccionarios normativos, frecuencia léxica, léxico básico y disponibilidad léxica. Por nuestra parte, descartamos el diccionario normativo, pues constituye un depósito de todo, o casi todo, el caudal léxico de una lengua, y añadimos el concepto reformulado de léxico fundamental.

Por otra parte, el denominado *umbral léxico* nos ofrecen un criterio muy útil para medir las necesidades receptoras de una persona que estudia español como lengua extranjera en cuanto a su vocabulario en la comprensión de textos. Este umbral sería el límite a partir del cual somos capaces de comprender un texto oral o escrito, esto es, el número o porcentaje de palabras que debemos conocer para llegar a un nivel de comprensión aceptable u óptimo. Aquí la respuesta suele establecer el 95% para un nivel aceptable y un 98% para un nivel óptimo (Cartaya Febres 2011). En relación a los porcentajes de

comprensión que alcanzamos con diferentes cantidades de palabras, sí se han apuntado datos sobre el español, ya que muchos de los vocabularios de frecuencias o básicos dedican algunas líneas a comentar los porcentajes de cobertura textual que ofrecen las palabras más frecuentes con respecto al corpus de textos del que se han extraído. Haremos a continuación un breve resumen de los dos tipos de estudios.

4.1.1. Cobertura léxica textual

En relación con la cobertura textual, hay diferentes datos sobre el número de palabras y los porcentajes de cobertura textual que cada uno de ellos proporciona (*vid. supra* § 2.2. y 3.4.1; *vid. infra* § 4.3.) A continuación retomamos algunos de ellos.

Juilland y Chang Rodríguez (1964) indicaban que en su diccionario de frecuencias sólo las 10 primeras formas suman ya el 36.55% de la frecuencia total, las 100 primeras el 66.20%, y las 1.000 primeras el 85.50% del total. Estas apreciaciones parecen corroborarse por estudios estadísticos más recientes.

En las listas de frecuencia elaboradas por Ávila Muñoz (1999) las primeras 5.000 palabras del corpus cubren el 99,7% de los textos de su corpus. Como ya vimos, al tratarse de un corpus de español hablado, el vocabulario es más reducido y simple. Estos datos están aplicados al léxico de frecuencia del español hablado en la ciudad de Málaga. Los datos corroboran que 5.000 palabras son suficientes para las necesidades básicas de un hablante, al que pueden ayudar a entender cualquier conversación.

En el *Léxico básico de la modalidad andaluza del español* (García Surrallés 2012: 10), dedica unas líneas a ofrecernos los cálculos esgrimidos por Closset (1958) y Esquer Torres (1968: 83). Ambos autores explican que el total del léxico de la lengua inglesa está constituido por unas 50.000 palabras, pero que 1.000 palabras bastan para abarcar el 85% de una conversación y para el uso corriente del idioma. Con otras 2.000 palabras más, solo subiríamos este porcentaje en un 10%, y las restantes 47.000 solo se usan en un 5% de las producciones, por lo que su frecuencia es muy baja. Es decir, que con 3.000 palabras se cubre un 95% del uso corriente de la lengua inglesa, por lo que

Esquer argumentaba que “salta a la vista la conveniencia de comenzar el aprendizaje léxico por aquellas mil o tres mil palabras que acaparan tan abrumadoramente el uso idiomático”. No obstante, estas cifras no coinciden con las que nos ofrecen otros autores. Por ejemplo, Patterson y Urrutibéheity (1975) valoran que nuestras 100 palabras más frecuentes constituyen más del 30% del material léxico de cualquier texto, con las 1.000 se alcanza el 50% de todo texto, y con las 5.000 más abundantes se sobrepasa el 90%. Quizá la diferencia resida en el tipo de textos y conversaciones que se usan para establecer estos porcentajes, en el sentido de que aquellos buscan los datos en “una conversación y para el uso corriente del idioma”. Los resultados en la lengua oral y el lenguaje coloquial, siempre van a ser menores, pues, como ya hemos apuntado, está demostrado que la cantidad de palabras que se usan normalmente en este tipo de interacciones es más reducida y de menor complejidad.

En el diccionario de frecuencias de Davies (2006) los autores de su prefacio resumen estas cuestiones cuando explican que esta obra intenta suplir las carencias de los manuales de español, aportando léxico real y frecuente a los aprendices. Una idea que no es nueva en otras lenguas como el inglés, pero que en el ámbito del español parece ser menos difundida y usada, según ellos:

While English has long been well provided with such data, there has been a relative paucity of such material for other languages. This series aims to provide such information so that the benefits of the use of frequency information in syllabus design can be explored for languages other than English. [Davies 2006: 7]

Se intenta, pues, extrapolar los datos que se habían ofrecido ya para el inglés y se exponen los que Nation (1990) ofrecía sobre la cobertura textual de las palabras más frecuentes —recordemos que Davies ofrece las 5.000 más frecuentes, según su frecuencia, uso y dispersión—, que llega hasta el 85% con las 1.000 primeras y alcanza un 95% de un texto escrito con las 4.000 o 5.000 palabras más frecuentes.

Por su parte, Izquierdo Gil (2004: 337) recoge algunos datos en referencia a la cobertura léxica o textual y muestra los resultados de varios estudios de léxico-estadística. Apunta también que cualquiera que sea la lengua y corpus que consultemos, los resultados sobre la cobertura textual suelen aproximarse, por

mucho que se puedan alterar estas proporciones en función del tamaño de la muestra, de su carácter oral o escrito, o bien del procesamiento de los datos.

En el ámbito del inglés, por ejemplo, Nation y Waring (1997) aportan cifras tomadas del *Brown Corpus of Standard American English*, en el que se especifica que con 1.000 lemas cubrimos el 72% de las palabras del corpus, con 2.000 el 79%, con los 3.000 el 84%, con 4.000 el 86,8% y con 6.000 el 89,9%. Para cubrir el 97,8% del corpus hacen falta 15.851 lemas. Por su parte, el *Cambridge International Corpus* dice que con 10 palabras (*the, of, to, and, a, in, that, is, for, it*) se cubre el 22% de este corpus, con las 50 primeras el 37%, con las 100 primeras el 44% y con las 2.000 palabras más frecuentes el 80%.

En la introducción del *Collins COBUILD English Language Dictionary* (1987), que fue creado a partir de un corpus de cerca de 20 millones de palabras extraídas de libros, revistas, periódicos, panfletos, folletos, conversaciones, programas de radio y televisión, el editor, John Sinclair, llama ya la atención sobre el hecho de que un número relativamente pequeño de palabras sea suficiente para cubrir un porcentaje de la lengua escrita y hablada diariamente. El COBUILD, que reúne cerca de 40.000 palabras, centró la atención en las palabras más frecuentes, que calcularon que son las primeras 2.000 o 3.000. Las entradas para esas palabras se acompañan de información sobre su uso. El editor advierte también que no se encontrarán palabras muy técnicas, ni términos obsoletos o dialectales, pues el objetivo del diccionario es el lenguaje del día a día.

Por su parte, el corpus que dio lugar al *Cambridge International Dictionary of English* (1995) recogía más de 100 millones de palabras, incluyendo variedades habladas y escritas de varios tipos y fuentes. Un aspecto interesante de la obra es la inclusión entre las fuentes de textos orales y escritos producidos por aprendices de inglés. Según los autores, esos textos fueron usados para atender a las necesidades de los estudiantes de la lengua inglesa.

Ambas obras retomaron el interés por la frecuencia de las palabras, un campo de estudio cuyo origen se remonta al final del siglo XIX, cuando Sweet postulaba que el vocabulario debería ser firmemente controlado por un método racional. Según explica Howatt, Sweet afirmaba que 3.000 palabras serían suficientes para la enseñanza de la lengua inglesa, exceptuando el vocabulario

especializado (Howatt 1984: 187). En la década de los años 20 del siglo pasado, se iniciaron los estudios lexicográficos que resultaron en la creación de listas de las palabras más usadas. Como ya hemos mencionado en los capítulos anteriores, en 1930 Ogden publicó un libro didáctico llamado *Basic English*, que utilizaba solo 850 palabras: 150 adjetivos, 600 sustantivos, 100 verbos. En 1937, Palmer presentó junto a Hornby una selección de 1.000 palabras en su volumen *Thousand-word English*. En 1943, Eckersley publicó la serie didáctica *Essential English* que contenía 500 palabras nuevas en cada una de sus cuatro prácticas. En 1953, West publicó su lista de cerca de 2.000 palabras (excluyendo las derivadas), denominada *General Service List*. Según Morris (1964: 31), West se basó en los siguientes criterios para seleccionar las palabras:

- Word frequency
- Structural value
- Universality in respect of geographic area
- Subject range
- Value of purposes of definition
- Values for word-building
- Stylistic functions of a word

En la década del los 70 las listas empiezan a perder su prestigio. Tenemos un testimonio de esta pérdida en las afirmaciones de Wilkins (1972: 117-8):

The original frequency data seem in danger of getting lost among these other considerations. No doubt those items that would appear on a list of the 2.000 most common words in a language are indeed useful words for a foreigner to acquire. But, as we have seen, learning could not be confined to these items. The teacher himself does not normally choose what vocabulary items his pupils will learn. (...) The more language teaching is oriented towards meeting the needs of the learners the more likely it is that the situations used for teaching will produce 'useful' language without it having been necessary to draw up an inventory of lexical items beforehand. A reasonable behavioural prediction may do the work that a vocabulary frequency counts sets out to do.

Sin embargo, las previsiones de Wilkins no se hicieron realidad y con el uso de corpus informatizados el interés por la frecuencia de las palabras volvió a manifestarse. Por ejemplo, Jane y Dave Willis produjeron una serie didáctica, *Collins COBUILD English Course*, basándose en el corpus del proyecto COBUILD. La serie, reunida en tres volúmenes, incluía textos reales alcanzando un total de 2.500 palabras. Al final de cada unidad, las voces nuevas eran listadas y se

añadían a las anteriores. Ya en este siglo, McEnery y Wilson (2001: 121) también se posicionan en relación al uso de corpus:

Corpus examples are important in language learning as they expose students at an early stage in the learning process to the kinds of sentences and vocabulary which they will encounter in reading genuine texts in the language or in using the language in real communicative situations.

Sobre el número de términos enseñados en un curso, Gairns y Redman (1986: 66) sugieren que en una clase de 60 minutos, el *input* ideal es de entre 8 y 12 palabras productivas.

Recopilamos en la siguiente tabla, de elaboración propia, algunos datos sobre cobertura léxica, en inglés y en español, para poder mostrar y cotejar los porcentajes a los que hemos podido acceder³².

³²Las obras consideradas son las siguientes:

- Juilland, A. & Chang-Rodríguez, E. (1964). *Frequency Dictionary of Spanish Words*. La Haya: Mouton.
- Carroll, J. B.; Davies, P. & Richman, B. (1971). *The American Heritage Word Frequency Book*. Boston: Houghton Mifflin Company.
- Patterson, W. & Urrutibéheity, H. (1975). *The lexical structure of Spanish*. The Hague: Mouton.
- Francis, W. N. & Kučera, H. (1964, 1971, 1979). *The Brown University Standard Corpus of Present-Day American English*. Providence, Rhode Island: Brown University.
- Francis, W. N. & Kučera, H. (1982). *Frequency Analysis of English Usage: Lexicon and Grammar*. Boston: Houghton Mifflin.
- Morales, A. (1986). *Léxico básico del español de Puerto Rico*. San Juan de Puerto Rico: Academia Puertorriqueña de la Lengua Española.
- Sinclair, J. M. (Ed.) (1987). *Collins COBUILD English Language Dictionary*. London: Collins ELT.
- Juilland, A.; Brodin, D. & Davidocitch, C. (1990). *Frequency Dictionary of French Words*. La Haya/Paris: Mouton.
- Crystal, D. & Moreno Cabrera, J. C. (1994). *Enciclopedia del Lenguaje de la Universidad de Cambridge*. Madrid: Taurus, [edición original: Crystal, D. (1987). *The Cambridge Encyclopedia of Language*. Cambridge: Cambridge University Press].
- Justicia, F. (1995). *El desarrollo del vocabulario. Diccionario de frecuencias. (niños de 6 a 13 años de Andalucía oriental)*. Granada: Universidad de Granada.
- Ávila Muñoz, A. M. (1999). *Léxico de frecuencia del español hablado en la ciudad de Málaga*. Málaga: Servicio de Publicaciones de la Universidad de Málaga.
- Terrádez Gurrea, M. (2001). Frecuencias léxicas del español coloquial: análisis cuantitativo y cualitativo. *Cuadernos de Filología, Anejo XLI*. Valencia: Universitat de València.

Obras ----- Nº de palabras	Juillard & Chang Rodríguez (1964)	Carroll, Davies & Richman (1971)	Patterson & Urrutibéh eity (1975)	Francis & Kučera (1982)	Morales (1986)	Sinclair COBUILD Corpus (1987)	Juillard <i>et</i> <i>al.</i> (1990)	Crystal (1994)	Justicia (1995)	Ávila Muñoz (1999)	Terrádez Gurrea (2001)
10	36.55%	25%					30.52%				25%
15								25%			
50										63%	50%
100	66.20%	50%	30%		53%		62.66%	60%		73%	60%
500					68%					89%	79%
700						70%					
1.000	85.5%	75%	50%	72%			85.58%	85%	90.8%	93%	84%
1.500						80%				95.5%	88%
2.000		81%		80%						96.6%	90%
2.500						86%				97%	92%
3.000		85%									
4.000								97%			
5.000		89%	90%							99.7%	

TABLA 8. Relación entre el número de palabras frecuentes y el porcentaje de cobertura textual obtenido con ellas según los estudios

4.1.2. Lemas, vocablos (*types*), palabras (*tokens*) y familias de palabras

Como hemos dicho, en función del tamaño de las muestras, estas proporciones pueden verse alteradas, pero en general sabemos que, cualquiera que sea la lengua y corpus que consultemos, los resultados sobre la cobertura textual suelen aproximarse. También conviene aclarar algunas cuestiones en relación con los conceptos y la terminología, que a veces también explican estas divergencias en las cantidades. Así, al hablar de cobertura léxica en léxicos de frecuencia o corpus, comentamos cuántas palabras son necesarias de las que se encuentran en las primeras posiciones, es decir, de las más frecuentes para cubrir determinado porcentaje de cobertura textual del total del texto.

Pues bien, además de las diferencias que ya comentamos en su momento a propósito del concepto de “palabra”, a la hora de contar las unidades de un corpus, existen otras, que afectan de manera directa al resultado de los recuentos. Por ejemplo, un corpus puede contar lemas o no, dependerá de si está lematizado o no y de cuál es la unidad básica de recuento. Teniendo en consideración que el objetivo es contar formas idénticas, conviene que el corpus esté lematizado y que etiquete todas las ocurrencias y las diferentes formas de un lema bajo una misma forma. Si cuenta lemas, entonces es que ha sometido las palabras a un proceso mediante el cual se representan bajo una única forma léxica todas las posibilidades paradigmáticas de un término. Esta lematización reduce las formas declinadas y conjugadas a formas básicas o radicales. El elemento resultante de este proceso reductivo es el *lema*. Se puede entender que varias formas léxicas distintas pertenecen a un mismo lema si después de borrar su segmento común en el principio de la palabra, los segmentos que quedan figuran en una lista de desinencias posibles y forman parte de una serie de desinencias compatibles, explica De Kock (De Kock 1983: 20 y Gómez Molina & De Kock 1982). El lema es la unidad sobre la que opera la estadística léxica en la mayor parte de las ocasiones (Ávila Muñoz 1999: 58). Realizar un recuento de frecuencias supone necesariamente descomponer un texto en unidades más simples a las que se concede significación y utilidad. Lógicamente el concepto de lema no es producto de la lingüística computacional, pues en lexicografía ha tenido una tradición arraigada durante siglos. Además, a lo largo de la historia,

los realizadores de diccionarios han denominado de manera diversa el mismo concepto: *voz guía*, *voz clave*, *entrada*, etc.

En otras ocasiones, la estadística se aplica directamente sobre cada una de las distintas presentaciones formales de los términos sin que se lleve a cabo el proceso de lematización, por ejemplo en los recuentos de Rodríguez Bou (1952) o Alameda y Cuetos (1995: 117), aunque en este caso los autores reconozcan la desventaja que causa la no lematización. Para hacernos una idea, este tipo de recuentos supone contar como palabras diferentes las distintas formas flexivas de un verbo, por ejemplo “fui” o “soy”, en lugar de ofrecernos los datos sobre el verbo “ser”, aunque luego se nos indique dentro de él las frecuencias concretas de cada persona y tiempo, que sería lo más útil.

Esta labor no se realiza sobre un simple listado de palabras, sino sobre la colección de textos que componen el corpus, a partir de la cual, mediante un proceso de reducción, obtenemos una unidad de análisis creada *ad hoc* para operar de manera computacional, a la que llamamos *lema*, que, como ya sabemos, hace referencia a las distintas formas flexivas con que puede presentarse una determinada palabra en un texto dado. En la estadística léxica el término lema suele usarse como sinónimo de vocablo o tipo léxico (*type*). Se trata de elementos integrantes del léxico y proporcionan información sobre los significantes de los lexemas, sobre la clase formal a la que pertenecen y sobre las categorías gramaticales con que se actualizan en el discurso. El proceso de lematización es, posiblemente, el asunto de mayor envergadura referido al análisis computacional de textos y casi todos los trabajos de lexicografía computacional combinan procesos de lematización automáticos con otros manuales, por lo que es importante su justificación teórica para exponer la metodología y objetivos, que casi todos los corpus y recuentos ofrecen. Uno de los corpus más recientes del español, el de Davies (2006), se detiene en esta cuestión en su introducción, en la que critica a los que no han trabajado con lemas y explica el proceso de lematización llevado a cabo en su diccionario: “In order to create a useful and accurate listing of the top 5,000 words in Spanish, the entire 20 million words of text needs to first be tagged and lemmatized” (Davies 2006: 3).

Y prosigue, explicando que etiquetar significa asignar una categoría

gramatical a cada palabra del corpus, para crear un lexicón del español que contiene más de 400.000 formas diferentes, con sus categorías y lemas. El autor explica que el término *lema*:

refers to the “base word” or “dictionary headword” to which each individual form belongs. For example, the following are five word forms from the 400,000 word lexicon: lápices / lápiz / noun_masc_pl, tengo / tener / verb_present_1pers_sg [Davies 2006]

Donde “lápices” y “tengo” son las *word forms*, “lápiz” y “tener” son los *lemas* y “noun_masc_pl” y “verb_present_1pers_sg” es la información sobre la *part of speech*. El autor continúa con los detalles sobre el proceso de desambiguar las palabras más difíciles y con la explicación de cómo se desarrolla la parte de los ajustes técnicos e informáticos.

Así pues, además de las diferencias debidas a los distintos criterios seguidos para confeccionar los corpus y seleccionar los textos usados, existen otras que también son importantes, como la consideración de lo que es una “palabra” en cada estudio, ya que muchos autores se refieren a familias de palabras (*headwords and their main inflections and derivations*), un concepto que incluye varios *types* y, generalmente, más de un lema. Un autor que usa recurrentemente el término “word family” es Nation, que explicó el concepto en una publicación con este mismo nombre (Bauer y Nation 1993):

Dealing with growth in morphological knowledge involves consideration of the idea of a “word family”. From the point of view of reading, a word family consists of a base word and all its derived and inflected forms that can be understood by a learner without having to learn each form separately. So, watch, watches, watched, and watching may all be members of the same English word family for a learner with a command of the inflectional suffixes of English. As a learner’s knowledge of affixation develops, the size of the word family increases. The important principle behind the idea of a word family is that once the base word or even a derived word is known, the recognition of other members of the family requires little or no extra effort. [Bauer & Nation 1993: 253]

Por este motivo el conocimiento morfológico determina también el desarrollo de nuestras capacidades lectoras y es importante sistematizarlo cuando enseñamos vocabulario y leemos textos. El conocimiento morfológico determina que podamos conocer muchas palabras de una misma familia como lectores. Y, mejor aún, determina que podamos reconocer palabras de una misma

familia sin haberlas aprendido, pero es un proceso que, conociendo la palabra base y los afijos por separado, requiere muy poco esfuerzo. En inglés, por ejemplo, enseñar una palabra con su familia depende de criterios como frecuencia, regularidad, productividad y predictibilidad, que se aplican a los afijos clasificándolos en 7 niveles diferentes. Estos niveles sirven como referencia para la enseñanza y para la creación de diccionarios. Bauer y Nation nos muestran algunos ejemplos de estos niveles para mostrar cómo se desarrolla una familia de palabras y lo fácil que puede resultar reconocerlas. Reproducimos uno aquí (Bauer & Nation 1993: 254):

word family

2	develop develops developed developing
3	developable undevelopable developer(s) undeveloped
4	development(s) developmental developmentally
5	developmentwise semideveloped antidevelopment
6	redevelop predevelopment

CUADRO 5. Ejemplo de una familia de palabras ordenada en niveles, según los criterios de Bauer & Nation (1993: 254)

Por otro lado, tenemos dos conceptos propios de la lingüística del corpus, sobre todo en el ámbito inglés, y que se emplean comúnmente en la disciplina. Son los términos *type* y *tokens*, que ya hemos presentado más arriba (*vid. supra* § 3.4.1.). Al hablar de *type* nos referimos a las palabras diferentes que existen en un corpus, que no tienen que coincidir necesariamente con los lemas o las palabras base, y que normalmente se acompañan de datos sobre frecuencia y

uso. En español, se ha optado por usar el término *vocablo* en estos casos. Así, *type* o *vocablo* es la única forma bajo la que se agrupan los *tokens* de un corpus, que son todas las palabras, formas u ocurrencias. La suma de todos los *tokens* equivale al tamaño del corpus.

Estos términos han sido usados para medir la variación léxica y la frecuencia de las palabras en un texto. Así, utilizamos la frecuencia para analizar la dificultad de un texto, ya que, si la mayor parte de las palabras se repiten, entonces necesitaremos pocos vocablos (*types*) para entenderlo, pero si son pocas las palabras que se repiten, entonces necesitaremos conocer un mayor número de *types*. El grado de diversidad léxica se suele expresar con la ratio de *type-token* (TTR o *type-token ratio*), que se calcula con la siguiente fórmula (Laufer & Nation 1995):

$$\text{Variación léxica} = \frac{\text{n}^{\circ} \text{ de palabras diferentes (types)}}{\text{n}^{\circ} \text{ total de palabras en el texto (tokens)}} \times 100$$

La TTR se usa para comparar dos corpus en términos de complejidad léxica. Si el resultado está cerca de 100, entonces será mayor el grado de complejidad, y si está cerca de 10 se repetirán más palabras en el texto y resultará, en principio, más sencillo. Esta sería una forma aproximativa de medir la complejidad del corpus y de establecer comparaciones. No es una fórmula compleja y sí puede resultar de utilidad. Posteriormente se han perfeccionado estos cálculos por otros autores para medir la riqueza léxica de un texto o de un estudiante de una L2. Un ejemplo es la STTR (*standardised type/token ratio*).

Podemos ilustrar con un ejemplo este índice. Así, si un párrafo contiene 245 palabras totales (*tokens*) pero solo 132 palabras diferentes (*types*), la ratio será de 53.88 ($132/245 \times 100$), lo cual supone un tratamiento léxico complejo. Normalmente los textos escritos tienen una ratio superior a 40, que puede fluctúa entre los 36 y los 57 puntos, por ejemplo, mientras que los textos orales suelen estar por debajo de 40. Obviamente, este índice no siempre es indicador de la dificultad de un texto, porque se puede dar el caso de que se repitan mucho las palabras, pero que se trate de palabras técnicamente complejas, por ejemplo en un artículo científico. Por otro lado, hay textos sencillos que, al ser muy largos, arrojan una TTR que no refleja su grado de complejidad. Por ejemplo, una

historia para niños como *Mi vecino Totoro* (Miyazaki 1992) en su traducción al inglés tiene una TTR de 55.59, porque el número de *types* no se incrementa normalmente tan rápido como el de *tokens* a medida que el texto se alarga (Schmitt 2000: 75).

No obstante, no todas las palabras de un texto son iguales, por lo que para medir la complejidad léxica podemos recurrir al índice de densidad léxica, que toma en consideración el número de palabras con contenido léxico (sustantivos, adjetivos, verbos, adverbios) y deja de lado las palabras gramaticales. La fórmula para medirlo es (Laufer & Nation 1995):

$$\text{Densidad léxica} = \frac{\text{n}^{\circ} \text{ de palabras con contenido léxico}}{\text{n}^{\circ} \text{ total de palabras en el texto (tokens)}} \times 100$$

Los textos con muchas palabras léxicas tienen una densidad léxica mayor, es decir, mayor complejidad, y viceversa.

La diferencia entre los conceptos *token*, *type*, *lema* y *word family* es importante para establecer a qué cantidad de palabras se refieren estas cifras a la hora de la elaboración de listados de frecuencias. Un *lema* incluye un menor número de palabras que una *word family* y, por lo tanto, un alumno que conozca 2.000 *word families* sabrá muchas más palabras diferentes (*types*) que si considerásemos los *lemas*. Cuando hablamos de cobertura léxica textual de un corpus, nos estamos refiriendo a esta cuestión de *lemas*, *types* y *tokens*. Es decir, al describir que con los 5.000 vocablos más frecuentes de un corpus obtenemos una cobertura de 97%, estamos manifestando que conseguimos cubrir el 97% de las *tokens*. Estos vocablos pueden entenderse bien como *lemas*, bien como *types*, en función de la unidad básica de medida que establezca ese diccionario de frecuencia en concreto. Debemos tener en cuenta que el número de *types* y de *lemas* no es exactamente el mismo, ya que en un texto contamos más de los primeros que de los segundos. Por ejemplo, Cobb emplea estos conceptos para valorar las necesidades de cobertura léxica:

[...] in English just a few word types account for most of the word tokens in any text. Ten words account for 23.7 % of the ink on any page (repeated words like "the" and "of"). Just 1000 word families account for more than 70% of the words or ink, and 2000 account for about 80%. So you need to find out what these

2.000 word families are and be sure you know them. [Cobb 2003: 2]

Para comparar la equivalencia en términos numéricos entre estos diferentes conceptos, podemos recurrir a un estudio hecho sobre el inglés que ofrece tablas comparativas de algunos manuales (Criado & Sánchez 2012: 83). Tomamos, en concreto, la que analiza el manual de nivel A2 *Face2Face Elementary Student's Book (F2F)* de Redston y Cunningham (2005). El concepto de *Range* se toma de Nation³³, quien usa una herramienta computacional diseñada por él mismo para determinarlo. Esta herramienta cuenta y clasifica el vocabulario en categorías de frecuencias, catalogándolas en listas de 1.000 palabras por su frecuencia en el inglés general. Las categorías de frecuencias se llaman *rangos*, así que la palabra rango (*Range*) se usa para designar cada grupo de 1.000 palabras en estas listas de frecuencias, que pueden llegar a ser 14.

Partiendo de esto, Criado y Sánchez determinan que el manual *F2F* contiene 36.721 ocurrencias (*tokens*), 2.509 de las cuales son palabras diferentes (*types*). Las familias de palabras ascienden a 1.719. Los *types* se distribuyen de esta manera:

- 1.207 *types* (más de 845 lemas) pertenecen al Rango 1 (las primeras mil palabras de mayor frecuencia);
- 365 *types* (sobre 255 lemas) pertenecen al Rango 2 (las segundas mil palabras más frecuentes);
- 136 *types* (cerca de 95 lemas) pertenecen al Rango 3 (las terceras mil palabras más frecuentes).

Además, hay que tener en cuenta todavía 801 *types* que están fuera de estos tres rangos, por lo que no pertenecen a las tres primeras listas de las 3.000 palabras más frecuentes. Todos estos datos se pueden ver resumidos en la siguiente tabla, que muestra *tokens*, *types*, *word families* por Rango de 1, 2 y 3 en el manual *F2F* (tomada de Criado & Sánchez 2012: 83):

³³ Se puede consultar en el sitio: <http://www.victoria.ac.nz/lals/staff/paul-nation/nation.aspx>

WORD range	TOKENS/%	TYPES/%	FAMILIES
one	31,752/86.47	1,207/48.11	691
two	1,886/5.14	365/14.55	278
three	552/1.50	136/5.42	109
off-ranges	2,531/6.89	801/31.93	(estimated) 641⁴
Total	36,721	2,509	1,719

TABLA 9. Relación entre *tokens*, *types* y *word families* del manual *F2F* organizados por *Range* según Criado & Sánchez (2012: 83)

En un artículo posterior, Nation (2006: 62-67) ofrece un nuevo estudio basado en estas 14 listas de 1.000 palabras frecuentes. En él, antes de explicar su estudio sobre cuánto vocabulario es necesario para entender concretamente diferentes tipos de textos orales y escritos, comenta las diferencias entre *tokens*, *types* y familias de palabras que existen en cada uno de estos 14 rangos, y los problemas que puede haber al trabajar con familias de palabras basándose en corpus que clasifican por frecuencia, rango y dispersión de *types* y lemas. El autor elabora las 14 listas basándose en el BNC (*British National Corpus*³⁴), creado en los años 90 y revisado en la década siguiente. Ofrece la información en lemas y *types*, y coteja los datos con los de otros corpus para validar la fiabilidad de sus listas. En el Cuadro 6 reproducimos la información sobre la correspondencia. En la primera columna está representada su lista de bloques de 1.000 palabras y a la derecha la equivalencia de *tokens*, *types* y familias de palabras con el LOB (*Lancaster-Oslo/Bergen Corpus*, elaborado en los años 70)³⁵.

³⁴ El *British National Corpus* (BNC) es un corpus de 100 millones de palabra con muestras escritas y orales de una amplia gama de fuentes en inglés. El corpus cubre el inglés británico de final del siglo XX. Sus textos son de una amplia variedad de géneros con la intención de ser suficientemente representativo del inglés británico hablado y escrito de ese tiempo. El proyecto para la creación del BNC contó con la colaboración de tres editores (Oxford University Press, Longman y W. & R. Quartes; el primero como líder), dos universidades (la University of Oxford y Lancaster University) y la *British Library*. La creación del BNC empezó en 1991 y el proyecto estuvo acabado en 1994. No ha habido ninguna adición de nuevas muestras después de 1994, pero el BNC experimentó revisiones leves antes de la liberación de la segunda edición *BNC World* (2001) y la tercera edición *BNC XML Edition* (2007).

³⁵ *Lancaster-Oslo/Bergen Corpus* (comúnmente abreviado como LOB Corpus) es una colección de un millón de textos del inglés británico. Fue acabado en 1970 gracias a la colaboración entre la University of Lancaster, la University of Oslo y el Norwegian Computing Centre for the Humanities de Bergen. Este corpus nació como homólogo del Brown Corpus, creado por Henry Kučera y W. Nelson Francis para el inglés americano en 1960.

Una de las evidencias es que, según avanzamos en las listas, el número de *tokens* y de *types* se reduce continuamente y se aproxima más. Esto quiere decir que en el bloque de las primeras 1.000 palabras hay más *tokens* y *types* que en las del bloque número 14. En este cotejo se manifiesta una incoherencia entre los *tokens* de la lista 9 y 10, que no siguen esta progresión decreciente. Finalmente, la número 15 se corresponde con nombres propios. En el cuadro también vemos que disminuye la correspondencia de las familias de palabras creadas sobre el *BNC* en relación con las del *LOB*.

Word list (1,000)	Token (%)	Types (%)	Families
1	78,944 (77.86)	4,487 (10.1)	998
2	83,477 (8.23)	4,131 (9.34)	998
3	37,511 (3.70)	3,239 (7.32)	998
4	18,198 (1.79)	2,683 (6.07)	998
5	10,495 (1.04)	2,226 (5.03)	969
6	7,080 (0.70)	1,789 (4.04)	928
7	6,633 (0.65)	1,542 (3.49)	887
8	4,096 (0.40)	1,382 (3.12)	836
9	3,217 (0.32)	1,118 (2.53)	734
10	3,228 (0.32)	1,025 (2.32)	719
11	1,609 (0.16)	753 (1.70)	587
12	1,434 (0.14)	646 (1.46)	498
13	1,211 (0.12)	529 (1.20)	441
14	973 (0.10)	339 (0.77)	288
15	18,519 (1.83)	2,878 (6.51)	2,878
Not in the lists	26,821 (2.65)	15,463 (34.96)	?????*
Total	1,013,9469	44,230	13,747

TABLA 10. Número de *tokens*, *types* y *word families* que contienen las 15 listas elaboradas por Nation (2006: 62-67)

A pesar de las dificultades que puedan existir para trabajar con listas de familias de palabras, el autor justifica esta elección de la siguiente manera:

The assumption that lies behind the idea of word-families is that when reading and listening, a learner who knows at least one of the members of a family well could understand other family members by using knowledge of the most common and regular of the English wordbuilding devices. There is research evidence from native speakers that the wordfamily is a psychologically real unit. Most L2 learners, however, will not have word-families as inclusive as those of native speakers. It should also be noted that adult native speakers will have much larger wordfamilies than the ones used in the present study. [Nation 2006: 67]

Así pues, detrás de la idea de las familias de palabras para las destrezas

receptivas (leer y escuchar) está la asunción de que quien conoce al menos uno de los miembros de una familia puede entender otros mediante la aplicación de las reglas más comunes de creación de palabras con los afijos conocidos, como ya hemos comentado. Hay investigaciones con hablantes nativos, aunque parece que el número de palabras de una misma familia que un hablante extranjero conoce es menor con respecto a los que tienen esta lengua como materna. En el estudio de Nation se tuvo esto en cuenta al trabajar con familias de palabras para evaluar la cantidad de las que son necesarias para comprender textos, novelas, películas y periódicos de forma adecuada, y se trabajó con un número de miembros de una familia de palabras más reducido que el que los nativos demuestran conocer.

Lo que sí se puede destacar es que, al trabajar con la evaluación de las destrezas receptivas, la unidad básica de medida recomendable puede ser entonces la familia de palabras y, dado que los diferentes ítems que pertenecen a una misma familia se perciben como si fueran una sola unidad al leerlas y oír las, no resulta descabellado equipararlas con otros estudios que hablan de lemas, *types* o palabras. La frecuencia, rango y dispersión de los lemas pueden verse como homólogos y aplicables también a sus familias de palabras, aunque estas incluyan varios lemas.

Otra cuestión que apunta Nation en relación con sus 14 listas de palabras frecuentes es que considera la lista número 6 como la que se corresponde con un estadio avanzado de aprendices. Laufer y Ravenhorst-Kalovsky (2010) comparten los mismos postulados que Hu y Nation (2000) y Nation (2006), a pesar de que los primeros hablen de palabras y estos de familias. Para las capacidades productivas, no obstante, la recomendación es la contraria, y se prefiere tomar como unidad básica de medida el lema o los *types*, pues se consideran más representativos. Además, cada lema puede tener diferentes colocaciones y normas sintácticas para su combinación y uso (Chung & Nation 2003).

4.2. Conocimiento léxico y comprensión lectora

Al hablar de comprensión lectora, tanto en lengua materna (Anderson & Freebody 1981, Stanovich 1986, Sternberg 1987) como en L2 (Alderson 1984, Llinares 1990; Laufer 1989, 1992, 1997a, 1997b), existe una serie de factores que determinan el éxito de la lectura. Además del conocimiento del vocabulario, entre ellos encontramos el conocimiento previo del tema y la aplicación de múltiples estrategias de lectura, como la predicción del contenido, las inferencias, la identificación del tipo de texto y su estructura, las ideas principales y secundarias del texto y de cada párrafo, etc. No obstante, en diferentes estudios desde los años 80 se ha demostrado la importancia que adquiere el conocimiento léxico (Cartaya Febres 2011: 104-5). Así, en alguno de los estudios que se han hecho de la relación entre estos factores y la comprensión lectora, se ha concluido que la variable que presenta la más alta correlación en el caso de una L2 es la amplitud del conocimiento léxico (Nassaji 2003). En conclusión, el éxito de la comprensión lectora depende directamente del conocimiento léxico. Y no debemos olvidar que hay estrategias de lectura que no se activan si este no es suficientemente alto, como ocurre por ejemplo con las inferencias léxicas.

En un estudio sobre la adquisición del vocabulario en una L2 y la lectura efectiva llevado a cabo en Venezuela con estudiantes de inglés se explora también la importancia del umbral léxico de comprensión lectora y la eficacia de diferentes métodos y estrategias de desarrollo de la competencia léxica de los aprendices (Llinares *et al.* 2008). Las autoras de este proyecto de investigación parten del conocimiento de los datos de estudios previos, que sugieren que los estudiantes que obtienen entrenamiento en el uso de estrategias aumentan su vocabulario (O'Malley & Chamot 1990) y que todos los discentes favorecen el uso de estrategias mecánicas para aprender léxico desconocido en la L2, especialmente aquellos con niveles más bajos de suficiencia (Schmitt 1997, Riazi & Alvari 2004). La investigación se lleva a cabo con estudiantes que habían acabado la enseñanza secundaria y pretendían cursar estudios superiores de tipo científico y técnico en inglés. Se pretendía probar su nivel de conocimientos en

esta lengua y orientar a los docentes para ajustar sus prácticas pedagógicas a las condiciones de su alumnado. Durante un período de doce semanas se dividió a 129 estudiantes en grupos de trabajo. A todos se les ofreció como material didáctico un glosario L2 -L1 de palabras de alta frecuencia del inglés general y académico, además de un apéndice sobre la importancia del conocimiento léxico para el desarrollo de las destrezas de comprensión de lectura y se les presentó un completo inventario de estrategias de aprendizaje léxico. Para medir los resultados en los sujetos que participaron en este estudio se les diseñó un plan de evaluación con pruebas de vocabulario que contaban un 20% en el total de las evaluaciones de la asignatura. Con los tres grupos de trabajo se usaron enfoques pedagógicos diferentes para el desarrollo de técnicas y estrategias con las que los estudiantes pudiesen avanzar en su conocimiento léxico. Con un grupo se estimuló la motivación extrínseca a través de la división y estructuración del glosario en sublistas que se evaluarían periódicamente de manera independiente (denominada *Intensive Vocabulary Testing – IVT*). Con otro grupo se llevó a cabo el entrenamiento de los estudiantes en el uso de una técnica específica de memorización de léxico desconocido (*Memory Enhancement Condition – MEM*). Y con el tercer grupo se potenció la práctica de distintas modalidades de ejercitación en estrategias de aprendizaje léxico que estimulen el procesamiento de la nueva palabra a diferentes niveles cognitivos (*Personalized Online Exercises – POE*) (Cartaya Febres 2011: 111-114). Los resultados hablan por sí solos, pues en los tres grupos con los que se pusieron en práctica estas técnicas el desarrollo léxico fue considerablemente mayor que en el grupo de control, que tuvo un aumento exiguo (de un 7.18). Los otros tres grupos muestran que los diferentes enfoques contribuyeron en diferente medida para mejorar el conocimiento del vocabulario: el método MEM en un 27.34, el de POE en un 20.55 y el de IVT un 18.54. Como se puede observar, el enfoque pedagógico que produjo mayor progreso fue el que entrenaba a los sujetos en el uso de una técnica específica de memorización de léxico desconocido, resultados que ya había apuntado Cobas (2003). Además, el entrenamiento en el uso de la técnica de memorización léxica, que es la estrategia pedagógica que produce los mejores resultados, presenta una serie de ventajas: requiere poco esfuerzo por parte del docente, consume muy poco tiempo de clase y le proporciona al estudiante una herramienta que puede

utilizar de manera autónoma en sus futuros esfuerzos para incrementar su conocimiento léxico en la L2 (Cartaya Febres 2011: 113). Por consiguiente, Cartaya Febres (2011) afirma que:

[...] resulta esperanzador constatar que la adopción de sencillas estrategias pedagógicas encaminadas a propiciar el desarrollo léxico produce a corto plazo incrementos en la amplitud del conocimiento léxico del estudiante que resultan estadísticamente significativos (Cartaya Febres 2011: 114-115).

Por este motivo apuesta por incluir en los programas de los cursos de una lengua extranjera a nivel universitario contenidos orientados a adquirir las palabras de mayor frecuencia en textos académicos. Esto es así porque si el estudiante es capaz de reconocer automáticamente el significado de unidades léxicas tanto de la lengua general como pertenecientes al léxico semiespecializado conseguirá también alcanzar el umbral léxico que le permita evitar los obstáculos léxicos que interfieren en la comprensión de un texto (Laufer 1997a y 1997b), aumentar el porcentaje de cobertura textual (Laufer & Ravenhorst-Kalovsky 2010, Nation 2006), transferir de manera exitosa las estrategias de lectura que ya domina en su L1 (Laufer 1997a y 1997b) y continuar ampliando su vocabulario en la L2 a través de la ampliación léxica que de manera incidental obtiene por medio de la inferencia contextual (Pulido 2009).

4.3. Cobertura textual y tipos de texto

Podemos también partir de las listas de frecuencia establecidas por Nation (2001) para intentar contabilizar los porcentajes de comprensión que ofrecen los diferentes tipos de textos en función del léxico que usan. Para ello, se parte de la *Academic Word List (AWL)* establecida por Coxhead (2000), que sustituyó al listado anterior: la *University Word List (UWL)*. La AWL contiene 570 familias de palabras que son especialmente frecuentes en los textos académicos, pero no se encuentran entre las 2.000 familias de palabras más frecuentes fijadas por Nation en sus listas. Aunque se trate de estudios sobre el inglés y la equivalencia no sea exacta, se aproxima lo suficiente para dar por válidos sus resultados como guía también para el español, por lo menos hasta que contemos

con datos propios. Los resultados sobre las frecuencia de las palabras en diferentes tipos de texto se muestran en la tabla siguiente (Nation 2001: 17):

Niveles	Conversación	Ficción	Periódicos	Textos académicos
1 - 1.000	84.3%	82.3%	75.6%	73.5%
1.000 - 2.000	6%	5.1%	4.7%	4.6%
Palabras académicas	1.9%	1.7%	3.9%	8.5%
Otras	7.8%	10.9%	15.7%	13.3%

TABLA 11. Número de palabras y relación con su cobertura léxica textual en los diferentes tipos de textos (Nation 2001: 17)

Estas 2.000 familias de palabras rondan el 80-85% en la mayoría de los textos escritos, y si se les suman las 570 de la lista de las más frecuentes en los textos académicos, se puede alcanzar una cobertura de más del 85% también en este tipo de textos. Si a estas, además, se unen algunas palabras técnicas hasta llegar cerca de las 4.000, este porcentaje puede subir a un nivel considerado óptimo para la lectura de textos relativamente complejos. Un estudiante de una L2, por lo tanto, puede leer bien un texto académico con un vocabulario relativamente reducido, si está bien escogido. No obstante, conviene tratar estos datos con mayor prudencia cuando se trasladan al ámbito hispano, ya que lo normal en este no es trabajar basándose en familias de palabras, sino en palabras, y las primeras suponen más unidades que las segundas. De esta forma se explica que hablemos de familias de palabras en muchos de los trabajos realizados sobre el inglés, mientras que usamos el término palabras o vocablos en los que se refieren al español.

La Tabla 10 muestra que las primeras 2.000 familias de palabras son más frecuentes en el inglés hablado que en el escrito, lo cual ocurre también en el español, como muestran los estudios en los que se comprueba que la cobertura textual evoluciona más rápidamente en los corpus elaborados a partir de textos orales que en los elaborados a partir de textos escritos. Recordemos, por ejemplo, que las 100 primeras palabras más frecuentes en el corpus oral de Ávila Muñoz (1999) representan el 73 % del total de palabras considerado, mientras

que las 100 primeras palabras en el corpus escrito de Morales (1986) representan el 53% del total. Son datos que coinciden en recuentos llevados a cabo en otras lenguas. Nation (2001) constataba que las 1.000 primeras palabras del listado oral de Schonell *et al.* (1956), *Oral Vocabulary of Australian Worker (OVAW)* cubrían un 94% del material y en el diccionario de frecuencias de Justicia (1995: 155) las 1.000 primeras palabras alcanzan un 90,8% de la muestra total. De ahí que las 50 o 100 palabras más frecuentes cubran una mayor proporción de *tokens* en el discurso oral que en el escrito. Otra característica de los listados orales es su más rápida estabilización. Mientras que en los listados de frecuencias generales la cobertura léxica tiende a estabilizarse a partir de las 1.000 primeras unidades léxicas, con lo que crece muy lentamente, en los corpus orales, por ejemplo el de Ávila Muñoz (1999: 93) y el de Terrádez Gurrea (2001: 67), esto ocurre a partir de las 500. Esto significa que con dicha cantidad se da cuenta de la mayoría de ocurrencias totales. El vocabulario que no se ha cubierto en ese tramo se considera, además, de baja frecuencia.

Según esto, los hablantes de una lengua elaboran una gran parte de sus mensajes orales con un limitado número de vocablos, pues hay una tendencia a la economía lingüística en general. Aplicado esto al léxico, el registro oral se caracteriza por la repetición de términos, lo cual significa que se requieren menos unidades léxicas para desenvolverse adecuadamente en el discurso oral informal que en el escrito. A esto se debe que el aumento rápido de cobertura textual constituya un rasgo característico de los recuentos basados en la lengua hablada.

A propósito de la cobertura textual y los tipos de texto, Nation (2006) ofrece datos que ponen en relación sus listas —con 14 niveles de 1.000 palabras más frecuentes cada una obtenidas del BNC, recordemos— para medir la cobertura textual de interacciones orales, lecturas de ficción graduadas, periódicos, películas infantiles y novelas de ficción famosas. Sobre las conversaciones de radio —interacciones, entrevistas, comentarios espontáneos y discusiones— y las conversaciones amistosas entre familiares y amigos, los datos muestran que con las 2.000 familias de palabras alcanzamos una cobertura del 89.41 en las conversaciones y entrevistas de radio y de un 89.35 en las

conversaciones amistosas. En las 3.000 familias de palabras siguientes las cifras son de 96.52 y de 96.03, respectivamente, y con las 6.000 alcanzamos el 98.26 y el 97.67. A estos cálculos se les suma los porcentajes de cobertura que ofrecen los nombres propios, que se calculan en un 1.29 para las intervenciones de radio y en un 1.03 en las conversaciones familiares. Estos resultados revelan que con las 3.000 familias de palabras llegamos al 95% de cobertura y con las 6.000 o 7.000 al 98%. Advierte el autor que la lengua oral hace un uso mucho mayor de las palabras de mayor frecuencia que la lengua escrita y que, no obstante, necesitamos un porcentaje de cobertura del 98% para hacer frente eficazmente a los intercambios orales espontáneos y transitorios, pues no podemos, como en los textos escritos, volver a leer las palabras ni tenemos tiempo para reflexionar demasiado.

Al medir el número de familias de palabras que necesitamos para comprender de forma natural la mayor parte de las palabras de una película para público infantil —en este caso se trata en concreto de *Shrek*— los resultados muestran que con las primeras 1.000 familias de palabras más frecuentes cubrimos un 81.54 del total de *tokens* y con las 4.000 llegamos al 95.27% de cobertura. Estos datos sobre la cobertura en interacciones orales difiere bastante de los que se observan en la comprensión de periódicos³⁶ y de novelas conocidas³⁷ ya que con las 2.000 familias de palabras más frecuentes solo se cubre entre el 81.79 y el 84.33 de los textos, en el caso de los periódicos, y el 86.95% y el 91.71, en el de las novelas. Con las 4.000 se alcanzan ya las cifras que rondan el 95% de cobertura y con las 8.000 y las 9.000 se llega al 98%. Cuando cubrimos el 98% con 8.000 o 9.000 palabras estamos hablando de encontrar un vocablo desconocido por cada 50. En cambio, si las lecturas son graduadas, como en el caso de la novela *The Picture of Dorian Gray* adaptada, entonces los porcentajes de cobertura mejoran incluso con respecto a la lengua oral, pues hablamos de 91.20 con las 1.000 familias de palabras, de 96.75 con las 2.000 y de 98.86 con las 3.000.

³⁶ Trabaja con 5 periódicos diferentes y extrae los datos de la sección dedicada a ellos en varios corpus diferentes.

³⁷ *Lord Jim* de Joseph Conrad, *Lady Chatterley's Lover* de D.H. Lawrence, *The Turn of the Screw* de Henry James, *The Great Gatsby* de F. Scott Fitzgerald, y *Tono-Bungay* de H.G. Wells.

Teniendo todos estos datos en mente, Nation intentó establecer una correlación aproximada entre sus listas de las palabras más frecuentes y los porcentajes de cobertura en textos escritos y en textos orales. Los resultados que ofrece son estos (Nation 2006: 79):

Nivel y número de familias de palabras	Porcentaje medio aproximado de cobertura en textos escritos	Porcentaje medio aproximado de cobertura en textos orales
1º - 1.000 palabras	entre 78%- 81%	81% - 84%
2º - 2.000	+ entre 8%-9%	+ entre 5% -6%
3º - 3.000	+ entre 3% - 5%	+ entre 2% - 3%
4º-5º entre 4000 y 5.000 palabras	+3%	+ entre 1,5% - 3%
6º-9º - entre 6.000 y 9.000 palabras	+2%	+ entre 0,75% - 1%
10º-14º entre 10.000 y 14.000 palabras	<1	+0,5%
Nombres propios	+ 2% - 4%	+ entre 1% - 1,5%

TABLA 12. Relación entre los porcentajes de cobertura en textos escritos y orales y las listas de las palabras más frecuentes de Nation (2006: 79)

4.4. Umbral léxico de comprensión lectora y comprensión oral

Hasta ahora nos hemos detenido en explicar la relación entre el número de palabras conocidas por un hablante, nativo o no, y los porcentajes de comprensión lectora y de comprensión oral que consigue con diferentes cantidades de palabras, para lo que usualmente se usa la relación entre lemas y *tokens* o entre familias léxicas y *tokens*. No obstante, tan importante como conocer estos porcentajes es saber a partir de qué niveles la comprensión permite entender realmente el contenido de un texto o de una conversación. Sobre cómo decidir el número de palabras del vocabulario receptivo necesario para leer y entender una conversación, Nation consideraba tres opciones diferentes (Nation 2006): intentar medir el léxico de la lengua, que podríamos cuantificar según él en 114.000 familias léxicas; intentar medir el vocabulario

receptivo de un nativo culto (lo que él llama “well-educated”), que se puede cifrar entre 8.000 o 9.000 familias léxicas; o bien estudiar cuánto vocabulario necesitamos para hacer un uso adecuado de las capacidades de comprensión en tareas habituales como leer un texto, ver una películas, leer el periódico, etc. La vía más adecuada es esta última, en la que se centraron tanto él como otros autores en diferentes trabajos, que examinaremos en las siguientes líneas.

Para medir qué número de palabras, o bien qué porcentaje de cobertura textual, necesitamos para leer un texto o comprender adecuadamente una conversación, debemos intentar definir qué se entiende por comprensión razonable, adecuada u óptima. Esto varía en función de las personas, pues no todas conseguimos entender del mismo modo un texto con idéntico porcentaje de palabras conocidas, lo que supone una de las primeras dificultades a esquivar. Para ello, los estudios trabajaron con muestras de variados individuos, sobre todo en el ámbito de una L2 y del inglés. Los porcentajes de mayor éxito en las pruebas de comprensión son los que nos indican cuál es la puntuación a partir de la cual consideramos que los individuos pueden llevar a cabo una lectura de forma “adecuada”, pero siempre son datos relativos, no totalmente fijos. Una de las primeras en hacer su aportación a este campo fue Laufer (1989), que trabajó con estudiantes de inglés con fines académicos. Ella relacionó la cobertura léxica con la comprensión lectora y pidió a su grupo de muestra dos tareas: subrayar las palabras desconocidas de un texto y llevar a cabo una prueba de traducción que incluía las palabras de menor frecuencia del texto. Luego comparó los resultados de las dos pruebas para ver las posibles discrepancias. La cobertura léxica era el total de palabras menos el número de palabras desconocidas en la práctica, convertido en porcentaje. En el test de comprensión lectora se fijó en que el grado de entendimiento “adecuado” se daba a partir de una puntuación de 55%. Además, los resultados mostraron que con una cobertura textual del 95% era donde había significativamente más participantes con una puntuación por encima de 55 en el test. Con todo, como ya hemos dicho, las puntuaciones no son las mismas en todos los participantes aunque trabajen con textos con un mismo porcentaje de cobertura textual. Pero si existe un mayor número de casos en un determinado nivel, parece el punto más plausible para establecerlo como el límite de la comprensión lectora “adecuada”.

Posteriormente, Hu y Nation (2000) dedicaron también un estudio a estas mismas cuestiones. Ellos crearon 4 grupos de estudiantes para trabajar con textos, en los que suprimieron palabras para sustituirlas por otras sin sentido. El resto de las palabras que dejaron en los textos pertenecían a las 2.000 más frecuentes. Al igual que en el anterior estudio de Laufer, ellos usaron dos tests, uno de elección múltiple y otro de escritura, para medir a partir de qué puntuación y en qué nivel de cobertura textual se producía una comprensión adecuada. No obstante, en la segunda prueba los participantes también tenían que demostrar su capacidad de escritura, por lo que las puntuaciones no son exactamente equivalentes a las de otras investigaciones. En el caso de Hu y Nation se obtuvo como resultado que nadie podía leer adecuadamente con una cobertura textual del 80%, algunos participantes podían hacerlo con el 90% y el 95%, aunque la minoría. Solo con el 98% de cobertura léxica se producían una comprensión adecuada en la mayor parte de los estudiantes. Por consiguiente, comparando los resultados de ambos estudios, Nation (2001: 147) sugirió que:

The probabilistic threshold is 98%. With this coverage almost all learners have a chance of gaining adequate comprehension. If, instead of adequate comprehension, a standard of minimally acceptable comprehension is applied (as Laufer did in her study), then 95% coverage is likely to be the probabilistic threshold. [Nation 2001: 147]

Laufer realizó una nueva aproximación para encontrar cuál es el número de palabras con el que alcanzamos el umbral de comprensión léxica adecuado. Al igual que en los resultados de estudios anteriores, se pudo ver como la noción de umbral léxico depende de lo que consideremos comprensión “razonable” o “adecuada”, ya que se obtuvieron diferentes puntuaciones en los tests de comprensión lectora. Los resultados muestran que con 3.000 familias de palabras las puntuaciones de un grupo fueron de 56%, el nivel más bajo de vocabulario a partir del cual había más lectores que entendían el texto que los que no. Con un vocabulario de 4.000 el nivel subía hasta los 63% y con 5.000 ascendía al 70%. Por lo tanto, volvemos a oscilar entre tres escalones diferentes.

Estos umbrales de comprensión coinciden con los que Nation obtuvo en un estudio posterior (2006) y se vieron refrendados por Laufer y Ravenhorst-Kalovski (2010). No obstante, los autores difieren en lo que se consideran el

umbral léxico de comprensión ideal pues, mientras Nation se decanta por el 98% de cobertura textual, es decir, un vocabulario de entre 8.000 y 9.000 familias de palabras para la lectura y alrededor de 6.000 y 7.000 para la comprensión auditiva, Laufer asume esta cota como la superior y propone el 95% como un segundo nivel de comprensión, si no ideal, sí aceptable. En lo que sí coinciden es en fijar estos dos niveles, explicando que el 95% es en el que la comprensión empieza a ser significativa y el 98% en el que trabajamos con una cota de comprensión más segura para la mayor parte de los estudiantes de una LE.

Nation explica, además, que el 98% es el nivel de cobertura textual en el que los aprendices no se sirven del apoyo de diccionarios u otros recursos externos al texto en sí, por lo que también este motivo parece apuntar a este porcentaje como el adecuado. Con un porcentaje del 95%, entonces, nos encontraremos 1 palabra desconocida en cada 2 líneas, lo que es lo mismo que 1 palabra desconocida por cada 20 palabras o 7 en cada minuto de discurso oral, a una media de 150 palabras por minuto (Nation 2006: 78-9). En relación con la cobertura textual y con la densidad de palabras (entendidas como *tokens*, es decir, ocurrencias) desconocidas de un texto, Hu y Nation (2000) expusieron la relación entre ambas cuestiones del siguiente modo:

% de cobertura textual	Densidad de palabras conocidas y desconocidas	Número de líneas del texto por cada palabra desconocida
99%	1 palabra desconocida cada 100	1 cada 10 líneas
98%	1 desconocida cada 50 palabras	1 cada 5 líneas
97%	1 desconocida cada 33 palabras	1 cada 3.3 líneas
96%	1 desconocida cada 25 palabras	1 cada 2.5 líneas
95%	1 desconocida cada 20 palabras	1 cada 2 líneas
90%	1 desconocida cada 10 palabras	1 cada 1 línea
80%	1 desconocida cada 5 palabras	1 cada 0.5 línea

TABLA 13. Relación entre el porcentaje de cobertura textual, la densidad de palabras desconocidas y el número de líneas del texto por cada palabra desconocida.

4.4.1. Vocabulario visual

En cuanto a la comprensión lectora y dentro del ámbito del vocabulario receptivo, podemos ser más específicos y delimitar lo que se ha acuñado con el término de “vocabulario visual” o “vocabulario receptivo de vista” (en inglés se suele denominar *sight vocabulary*). Este se refiere a las palabras cuyo significado nos es tan familiar que podemos entenderlo fuera de contexto y, si lo encontramos en el interior en un texto, entonces lo reconocemos y decodificamos rápidamente y sin ningún esfuerzo cognitivo (Laufer & Ravenhorst-Kalovski 2010: 16). Schmitt y McCarthy definen los elementos de este vocabulario como las palabras que, en la lectura, “are known well enough to be quickly, automatically and accurately recognised” (Schmitt & McCarthy 1997: 331).

Este tipo de vocabulario contribuye a una lectura fluida y libre de esfuerzo cognitivo para interpretar textos, con lo que podemos volcar nuestro esfuerzo en la comprensión del contenido del texto concreto y sus implicaciones, obteniendo así un nivel de lectura mayor (Mezynski 1983, Pulido 2007 y Segalowitz 2007). En la misma línea, Laufer (1997a y 1997b) apuesta por que el aprendiz dirija sus esfuerzos hacia el desarrollo de este tipo de vocabulario, pues libera sus recursos cognitivos para invertirlos en darle sentido al vocabulario desconocido o poco familiar y para interpretar el sentido global del texto. Así pues, la cobertura textual y el nivel de nuestro vocabulario visual son factores que están directamente relacionados con nuestro umbral léxico de comprensión. Según Pulido (2009), un amplio vocabulario de vista (*sight vocabulary*) redundará en mayores ganancias léxicas a través del proceso de inferencia contextual y en mayores índices de retención de estas ganancias. Pulido concluye que:

Still, the rich tend to get richer, that is, L2 passage sight vocabulary and L2 reading ability play a significant role in lexical input processing, such that the learners with great text coverage and better reading skills also experience greater levels of vocabulary development during reading. [Pulido 2009: 78-79]

4.4.2. Inferencias léxicas

Algunos investigadores han puesto de manifiesto que la habilidad para inferir o adivinar el significado de palabras desconocidas en un texto puede depender del número de palabras frecuentes que conozca la persona que está en proceso de aprendizaje de una lengua:

These words make up the majority of tokens in any discourse, so if they are not known, language users will be unable to make accurate guesses about the meanings of the remaining less frequent words, many of which are likely to be unknown. [Schmitt 2000: 73]

Laufer (1997a y 1997b) explica que uno de los obstáculos para llevar a cabo una lectura comprensiva en los aprendices de una LE es la incapacidad para inferir palabras desconocidas, ya que la deducción del significado del vocabulario desconocido a través de las inferencias léxicas contextuales es casi imposible si no se conocen la mayoría de las palabras que componen el texto. Los factores que intervienen en este hecho son: la existencia de claves contextuales; la familiaridad con estas claves, ya que aunque estas aparezcan puede que el estudiante no las entienda y por lo tanto no haga una inferencia exitosa; la aparición de claves parciales o engañosas, pues a veces el estudiante puede confundirse; y, por último, la incompatibilidad entre el contexto y los esquemas cognitivos de la persona que lee el texto, pues se puede dar el caso de que los elementos no coincidan y se interprete el texto erróneamente (Cartaya Febres 2011: 108).

En las investigaciones sobre el nivel umbral se apunta también que este sería, según los autores, el punto a partir del cual una persona es capaz de inferir las palabras desconocidas, proceso que se conoce como inferencia léxica contextual, o, lo que es lo mismo, la capacidad de deducir palabras por el contexto conocido. Se ha establecido la existencia de este umbral o conocimiento mínimo de vocabulario que posibilita la comprensión de un texto auténtico. Así, estos porcentajes de cobertura léxica del 95% y del 98% que se han establecido como la horquilla de comprensión aceptable y óptima son los niveles a partir de los cuales estos procesos se consiguen llevar a cabo. Hemos visto que el número de palabras varía, en función de si se trata de la lengua oral o escrita, entre las

3.000 familias de palabras más frecuentes y las 8.000 o 9.000, a partir de las cuales el alumno podrá aplicar las estrategias de inferencia. Con todo, Laufer (1997a y 1997b) afirma que si lo que se pretende es que el aprendiz de una L2 efectúe inferencias contextuales precisas, debe conocer como mínimo el 98% de las palabras que componen el texto. Lo cual implica que el aprendiz debe poseer un vocabulario receptivo de al menos 5.000 familias de palabras, según Laufer y Ravenhorst-Kalovsky (2010), y de 6.000-7.000 (comprensión oral) o bien 8.000-9.000 (comprensión lectora), según Nation (2006).

En relación con el concepto presentado en el apartado anterior, Pulido y Hambrick (2008) afirman que si no se cuenta con una cantidad adecuada de “vocabulario de vista” relacionado con el fragmento que se esté leyendo, se puede experimentar una especie de “cortocircuito”, por el cual el aprendiz no acierta ni en establecer relaciones entre las ideas del texto, ni en el control y comprensión de la idea global, por lo que pierde la capacidad de inferir e integrar nuevos significados. Se pierde, por tanto, la oportunidad de desarrollar el léxico mediante esa lectura.

4.5. Palabras de baja frecuencia, cobertura léxica y lectura efectiva de textos

Hemos hablado hasta ahora de la relación entre el conocimiento léxico, la cobertura léxica textual y la comprensión eficaz, y nos hemos estado refiriendo casi siempre a las cifras de cobertura léxica que se obtienen con las palabras de mayor frecuencia, tanto de la lengua general como las de ámbito académico o técnico. No obstante, la relación entre cobertura, vocabulario y comprensión implica también que incluso un ligero incremento en la cobertura léxica (entre un 1% o un 2%) puede ser tan beneficioso para la lectura como lo sería un aumento mayor. Es decir, un pequeño aumento en la cobertura léxica se asocia con una mejora sustancial en los resultados de los tests de lectura y comprensión, dando como resultado mejores puntuaciones. Laufer y Ravenhorst-Kalovsky (2010) han aportado las siguientes cifras sobre la cuestión:

However, the relationship between coverage, vocabulary and reading implies that even a small increase in lexical coverage (1.19% from 4K to 5K) may be just as beneficial to reading as a larger increase in coverage (2.25% from 3K to 4K). Interestingly, a small improvement in coverage (0.8%) from 5K to 6K, or 1.3% from 5K to 7K was associated with the best improvement in the reading score (17%). [Laufer & Ravenhorst-Kalovsky 2010: 24-25]

La explicación a estos resultados puede ser que algunas palabras de baja frecuencia pueden actuar como palabras clave y por lo tanto ser cruciales para la comprensión, tanto en textos académicos y técnicos —para lo que necesitaremos un vocabulario seleccionado de tales ámbitos—, como en la lengua general, en principio no especializada —para lo que necesitamos las palabras con mayor carga semántica y asociadas directamente con el contexto comunicativo, como puede ser el vocabulario disponible—.

Laufer y Nation (2001), al analizar la relación entre la cantidad de vocabulario y la rapidez en decodificar el significado de las palabras, encontraron que el conocimiento adicional de algunas unidades léxicas infrecuentes contribuye a la mejora de la comprensión lectora, aunque no suponga un incremento en la cobertura textual. Esto debe ser tenido en cuenta, por lo tanto, para la selección del vocabulario para la enseñanza. Algunos corpus explican que el vocabulario infrecuente no compensa el esfuerzo de su aprendizaje, porque ofrece poca cobertura textual. Y aun así, los resultados muestran que, a pesar del poco incremento de la cobertura, es sustancial en el caso de la comprensión lectora, porque el beneficio de enseñar algún vocabulario infrecuente pero bien escogido por su utilidad puede ser mucho mayor de lo que se pensaba inicialmente. Y es en este camino en el que los métodos de selección léxica como la disponibilidad léxica se manifiestan como un componente imprescindible.

4.5.1. Número de palabras frecuentes para el nivel inicial

Como ya vimos en la tabla de cobertura textual (Tabla 8, página 155), según los listados de frecuencias, al margen de las diferencias en la elección de las muestras elegidas para el recuento, parece existir una convergencia

sustancial destacable —de aproximadamente un 80%— entre las unidades léxicas consideradas más frecuentes en diversos listados de frecuencias léxicas en una misma lengua (McCarthy 1990: 80; Hwang & Nation 1995: 36; Picoche 1993: 58, 1999: 432; Corda & Marelló 1999: 165; Nation 2001: 15; Izquierdo Gil 2004: 339). Este dato evidencia dos cuestiones: una es la estabilidad y relevancia de ciertas unidades léxicas y otra que nos podemos fiar de la frecuencia léxica como un criterio de selección del vocabulario de primera importancia, tanto para el desarrollo de la comprensión como el de la producción.

Se suele afirmar que el límite entre las palabras frecuentes y las de más baja rentabilidad se sitúa a partir de las 2.000 unidades. Para Nation (1990) el 87% de las palabras de un texto están contenidas dentro de las 2.000 más frecuentes y aparecen en la lengua corriente utilizada en la vida diaria en los textos que se leen o se escuchan normalmente. Así pues, quien cuente con estas 2.000 palabras más frecuentes en su vocabulario está en una situación que le posibilitaría entender el 87% de un texto real, no adaptado ni diseñado con fines pedagógicos. Del mismo modo, y como ya hemos comentado, si a estas dos mil palabras más frecuentes se suman las que se contienen en la *University Word List* (800 *headwords*), entonces se alcanza un 95% de comprensión de un texto académico.

Carter afirma también que el conocimiento de esas primeras dos mil palabras más frecuentes facilitará la comprensión de cualquier texto auténtico y al mismo tiempo estimulará la motivación del alumnado, ya que constituirá un aprendizaje rentable. Además, los aprendices las encontrarán constantemente en los diversos textos a los que se enfrenten a lo largo de su proceso de aprendizaje de la LE:

It is claimed that knowing these words gives access to about 80 per cent of the words in any written text and thus stimulates motivation since the words acquired can be seen by learners to have a demonstrably quick return. [Carter 1996: 163]

Ya hemos comentado que debido a la estrecha relación existente entre el vocabulario frecuente y los porcentajes de cobertura textual, parece indispensable efectuar en el nivel inicial una cuidadosa elección de las unidades léxicas más frecuentes a partir de diversos recuentos de frecuencias léxicas del

español, con el fin de que la persona que aprende español pueda alcanzar progresivamente un nivel de comprensión adecuado al leer cualquier texto:

If you carefully choose what words to learn, you can get a good coverage of the words on a page. For example, if you choose the most frequent 1000 words of English, which includes words like *put*, *end*, *difficult*, *come* and *material*, you would cover about 75% of the words on a page (Nation 1993: 119).

Pues bien, otros autores (Nation 2001: 12; Baker, Hill & Devylder 2000) indican que existe un pequeño núcleo de palabras de alta frecuencia, recurrentes en cualquier listado de frecuencias léxicas, independientemente de las muestras utilizadas para el recuento, los temas que predominen en ellas, el género, cualquiera que sea la edad, sexo o nivel sociocultural de los sujetos entrevistados. Ese núcleo de vocabulario compuesto por unidades léxicas fundamentalmente gramaticales debería, pues, ser objeto de estudio en el nivel elemental o inicial (Izquierdo Gil 2004).

Nation (2001) alega además que:

To reach 95% coverage of academic text, a vocabulary size of around 4,000 word families would be needed, consisting of 2,000 high-frequency general service words, about 570 general academic words (the Academic Word List) and 1,000 or more technical words, proper nouns and low-frequency words (Nation 2001: 147).

Esas 2.000 familias de palabras de la lengua general de alta frecuencia a las que se refiere son las que componen la *General Service List of English Words* (West 1953, Baumann & Culligan 1995) y las 570 familias de palabras que pertenecen al vocabulario académico general o semiespecializado se refieren a la *Academic Word List* de Coxhead (2000). Entonces, la utilidad de la frecuencia para la selección léxica se podría establecer en estas 2.000 palabras, que alcanzan una cobertura textual del 80%, y a partir de ahí el objetivo sería “enriquecer nuestro vocabulario” con otros métodos, sobre todo para buscar aquel tipo de unidades léxicas que no nos ofrecen los listados de frecuencias y que se revelan altamente productivas, por resultar palabras de alto contenido léxico que operan como palabras clave. En los textos y conversaciones de nivel coloquial, estas podrían ser las palabras que las nociones específicas y los centros de interés nos ofrecen, mientras que en los textos e intercambios

comunicativos más académicos o técnicos, deberemos buscarlas en listados específicos de este tipo. Lo que parece cierto es que, si a esas 2.000 palabras más usuales para un nivel inicial o 5.000 para un nivel elevado le sumamos un número relativamente bajo de palabras con alta carga semántica, conseguiremos obtener una cobertura léxica que nos asegure una comprensión, si no óptima, sí aceptable. Sin duda, unas candidatas perfectas para constituir este conjunto complementario son las palabras disponibles.

4.5.2. Número de palabras, cobertura léxica y umbral de lectura para los diferentes niveles del *Marco común europeo de referencia para las lenguas (MCERL)*

Tomando como referencia los trabajos revisados en los apartados de este capítulo, hemos elaborado una tabla (Tabla 13) con las cifras que se pueden obtener sobre la cobertura léxica textual de las palabras más frecuentes. Si a estas cantidades de palabras les sumamos las disponibles, que actúan como palabras clave, los porcentajes de cobertura ciertamente ascenderán, con ello también aumentará la capacidad de comprensión lectora de las personas que estudian español como segunda lengua o como lengua extranjera. En la Tabla mostramos también las correspondencias con los niveles del *MCERL*.

Nº de palabras frecuentes	Cobertura léxica textual	Nivel de comprensión (según la hipótesis del nivel umbral)	Nivel del <i>MCERL</i>
500 palabras	Entre 68% y 79%	Nivel elemental. Por debajo del nivel umbral.	A1
1.000 palabras	Entre 78% y 81%	Nivel elemental. Por debajo del nivel umbral.	A1+
2.000 palabras	Entre 86% y 90%	Nivel elemental. Por debajo del nivel umbral.	A2
3.000 - 5.000 palabras	Entre 89% y 95%	Nivel de comprensión aceptable - Umbral mínimo.	B1
5.000 - 6.000 palabras	Entre 92% y 98%	Nivel de comprensión óptimo. Usuario independiente, sería el aceptable para cursar estudios superiores teniendo el español como lengua vehicular.	B2
6.000 - 9.000 palabras	Entre 94% y 99%	Nivel de comprensión óptimo. Usuario independiente, sería el aceptable para cursar estudios superiores teniendo el español como lengua vehicular.	C1

TABLA 14. Relación entre el número de palabras frecuentes, su cobertura léxica textual, el nivel de comprensión que se alcanza y los niveles del *MCERL*

Para tener datos más precisos sobre el léxico fundamental, y no solo sobre el frecuente, habría que hacer un estudio semejante a los que hemos analizado a lo largo del capítulo, cuyo objetivo sería medir la eficacia receptiva en español de los hablantes, tanto de una lengua materna como de una lengua extranjera. Un estudio como el que sugerimos serviría para avalar nuestra hipótesis de que el nivel de comprensión mejoraría sustancialmente. Con esto definiríamos de forma más certera a partir de qué cantidad de palabras y porcentaje de cobertura textual podríamos establecer un nivel umbral óptimo.

Estas cifras marcan los límites de la cobertura léxica y el número de unidades que definen el vocabulario fundamental del español, con el que un hablante puede hacer frente a conversaciones auténticas cotidianas de forma eficaz y a la lectura de textos no académicos ni especializados. Así mismo, y con la intención de aplicar este concepto al contexto de enseñanza-aprendizaje de ELE, sería útil establecer escalas de cantidades y cobertura, como los expuestos en la tabla anterior, en relación con los diferentes niveles del *MCERL*. Para probar qué datos se obtienen empíricamente debemos trabajar con la tipología de textos, intercambios comunicativos (las funciones comunicativas) y el vocabulario (nociones específicas) que son definidos por el *PCIC* para cada nivel. Además, sería conveniente seleccionar los grupos de participantes clasificados también por niveles. Creemos que el abordaje de estudios de este tipo para el español nos ayudaría a avalar y perfeccionar con criterios más científicos los datos y los parámetros que hemos expuesto. No obstante, sí contamos con suficiente información en esta línea a partir de estudios hechos sobre la lengua inglesa y sobre las palabras frecuentes, por lo que trabajamos a partir de sus conclusiones. Para ello, tomamos como punto de partida las cantidades demarcadas sobre la hipótesis del umbral léxico en la comprensión de una L2 (Cartaya Febres 2011), en las que parece que existe una cierta coherencia entre las diferentes investigaciones. Recordemos estos marcos aproximados:

<i>Oralidad:</i>		
3.000 palabras – cobertura léxica de un 95%	--	nivel de comprensión aceptable
5.000 palabras – cobertura léxica de un 98%	--	nivel de comprensión óptimo
<i>Escritura:</i>		
5.000 palabras – cobertura léxica de un 95%	--	nivel de comprensión aceptable
8.000 – 9.000 palabras – cobertura léxica de un 98%	--	nivel de comprensión óptimo

CUADRO 7. Relación entre los grados de comprensión de la hipótesis del nivel umbral y el número de palabras necesarias para lograrlo, distinguiendo oralidad y escritura

En inglés existe una obra lexicográfica destinada a estudiantes (jóvenes y adultos) con 8.000 entradas. Se trata del diccionario *COBUILD Essential English Dictionary* de Harper Collins. Se presenta como un libro que recoge las palabras y colocaciones que los aprendices de nivel A1 hasta B1 deben conocer. Las definiciones usan un lenguaje también sencillo, con ejemplos reales e ilustraciones. Este volumen sería un ejemplo de un diccionario referente para estudiantes de nivel elemental e intermedio básico de inglés como lengua extranjera. Como se puede observar, el número de entradas coincide en general con el límite superior de comprensión que hemos establecido en la tabla anterior para la escritura.

Capítulo 5. Los vocabularios básicos, el léxico frecuente y el léxico usual

En los apartados anteriores hemos venido ofreciendo información general sobre la frecuencia y la disponibilidad léxicas, al tratar las cuestiones relacionadas con el léxico y con los diferentes métodos de selección del vocabulario. En este capítulo y en los siguientes, retomaremos cada una de estas líneas de trabajo por separado para ampliar y condensar la información más relevante sobre ellas. De este modo, en las siguientes páginas nos centraremos en las diferentes fuentes que nos informan sobre la frecuencia de las palabras para estudiar las publicaciones que hay en este campo desde sus inicios hasta la actualidad. Intentamos comparar esta producción en español con incursiones en la que existe en otras lenguas, sobre todo en inglés, pues es la lengua pionera en este tipo de trabajos. Así, veremos cómo se ha desarrollado esta línea en diferentes ámbitos lingüísticos y el provecho que de ella se ha hecho en la enseñanza de lenguas extranjeras.

Describiremos más pormenorizadamente las obras de este tipo que existen en español, ya que nuestro objetivo es estudiar los diferentes recuentos y léxicos básicos que nos pueden ser útiles para seleccionar aquellas palabras frecuentes que necesitamos para la enseñanza-aprendizaje de ELE en los primeros niveles. Luego, intentaremos escoger aquellas que se ajusten a nuestros intereses para estudiarlas y compararlas de modo más preciso, y así poder centrarnos en una o dos de ellas. Presentaremos a continuación una lista de palabras frecuentes para trabajar en nuestra parte práctica sobre el vocabulario fundamental.

5.1. Los corpus de frecuencias

Recordemos que un corpus es una gran colección de textos almacenados en un ordenador. Los corpus son de gran tamaño para ser buenos representantes de la lengua en su globalidad y, desde hace algunos años, son utilizados como base para la elaboración de diccionarios, gramáticas y libros de texto. De hecho, en la elaboración de diccionarios resultan de gran utilidad debido a que, al ser analizados con programas de concordancias, ofrecen datos hasta entonces no disponibles, que se pueden usar para miles de comprobaciones y comparaciones, permitiendo un estudio más completo y detallado, así como una descripción exacta del lenguaje con un grado autoridad que era inimaginable con las viejas técnicas en papel o incluso de pequeños corpus. Las concordancias extraídas de un corpus amplio revelan patrones de uso con mucha mayor definición y representatividad. En los últimos años, los corpus han desterrado las viejas modas a la hora de crear diccionarios y se están aliando con la informática e internet para ofrecer nuevas herramientas lexicográficas que poco a poco van dejando atrás a los diccionarios clásicos. A esta cuestión se refiere la distinción que los estudios lexicográficos establecen entre *poner el corpus en el diccionario* (PCED) y *poner el diccionario en el corpus* (PDEC) (Kilgarriff 2009). La primera línea representa la lexicografía basada en corpus, mientras que la segunda es la defendida por quienes codifican la información en el corpus, tal y como se hace en algunos trabajos de desambiguación semántica automática o en cualquier trabajo de anotación de corpus que necesite información léxica (Alonso Ramos 2009).

En relación a la enseñanza de una L2/LE, la lingüística del corpus se erige como la base de aquellos movimientos que defienden la importancia del vocabulario en el aprendizaje de lenguas, así que es necesario idear métodos para seleccionarlo. De este modo, el vocabulario frecuente, usual o básico, ha destacado como fuente de palabras útiles desde los inicios. Recordemos por ejemplo el Movimiento de Control del Vocabulario, que comenzó de forma rigurosa durante la década de los años 20 y 30 y manifestaba su interés por los contenidos del vocabulario y su tratamiento sistemático en la enseñanza. La idea básica era concentrar los esfuerzos de aprendizaje en una selección de las

palabras que se usaban más y componían un vocabulario mínimo de palabras que deberían servir para satisfacer necesidades básicas de comunicación. Más recientemente, la denominada aproximación léxica ha vuelto a poner el énfasis en esta clase de palabras, por lo que ha tenido lugar una nueva defensa de su centralidad en el proceso de enseñanza-aprendizaje.

5.1.1. Corpus lingüísticos, léxico frecuente, léxico usual y léxico básico

La lingüística del corpus nos ofrece, ya lo hemos comentado, aportaciones de gran importancia para la investigación sobre las lenguas de modo general, y, como hemos dicho, en concreto en el ámbito lexicográfico permite elaborar recursos hasta ahora nunca imaginados a nivel cuantitativo y sobre todo cualitativo. Gracias a ella, ya desde inicios del siglo XX conseguimos extraer listados con las palabras más frecuentes de un conjunto de textos, que normalmente se agrupan en diferentes áreas.

Los corpus se puede usar de este modo para elaboración de diccionarios siguiendo una metodología léxico-estadística, que deja de lado las posibles parcialidades y subjetividades del quehacer lexicográfico a la antigua usanza. Pero con ellos conseguimos también ser más selectivos a la hora de elaborar obras léxicas. Así, los diccionarios normativos son más globales e intentan recoger el mayor caudal léxico posible (algunos incluso considerando arcaísmos, regionalismos, tecnicismos y cultismos), por lo que pierden su utilidad para la enseñanza de una lengua. En cambio, las listas de frecuencia muestran los vocablos que sí se utilizan y recogen el vocabulario que tiene vida en una determinada comunidad de habla, en función de su número de apariciones. Este se ordena según rangos, es decir, estableciendo una jerarquía de acuerdo con la cantidad de veces que los vocablos se repiten en un texto. La frecuencia, entre otras cosas, pone al descubierto el manejo real que los hablantes hacen del caudal léxico de su propia lengua.

Sin embargo, existen diferencias entre los listados de frecuencias, determinadas por los métodos de elaboración, así como por el tipo y la extensión de los materiales utilizados para formar el corpus a partir del cual se elabora el listado correspondiente. En segundo lugar, influyen también los criterios de

selección de las palabras. Por ejemplo, si al elaborar el corpus se ha incluido un conjunto elevado de textos periodísticos de determinada época, algunas palabras que aparecen en las noticias recurrentemente en esos años pueden acabar en las más frecuentes, cuando, en realidad, su uso responde más a una cuestión concreta o de moda. Coxhead menciona algunos ejemplos, como “yemení” o “lithuanian”, que serían palabras de este tipo presentes en el diccionario Collins COBUILD (1995), pues al relacionarse con los conflictos políticos de los años 90 aparecen reflejadas en el corpus del *Bank of English*.

En conclusión, la frecuencia de las palabras en un texto es un primer criterio para la elaboración de un listado, pero no debe ser el único, ya que la selección de los textos a partir de los que se realiza y, sobre todo, los criterios que se apliquen producirán resultados divergentes. Por ejemplo, el concepto de “palabra” que cada autor emplea le lleva a incluir y excluir determinados elementos de su listado, con lo que se obtienen resultados dispares. En ocasiones no se ajustan a las necesidades específicas de determinado grupo de alumnos o a un nivel concreto, por lo que, ya lo hemos apuntado anteriormente, es necesario tener este tipo de cuestiones en mente si pretendemos aplicar los recuentos a la docencia.

En cuanto a los criterios concretos que se pueden seguir para elaborar los listados, Nation y Waring (1997) consideran la representatividad, la frecuencia y la cobertura, las palabras base y las familias de palabras, las frases idiomáticas y las expresiones fijas y la variedad de la información. Así, consideramos cuestiones como que el corpus sea representativo de los variados usos de la lengua (por ejemplo, oral y escrita), o que las palabras escogidas aparezcan en una amplia gama de textos. También es necesario tener en cuenta a quién va destinado el listado y para qué usos, contemplar las frases y expresiones, así como información sobre formas y categorías gramaticales que hay en cada familia, los diversos significados, las colocaciones y las restricciones en su uso, bien debidas al registro, bien debidas a la distribución geográfica. Por otra parte, para la elaboración del *Academic Word List (AWL)* su autora tuvo en cuenta la frecuencia, la aparición especializada (las palabras que se incluyeron en la lista no debían estar entre las 2.000 palabras más frecuentes de la lengua según la

General Service List) y el rango —teniendo en cuenta la distribución, la palabra debía aparecer al menos 10 veces en cada una de las cuatro secciones principales del corpus académico (Letras, Comercio, Derecho y Ciencias), y en 15 o más de las 28 áreas especializadas—.

Los criterios y los métodos han ido cambiando y perfeccionándose a lo largo del tiempo. Desde los primeros recuentos de palabras de finales del siglo XIX hasta los diccionarios de frecuencias de mediados del siglo XX los procedimientos se han ido afinando. Aunque solemos usarlas como expresiones sinónimas, las denominaciones de “recuento”, “diccionario de frecuencias”, “léxico usual” y “léxico básico” pueden esconder también diferencias significativas que responden a la época y los métodos que se han aplicado para la elaboración de los listados. En este sentido, si tratamos con mayor profundidad esta cuestión se pueden distinguir matices entre todas estas denominaciones, observando su evolución cronológica a lo largo del siglo XX. Así, los primeros trabajos en esta línea se pueden considerar recuentos, mientras que en el otro polo están los léxicos básicos, pasando por los diccionarios de frecuencias léxicas, que se podrían considerar como las obras fronterizas entre ambos. No cabe duda de que los primeros recuentos nacieron condicionados por lo que los medios técnicos permitían en esos momentos, y que poco a poco la depuración de las técnicas que se aplicaban evolucionaron parejas a los desarrollos informáticos. El primer recuento léxico, de la autoría de Kaeding (1898) para la lengua alemana, fue elaborado de forma manual en su recopilación, análisis y cómputo. Con el nuevo siglo, surgieron otras propuestas de este tipo, como las de Keniston (1920), Thorndike (1921), Buchanan (1927), etc. No obstante, además de la falta de madurez que todavía presentaban en relación a la elección de los textos, la división de los dominios y las fórmulas matemáticas para analizar la frecuencia de los vocablos, estos primeros trabajos nos ofrecen listados de gran tamaño, con lo que pueden resultar poco funcionales en cuanto a su grado de selección. Pensemos, por ejemplo, que el *The teacher's Word Book* de Thorndike recoge 10.000 palabras, el *A Graded Spanish Word Book* de Buchanan considera una muestra total de 1.200.000 con una lista final de 18.331 vocablos y la primera lista elaborada en español por Rodríguez Bou (1952) es de 10.000

palabras frecuentes a partir de un vocabulario de 83.430. Dos de estos trabajos surgieron con una aplicación pedagógica, por lo que su tamaño no parece demasiado adecuado a su objetivo.

Thorndike publica en solitario su *The Teacher's Word Book* en 1921 y diez años después una versión ampliada, *The Teacher's Word Book of 20,000 Words* (1932). Esta incluye la frecuencia con la que cada palabra aparece dentro de un corpus que asciende a un millón de palabras. Informa también sobre el número de tipos de textos distintos. A partir de un corpus escrito de 18 millones de palabras extrae 30.000 lemas, o 13.000 familias de palabras.

En 1924 Henmon publica *A French Word Book Based on a Count of 400.000 Running Words*, sobre el que ofrece una muestra de las 5.000 palabras más frecuentes del francés, para lo que trabaja con la frecuencia y el rango. En su *A Graded Spanish Word Book*, Buchanan en 1927 discrimina siete áreas temáticas en su corpus en función de la procedencia de los textos: obras dramáticas, novelas, poesía, folclore, prosa miscelánea, textos técnicos y textos periodísticos. Para la posterior elaboración de su listado usa una fórmula en la que combina la frecuencia y el rango, aunque el problema al que se enfrentó fue que el rango solo resultaba útil en las palabras con baja frecuencia, mientras que perdía su eficacia en las de alta frecuencia. Vander Beke en 1935 sigue en su *French Word Book* en líneas generales los procedimientos de Henmon. En él da una mayor importancia al rango que a la frecuencia, pues valora más que una palabra aparezca una sola vez en diferentes ámbitos que varias en uno solo.

Un punto de referencia para la docencia de lenguas extranjeras y para la creación de manuales fueron las obras de Thorndike en coautoría con Lorge (1944) y la *General Service List of English Words (GSL)* de West (1953). Este último fue el primer listado de frecuencias de importancia y prestigio en inglés, además de ser ampliamente usado. Contiene 2.000 palabras base (*headword*) ordenadas alfabéticamente, con una breve definición y ejemplos. Para cada palabra se ofrece el número de veces estimado (representado como = e) que la palabra está contenida en un corpus de 5 millones de palabras. Se distingue el porcentaje de frecuencia que tiene cada sentido de una palabra y las formas

derivadas al lado de la palabra base, generalmente con su frecuencia. La *GSL* ha tenido una gran influencia durante muchos años y ha servido como base para crear diferentes tipos de materiales relacionados con la enseñanza del inglés. No obstante, es un listado elaborado a partir de textos publicados entre 1938 y 1949, por lo que ya no es representativo de la actualidad.

Como hemos mencionado, el primero de este tipo de estudios que se publica en el ámbito hispano es el *Recuento de vocabulario español* de Rodríguez Bou (1952). Un año más tarde García Hoz publica su *Vocabulario usual, común y fundamental* para determinar cuál es el léxico básico del español con idea de destinarlo, sobre todo, a la enseñanza de una L1. Este mismo autor publicaría también en 1976 un *Vocabulario general de orientación científica y sus estratos*, que se centra en el vocabulario científico básico para la enseñanza media. A lo largo de estos años y a través de esos trabajos fue depurándose la metodología: se ajustaron los criterios de selección de textos, se perfeccionaron las fórmulas estadísticas empleadas y, en general, fue avanzándose en la elaboración de corpus léxicos más representativos y diversificados. A pesar de estas iniciativas pioneras y de la pasión que levantaron las listas de frecuencias en los años treinta y posteriormente en los sesenta (auspiciadas por la metodología audio-oral en la enseñanza de lenguas extranjeras) la aparición de los ordenadores dio el impulso definitivo a los trabajos basados en corpus, pues no solo permitía el manejo de una ingente cantidad de datos con rapidez, sino también conseguir resultados más fiables.

Esta línea de investigación alcanzó tal madurez y estabilidad durante esos años que en 1959 Guiraud se detiene en estudiarla en su obra *Problèmes et méthodes de la statistique linguistique*, considerada como “la primera gran obra de síntesis de la estadística lingüística” (García Marcos 1985: 23). En su aportación, Guiraud ya perfila una distinción entre “palabra” y “vocablo”, que hoy podemos encontrar en los estudios sobre corpus lingüísticos y en los de disponibilidad léxica. Así, contrapone las “palabras totales” de un texto y las “palabras diferentes” dentro del conjunto, e indica que se deben contabilizar por separado (Guiraud 1959: 9). En general, en su propuesta realiza un repaso de la bibliografía sobre estadística lingüística hasta el momento. En este marco, resalta

la importancia de la frecuencia para la descripción lingüística pero, sobre todo, para la enseñanza de la lengua, a la que se asoció ya desde sus inicios.

Un antes y un después lo marca la publicación en 1964 de *Frequency Dictionary of Spanish Words* de Juilland y Chang-Rodríguez, por ser uno de los trabajos de mayor envergadura, mejor perfilados metodológicamente hasta el momento y que mayor influencia tuvo en trabajos coetáneos y posteriores. En él observamos una lista de 5.024 palabras seleccionadas a partir de un conjunto léxico integrado por 500.000 unidades, repartidas en 5 mundos textuales con 100.000 palabras cada uno. Se centraron en el vocabulario adulto, estándar, de la Península Ibérica de obras escritas entre 1920 y 1940. Aunque lo que resulta de mayor interés es su aparato estadístico, en el que emplean una fórmula que combina la frecuencia con la dispersión, lo que supuso una innovación, pues nadie la había usado hasta el momento. Con este índice de dispersión intentan controlar la estabilidad de la frecuencia a través de los distintos mundos de los que esta procedía. Al multiplicar la frecuencia por la dispersión obtenemos el “índice de uso” de cada palabra, que es en lo que se basan sus listados. Este índice de uso varía entre 0 y 1 y va creciendo a medida que la dispersión no se reparte equitativamente a través de todos los mundos considerados. Aunque se pueda criticar, como hizo Morales en una obra posterior (1986), tanto este diccionario de frecuencias como los que Juilland publicó en colaboración con otros autores en francés, italiano y rumano, llenaron un vacío existente hasta el momento en esta línea de investigación, y todavía hoy se reconocen sus logros en cuanto a la metodología.

Estas obras, que fueron llamadas “diccionarios de frecuencia” por sus autores, se pueden interpretar como la frontera entre los anteriores recuentos de frecuencia y los que posteriormente consideramos repertorios de léxico básico. Una de sus características es que su forma de separar los mundos de los que procede el vocabulario tiene repercusiones significativas en la selección del contenido y en los resultados finales de los corpus que analizan (Mateo García 1998: 35-6). El que podemos considerar el primer léxico básico del español, elaborado por Morales en 1986, toma estas obras como base de forma explícita. Así, la evolución en las técnicas que se usaban en la léxico-estadística dejaba

claro que era necesario recurrir al uso del vocabulario, y no solo a la frecuencia. Debido a esto, en la siguiente fase, los léxicos básicos intentaron perfeccionar los anteriores recuentos en este sentido, ya que fueron conscientes de que el aumento de la frecuencia no es sinónimo por sí solo de una mayor utilidad de las palabras. Al contrario, en los léxicos básicos se pondera la frecuencia y la dispersión, con lo que no se unifica todo tipo de textos a partir de los cuales se extraen las frecuencias, sino que se divide todo el universo léxico en diferentes mundos (*domains* o dominios en español) delimitados por su contenido o por sus condiciones formales. Por otra parte, los léxicos básicos no recogen una gran cantidad de palabras, ya que normalmente no van más allá de las 5.000 o las 8.000. En muchas ocasiones incluso menos. Esto se debe a que el léxico básico pretende recoger aquel núcleo de palabras más importantes y repetidas en un idioma, fruto de un proceso metódico de selección. En español no contamos con muchas publicaciones de este tipo. Entre las primeras obras se encuentra el *Léxico Básico de Puerto Rico* de Morales. Una década después, los proyectos *Léxicos básicos de España (LEBAES)* y *de Canarias (LEBAICan)*, con Benítez Pérez, Hernández Cabrera y Samper Padilla al frente, sentaron sus bases teóricas y sus principios metodológicos, aunque se trata de recuentos inconclusos. Sus directores partían de la necesidad imperiosa de este tipo de estudios, pues “Salvo el citado diccionario de Juilland, que ha quedado sumamente anticuado en sus resultados, nada se ha hecho en el país que se asemeje a un léxico básico” (Benítez Pérez, Hernández Cabrera y Samper Padilla 1995).

Si bien no contamos con el resultado, el proyecto puede resultar útil simplemente por el hecho de resumir las características de lo que debe ser un trabajo de este tipo: en cuanto a la selección de la muestra, (a) carácter sincrónico, centrándose en un periodo de 25 años, que empieza en 1970 y termina en 1994, ambos inclusive; (b) carácter escrito, pues las muestras orales son difíciles de clasificar; (c) carácter homogéneo, solo se trabaja con la lengua general; (d) carácter exhaustivo, pues se incluyen todos los títulos de un género determinado que han visto la luz en los años indicados; (e) la clasificación del conjunto de obras se divide en los cinco mundos que ya había considerado Juilland: drama, narrativa, ensayo, literatura técnica y periodismo. Para cada uno

hay un listado final de 100.000 palabras, con lo que tenemos cinco corpus distintos y una muestra total de 500.000 palabras (como hicieron Juilland y Morales). Sobre las características metodológicas, se utiliza la misma fórmula de dispersión de Morales (1986), elaborada a su vez sobre la de Juilland (1973), aunque con refinamientos:

$$D = 1 - \frac{\sqrt{n x_i^2 - T^2}}{2T}$$

donde $T^2 = (X_i)^2$ es la suma elevada al cuadrado de la frecuencia total de cada palabra y $n x^2$ es el producto del número de mundos o categorías por la suma de los cuadrados reales de las frecuencias de la palabra en cada mundo. Se pondera la frecuencia con la dispersión para alcanzar el índice de uso mediante la fórmula $\{U = F \times D\}$.

También toman determinadas decisiones en relación a la segmentación de la unidad de análisis, con presupuestos muy próximos a Morales (1986). De este modo, se establece partir de la palabra como unidad de segmentación, lematizar el corpus, distinguir las polisémicas y las homógrafas, asignar palabras a clases, desambiguar las homonimias sintácticas mediante la distinción de los homógrafos que representan clases gramaticales diferentes y dividir y agrupar cadenas gráficas (Benítez Pérez, Hernández Cabrera & Samper Padilla 1995).

En la elaboración y uso de corpus y listados de frecuencias, los ingleses nos llevan varios cuerpos de distancia en cuanto a la enseñanza de LE y muchos más si nos referimos al caso concreto del léxico. En esta lengua disponen de diversos manuales y diccionarios, aparte de una metodología bastante precisa que varía levemente en función de las diferentes escuelas lingüísticas. En este ámbito, por ejemplo, uno de los que se presume de los mejores en la actualidad es el *British National Corpus* (BNC), con 10 millones de palabras de muestras de lengua hablada. El *Cambridge Internacional Corpus* afirma contener más de un billón de términos y varios subcorpus, incluyendo el *Cambridge and Nottingham Corpus of Spoken Discourse in English* (CANCODE), de 5 millones.

Si reducimos los corpus actuales del inglés a un esquema, se pueden representar de la siguiente manera (López Mezquita 2005):

1. *General Service List*, versión de John Bauman (GSL(B))
2. *Brown Corpus* (BC)
3. *Cambridge International Corpus* (CIC)
4. *British National Corpus* (BNC)
 - a) versión de Adam Kilgarriff (BNC (K))
 - b) versión de Geoffrey Leech (BNC (L))
5. *Bank of English* (BOE)
6. *Cambridge and Nottingham Corpus of Discourse in English* (CANCODE).

Sus características básicas son las siguientes (López Mezquita 2005):

	GSL (B)	BC	CIC	BNC	BOE	CANCODE
tamaño del corpus	5 millones de palabras	1 millón de palabras	600 millones de palabras	100 millones de palabras	450 millones de palabras	5 millones de palabras
corpus oral /escrito	escrito	escrito	escrito y oral	escrito y oral	escrito y oral	oral
lista lematizada	sí	no	no	sí	sí	no

TABLA 15. Algunos corpus textuales del inglés y sus características

E incluso podemos comparar cuáles son las primeras 50 palabras de mayor frecuencia en cada uno de ellos:

Nº	GSL	BC	CIC	BNC(K)	BNC (L)	BOE	CANCO DE
1	the	the	the	the	the	the	the
2	be	of	of	be	be	be	I
3	of	and	to	of	of	of	you
4	and	to	and	and	and	and	and
5	a	a	a	a	a	a	to
6	to	in	in	in	in	in	it
7	in	that	that	to	to	to	a
8	he	is	is	have	have	have	yeah
9	have	was	for	it	I	to	that
10	it	he	it	to	it	for	of
11	that	for	was	for	he	I	in
12	for	it	he	I	for	on	was
13	they	with	on	that	not	with	is
14	I	as	with	you	that	he	it's
15	with	his	I	he	you	that	know
16	as	on	as	on	on	at	no
17	not	be	at	with	they	as	oh
18	on	at	be	do	do	by	so
19	she	by	by	at	by	but	but
20	at	I	his	by	she	it	on
21	by	this	but	not	at	do	they
22	this	had	have	this	his	it	well
23	we	not	from	but	we	from	what
24	you	are	are	from	that	his	yes
25	do	but	said	they	or	say	have
26	but	from	not	his	an	they	we
27	from	or	they	that	will	not	he
28	or	have	you	she	say	you	do
29	which	an	this	or	as	we	got
30	one	they	an	which	would	an	that's
31	would	which	had	as	can	or	for
32	all	on	has	we	there	will	this
33	will	you	or	an	all	their	just
34	there	were	one	say	if	't	all
35	say	her	which	will	get	which	there
36	who	all	will	would	her	year	like
37	make	she	were	can	make	who	one
38	when	there	their	if	go	this	be
39	can	would	who	their	who	she	right
40	more	their	we	go	one	up	not
41	if	we	would	what	see	when	don't
42	no	him	all	there	so	if	she
43	man	been	she	all	know	would	think
44	out	has	her	get	time	her	if
45	other	when	more	her	some	go	with
46	so	who	been	make	take	make	then
47	what	will	about	who	up	there	at
48	time	more	there	as	as	all	about
49	up	no	when	out	could	about	are
50	go	if	its	up	year	can	as

TABLA 16. Las primeras 50 palabras más frecuentes en algunos corpus del inglés

Hay algunos corpus modernos que están disponibles, como el *Bank of English Corpus* (BOE), el *Cambridge International Corpus* (CIC) y el *British National Corpus* (BNC), por lo que se han convertido en fuentes perfectas con las que elaborar listados de frecuencias actualizados y de alta fiabilidad. El BOE proporcionó las frecuencias que fueron usadas para elaborar el diccionario *Collins COBUILD*, mientras que los listados elaborados a partir de los datos extraídos del CIC están, por el momento, sólo a disposición de los autores que trabajan en materiales pertenecientes a la editorial *Cambridge University Press*.

También existen otros recursos elaborados a partir de los corpus de frecuencias. Por ejemplo, utilizando una variedad de fuentes como las anteriores, recordemos que Nation y algunos otros investigadores recopilaron el vocabulario más frecuente con el objetivo de elaborar varias listas de 1.000, 2.000, 3.000, 5.000 y 10.000 familias de palabra en inglés (Nation 2001). Nos hemos referido a ellas al hablar de la cobertura léxica y el número de unidades necesarias para una comprensión eficaz, pues este autor toca ambos temas (*vid. supra* § 2.2, 4.3. y 4.5.1.). Se las suele conocer como las “primeras 1.000”, las “primeras 2.000” y así sucesivamente. Coxhead (2000) también ha producido una Lista de Palabras Académicas (*Academic Word List - AWL*) de 570 familias de palabras que son especialmente frecuentes en textos académicos, pero no se encuentran entre las 2.000 más frecuentes. Este listado actualiza otro anterior llamado Lista de Palabras Universitarias (*University Word List - UWL*).

Nation también ha diseñado un software para producir “perfiles léxicos” de textos. Este recurso sirve para evaluar la riqueza léxica de los textos, o el porcentaje de palabras de las primeras 1.000, 2.000 y de los niveles del *AWL*. También es capaz de evaluar la dificultad de un texto e identificar palabras infrecuentes, e incluso puede ser usado para evaluar el uso del vocabulario en los textos escritos de los estudiantes. Usando este software, Littlewood & Liu (1996) descubrieron el 89% de las palabras usadas por una muestra de estudiantes universitarios estaban en la lista de las primeras 2.000 familias. Otros estudios mostraron que en estudiantes que eran hablantes nativos el porcentaje era de 83% y para los escolares era de 65% (Bauer & Nation 1993). Con todo, estos

trabajos se basaron en muestras pequeñas, lo cual explicaría la variación en los resultados.

Otra muestra del desarrollo de este tipo de estudios en lengua inglesa son los diferentes recursos en línea de los que disponemos, desde los corpus citados, hasta el software de Nation, así como otros relacionados con la frecuencia y los recuentos de palabras³⁸.

5.1.2. Rango, dispersión e índice de uso

Como hemos comentado en el anterior recorrido por la bibliografía de los trabajos de frecuencias léxicas, los vocabularios de frecuencias evolucionaron en una segunda fase al incorporar las nociones de rango y dispersión a su metodología, lo cual supuso un gran salto cualitativo³⁹. Ya Buchanan prefirió priorizar las palabras que aparecían en los varios tipos de muestras del corpus,

³⁸ Incluimos a continuación una lista con algunos de ellos:

British National Corpus: Con 100.000.000 formas, actualmente es el mayor corpus accesible y público. Se puede encontrar en él las frecuencias de las palabras y ver 50 ejemplos. También se vende a instituciones y particulares. <http://www.natcorp.ox.ac.uk/>

Cambridge Advanced Learner's Dictionary: probablemente el mejor diccionario en línea para aprendices de inglés como L2. Los aprendices más jóvenes también pueden escoger el *Cambridge Learner's Dictionary* en la misma página web. Ambos están hechos a partir del Cambridge International Corpus. <http://dictionary.cambridge.org/>

Cambridge International Corpus: contiene más de 1 billón de formas divididas en subcorpus. Usado en diccionarios, manuales y publicaciones académicas de Cambridge University Press. No es de uso público general.

Cambridge and Nottingham Corpus of Spoken Discourse in English (CANCODE): Es un corpus de inglés hablado usado para trabajos innovadores sobre el vocabulario y la gramática de esta variedad de lengua. <http://www.cambridge.org/elt/corpus/cancode.htm>

HKU English Centre: Vocabulary analysis tools: web basada en versiones de software de análisis de textos. Puede ser usada para producir listas de frecuencia y perfiles de vocabulario cortando y pegando textos. <http://ecdev.hku.hk/vocabulary/tools.htm>

Lancaster University: University Centre for Computer Corpus Research on Language: esta página tiene una interesante variedad de recursos, incluidas listas de vocabulario basadas en búsquedas en el BNC. <http://www.comp.lancs.ac.uk/ucrel/>

Range: versión completa del software para el perfil del Rango del vocabulario de P. Nation. En la página web de este investigador en la Victoria University of Wellington, Nueva Zelanda. Identifica las palabras entre las primeras 1.000, 2.000 y en la AWL. También incluye estas listas de palabras. Hay una versión de Range en HKU English Centre: Vocabulary analysis tools. <http://www.vuw.ac.nz/lals/staff/paul-nation/nation.aspx>

The Compleat Lexical Tutor: mantenido por Tom Cobb en la Universidad de Quebec, incluye herramientas para la contar el tamaño del vocabulario. <http://www.lex Tutor.ca/>

³⁹ Diferentes autores han tratado los conceptos de *rango* o *dispersión* como un criterio que necesariamente se debe unir al de la frecuencia para la selección del léxico. Entre ellos, McCarthy 1990: 69; Carter 1996: 44; Gairns & Redman 1996: 57-63; Mackey & Savard 1967).

aunque se repitieran poco, frente a las que se repiten muchas veces pero en un solo dominio. Al incluir rango y dispersión se observa que dos palabras pueden presentar una misma frecuencia, pero diferente distribución en los tipos discursivos o géneros textuales que se hayan distinguido (Izquierdo Gil 2004: 330). Si un término se repite en un solo ámbito textual, aunque sea un elevado número de veces, pero no aparece en otros, debemos deducir que es característico de ese género. En este caso la posibilidad de que aparezca en otro dominio es baja, por lo que su frecuencia absoluta no se corresponde con el peso y el valor real que tiene en la lengua en general. De hecho, si la distribución es regular en los diversos géneros podemos afirmar que la palabra es independiente de esas variables, por lo que su peso y utilidad es mayor (Ávila Muñoz 1999: 74).

Así pues, para corregir la medida de frecuencia en las investigaciones léxico-estadísticas, se comenzó a aplicar el rango para medir la dispersión de los vocablos en las diferentes fuentes. Los conceptos de rango y de dispersión —a la que otros autores llamaron *distribución*, *repartición* o *diseminación*— se usaron para ponderar la frecuencia y centrarse en la estabilidad de los vocablos, de tal forma que la dispersión revela el valor del término en general, determinando su presencia en todos los subconjuntos de textos de la muestra total. De esta forma, medir la distribución aporta una información valiosa en cuanto a la utilidad de las unidades léxicas en la lengua, con lo que obtenemos un vocabulario de uso general, más lógico y más exacto. Otra ventaja de tener en cuenta la dispersión es que se aumenta la representatividad de la muestra, en la medida en que conseguimos acercarla al idioma en su totalidad, y no parcialmente.

En la práctica, para ponderar la frecuencia se establecieron dos índices, el *índice de dispersión* y el *índice de uso*, ambos con una oscilación que va de 0 (nivel inferior) a 1 (nivel superior). La forma en que funcionan estos índices y sus resultados están relacionados, ya que la dispersión es uno de los factores considerados en las diversas fórmulas que se han usado para hallar el índice de uso. Si una unidad léxica tiene idéntica frecuencia en todos los ámbitos discursivos, entonces su índice de dispersión es 1, mientras que será 0 si aparece localizada solo en uno. Con el índice de uso confirmamos que, efectivamente,

estos vocablos son frecuentes en cualquier contexto de uso de la lengua. Por consiguiente, este índice discrimina aquellas palabras con una frecuencia muy elevada, pero con desigual distribución. Si se restringen a un mundo discursivo, independientemente de su frecuencia de aparición, su índice de uso es bajo. Las unidades más generales y estables suelen identificarse en las primeras posiciones con las *palabras gramaticales*, independientemente de los cambios de tema, de canal, etc. Las unidades de mayor índice de uso, a pesar de tener frecuencias más bajas, desplazan en muchas ocasiones a las de mayor frecuencia pero que solo están focalizadas en un mundo o tipo discursivo, debido a que aquellas cuentan con un mayor índice de dispersión. Por ejemplo, *necesitar* y *grabar* pueden tener frecuencia similar (222/203) pero la primera está presente en diferentes textos y la segunda concentra más su aparición, por lo que el índice de dispersión es de 0,70 y de 0,33 respectivamente. A su vez, estos índices de dispersión influyen en el índice de uso, de tal modo que el de *necesitar* será mayor (Izquierdo Gil 2004: 332).

No solo los trabajos de léxico-estadística que contemplaron la frecuencia, el rango y la dispersión en su horizonte fueron los basados en corpus, pues los estudios de léxico disponible también recurrieron a estos parámetros en su procedimientos. El modo en que esto se llevó a cabo ha ido mejorando a lo largo de los años con las diferentes fórmulas matemáticas propuestas, al igual que ocurrió en los diccionarios de frecuencias y léxicos básicos, pero esa es una cuestión sobre la que trataremos en el Capítulo 6.

5.1.3. Recuentos léxicos y lematización

Otro aspecto que distingue a los recuentos léxicos es si están sin lematizar, por ejemplo los de García Hoz (1953), Justicia (1995) y Molero (1999), o lematizados, como los de Juilland y Chang-Rodríguez (1964), Morales (1986) y Ávila Muñoz (1999). Como ya hemos visto, cuando un listado léxico está lematizado, las diferentes formas con una misma base léxica se agrupan bajo una sola entrada, lo cual ocurre por ejemplo en las formas conjugadas de un verbo, como por ejemplo *doy, daré, dábamos, dando, disteis, dar...* para unir las a su

infinitivo, *dar*, que es lo que encontramos como entrada. En los sustantivos y adjetivos, las flexiones de género, número, las formas apocopadas o los superlativos, también confluyen en el masculino singular. Debido a esto tendremos que buscar *casas* en la entrada *casa*, por ejemplo. Por el contrario, si los recuentos no están lematizados se mantienen entradas diferentes para las diversas flexiones de las palabras variables, como sustantivos y verbos. Por otra parte, en los recuentos no lematizados podemos obtener exclusivamente las frecuencias brutas de cada forma léxica, mientras que los recuentos lematizados operan por agrupaciones, ofreciendo la frecuencia relativa de cada forma léxica y la frecuencia absoluta, que consiste en la suma de todas las frecuencias relativas (Sinclair 1991: 173).

Estos dos diferentes tratamientos de la frecuencia léxica pueden tener repercusiones de calado en la presentación de la lista final. Por una parte, el uso del lema como unidad de recuento puede reducir el número de vocablos totales en un corpus (Nation, 2001: 8). Por otra, el orden de las formas léxicas también puede variar considerablemente de un recuento no lematizado a otro lematizado. Nation añade, además, la posibilidad de ordenar el listado en familias de palabras, lo cual constituye otra forma de presentar los datos. Con esta agrupación reunimos no sólo las formas conjugadas o las variaciones de género y número, sino también todas las formas derivadas de una misma unidad léxica.

5.2. Los listados de frecuencias más destacados

En las páginas anteriores hemos ido adelantando características generales sobre los listados de frecuencias léxicas y también hemos ido aportando los títulos más conocidos (*vid. supra* § 3.4.1.). En este epígrafe los estudiaremos más pormenorizadamente por orden cronológico de fecha de publicación.

La primera referencia con que contamos de un recuento de palabras según criterios de frecuencia data de 1898⁴⁰, cuando Kaeding publicó un listado

⁴⁰ Kaeding, F.W. (1898). *Haufigkeitwörterbuch der deutschen Sprache*, Berlín.

de frecuencias basado en un corpus de 11.000.000 de palabras, cuya recopilación, análisis y cómputo fue manual, para contribuir a la formación de taquígrafos, y que se analizó bajo la supervisión del Bureau Prusiano de Estadística. El diccionario de frecuencias de Kaeding sirvió como base para la elaboración del primer vocabulario básico alemán: el de Morgan, de 1928. Además, Estoup usó los datos de Kaeding para realizar algunos cálculos estadísticos sobre la frecuencia de palabras y formas en textos, entre 1905 y 1907 (con una cuarta edición en 1916, de mayor importancia).

- Keniston, H. (1920). Common words in Spanish. *Hispania* 3, 85-118.

En 1920 Keniston intentó establecer un inventario de lengua cotidiana y para ello creó una lista de 1.322 palabras divididas en ocho secciones y ordenadas por frecuencias, aunque este trabajo tiene en cuenta tanto el criterio de la frecuencia como el de la repartición. La repartición se podría identificar con la dispersión para otros autores, aunque en estas fechas todavía se aborda de forma rudimentaria. Según señala Keniston, las 185 palabras que aparecen en la primera lista se encuentran al menos en el 80% de los textos estudiados, mientras que las 221 de la lista octava, aparecen en el 33% de esos textos. Incluyó obras de teatro, periódicos, revistas, así como cuentos y novelas, pero no poesía. Es la primera lista de vocabulario especialmente destinada a la enseñanza de una lengua extranjera (Bogaards, 1994: 104). Frente a otros autores de la época, no incluyó en su corpus ni cartas ni pasajes bíblicos, como solía ser frecuente.

- Thorndike, E. L. (1921). *The Teacher's Word Book*. New York: Columbia University, Teachers College.

Con finalidad pedagógica se crearon también para el inglés listas de frecuencias para uso de los docentes. Thorndike creó una lista de diez mil palabras en 1921 y una versión ampliada en 1932. En colaboración con Irving Lorge publicó también *The Teacher's Word Book of 30.000 Words* (1944), que alcanzó fama y uso para la selección del vocabulario de libros de texto y lecturas

graduadas. Incluye la frecuencia con la que cada palabra aparece dentro de un corpus de un millón de palabras, y el número de tipos de textos distintos. Contiene 30.000 lemas, o 13.000 familias de palabras, extraídas a partir de un corpus escrito de 18 millones de palabras (Goulden, Nation y Read 1990). Thorndike empleó diversas fuentes, desde la Biblia u obras dramáticas de Shakespeare, hasta la correspondencia privada, incluso libros de texto, aunque no explica los motivos de escoger estas fuentes. Este vocabulario no se puede considerar de frecuencia ya que, aunque ofrece este dato de las palabras, dado el número de vocablos que registra, sobrepasa lo que se entiende por vocabulario usual, y en realidad incluye mucho que no lo es (García Hoz 1953: 17-8).

- Buchanan, Milton A. (1927). *A Graded Spanish Word Book*. Toronto: American and Canadian Committees on Modern Languages.

Se trata de un recuento basado en el lenguaje escrito (obras de teatro en su mayor parte, periódicos, novelas, revistas, narraciones), que gozó de cierto prestigio en los años siguientes. El autor seleccionó 30.000 palabras en cada una de las 40 categorías establecidas. La muestra total es de 1.200.000 palabras que dieron como resultado un vocabulario de 18.331 vocablos. En su corpus parte del siglo XVII e incluye textos hispanoamericanos.

- Eaton, H. (1940). *An English-French-German-Spanish Word Frequency Dictionary*. New York: Dover Publications, Inc.

Según se explica en la contraportada de la primera edición, es un listado multilingüe de 6.000 palabras frecuentes que se publicó gracias al apoyo del Comité de Lenguas Modernas del Consejo de América de Educación:

This indispensable language study aid, which was prepared under the auspices of the Committee on Modern Languages of the American Council on Education, is a semantic frequency list of the 6,000 most frequently used words for each of four languages: English, French, German and Spanish. It differs from ordinary work frequency lists in being based upon concepts rather than upon words alone, so that you will not be misled by sub meanings, synonyms or variant constructions [...]. [Eaton 1940: texto de la contraportada]

- Rodríguez Bou, I. (1952). *Recuento de vocabulario español, I*. Río Piedras, Puerto Rico: Universidad de Puerto Rico.

En la década de los 50 se publican varios trabajos, tanto en el ámbito del español como en el del francés, siempre con un interés pedagógico. La lista realizada incluía palabras del lenguaje oral y escrito de niños y adultos tomadas de composiciones y conversaciones transcritas de niños, periódicos, revistas, programas de radio, literatura religiosa y textos escolares. La muestra total es de 7.066.637 palabras, con un vocabulario global 83.430 y una lista de 10.000 palabras frecuentes. En este recuento se distingue el “vocabulario de expresión y de reconocimiento”, pero no se considera el criterio de dispersión. Parte del trabajo de Buchanan, por lo que recoge palabras que aparecían en su trabajo, así como textos dialectales y procedentes de Hispanoamérica. Esta obra está compuesta de varios volúmenes y ofrece diferentes listas, por lo que su consulta se vuelve muy compleja. En su elaboración trabajaron durante 8 años filólogos, estadígrafos y maestros, con lo que consiguieron aumentar sus posibilidades de subclasificación y desglose y ofrece una lista esencial de 105 palabras, que, según los autores, cuentan con la más alta frecuencia de uso y representan el 50% del vocabulario común latinoamericano. En el segundo volumen recoge las palabras de los 14 mundos o apartados en que divide los textos seleccionados. Al lado de las palabras más frecuentes se incluye un número que indica el rango de acuerdo con su frecuencia, para lo que emplea una clave numérica que se define al inicio del volumen. La obra consta de 2 volúmenes de más de 500 páginas cada uno, y en el primero ofrece la lista de las 10.000 palabras más frecuentes en orden de rango y otra en orden alfabético. Esto puede dar una idea de su complejidad y envergadura.

- García Hoz, V. (1953). *Vocabulario usual, común y fundamental: determinación y análisis de sus factores*. Madrid: C.S.I.C.

Recoge un vocabulario aproximado de 15.000 palabras, que era la cantidad que se suponía debía dominar una persona adulta. El corpus es de unas 400.000 unidades divididas en cuatro aspectos: a) vida familiar —cartas privadas—, b) vida social indiferenciada —periódicos—, c) vida social regulada —documentos oficiales del estado y eclesiásticos, así como folletos de

organizaciones sindicales—, y d) vida cultural —textos de los libros más vendidos—. Da cabida a textos dialectales, léxico religioso e infantil. A partir de dicho corpus, García Hoz estableció tres estratos o niveles léxicos: vocabulario usual —12.428 voces con una frecuencia media de 100—, vocabulario común —de 1.971 voces, presentes en los cuatro bloques seleccionados y con frecuencia no menor a 40— y fundamental —de 208 voces de frecuencia no menor a 400—. De estos tres, el “vocabulario común”, recoge las palabras que se hallan en todos los tipos de vocabulario y son las más importantes, ya que se caracterizan por ser las más generales y tener presencia en todos los tipos de vocabulario. En cuanto a la aplicación de este vocabulario común en la lectura, ya señalaba su autor en los años cincuenta que estaba en manos de “los autores de libros escolares, quienes, en vez de usar palabras cultas, que en muchas ocasiones reflejan una retórica rebuscada y empalagosa, debieran emplear con preferencia las palabras de uso corriente” (García Hoz 1953: 385). Para él, pues, el vocabulario común (extraído del vocabulario usual del adulto) posee una clara aplicación pedagógica en los primeros años de enseñanza del español, ya que está constituido por palabras que se hallan en todos los tipos de vocabulario, sea cual sea la situación en que se encuentre un individuo.

El “vocabulario fundamental” es el constituido por las palabras que tienen sus frecuencias prácticamente distribuidas en partes iguales en uno y otro tipo de campo. Así llega a determinar las 208 palabras que lo constituyen. Fuera de él quedan las que, aun siendo comunes, son más propias de un tipo de mundo que de otros, pues tienen frecuencias desiguales.

Además de los tres capítulos de vocabulario usual, común y fundamental donde define lo que entiende por cada uno de ellos y ofrece las listas de palabras, García Hoz hace un verdadero ejercicio de compilación y explicación metodológica para fundamentar su publicación. Empieza tratando la cuestión de lo que entiende por vocabulario usual, del que destaca su valor pedagógico y psicológico, frente a las aportaciones de otros autores, como Thorndike que le parece que sigue un método demasiado inclusivo y masivo. Él se cuestiona los criterios que debe seguir y llega a la conclusión de que debe documentar las palabras usadas por los adultos, que son la expresión lingüística de la sociedad. A continuación explica pormenorizadamente los textos que incluye y que excluye

(el teatro costumbrista, por ejemplo) en su afán de registrar el lenguaje coloquial escrito. Comenta en detalle las publicaciones, obras y fechas que selecciona.

En otro apartado trata la cuestión del número de palabras que existen en el vocabulario de los diferentes individuos como referente para escoger el número de palabras del usual. Comenta las que seleccionaron trabajos anteriores a los suyos y, lo más importante, argumenta la organización de su corpus en los grupos de 100.000 palabras para cada tipo de fuente, con los cálculos que lo llevaron a tomar sus decisiones finales.

Es interesante el capítulo que denomina “Técnicas de registro”, en el que define lo que entiende por “palabra” y los problemas técnicos para su contabilización. Llega a la conclusión, siguiendo a Casares (1950), de que no hay criterios objetivos y científicos para segmentar palabras debido a su diversidad, por lo que sigue el pensamiento de García de Diego y se rige por lo que la Real Academia Española registra en su *Diccionario Vulgar* como palabras, sin distinguir las diferentes acepciones para contabilizarlas. Emplea complejas fórmulas de análisis factorial y métodos estadísticos para medir las correlaciones entre las palabras del vocabulario usual y común en los cuatro mundos.

Finalmente, incluye un capítulo sobre el análisis factorial y desarrolla casi todo un capítulo para explicar la Técnica de Spearman y la Técnica de Thurstone sobre correlaciones con complejas fórmulas matemáticas, gráficas y tablas (para lo cual se usan páginas desplegadas), que le sirven para explicar los factores que intervienen, en su opinión, en los distintos tipos de vocabulario, tanto en los específicos como en los comunes.

- West, M. (1953). *General Service List*. London: Longman, Green and Co.

Se publica en el mismo año el primer listado de frecuencias de auténtica importancia y reconocido prestigio en lengua inglesa, que ha sido ampliamente utilizado. Es un corpus de 5 millones de palabras que ofrece 2.000 palabras base (*headword*) más frecuentes ordenadas alfabéticamente, con una breve definición y ejemplos de oraciones en las que aparecen. Al lado de cada una se muestra una cifra que representa el número de veces estimado de apariciones, y también el porcentaje de la frecuencia que tiene cada significado de la palabra. La formas

derivadas también se ofrecen bajo la palabra base en negrita minúscula, acompañadas generalmente de su frecuencia. Las primeras mil palabras de la lista son las que se registran 332 o más veces en 5 millones de palabras, más los meses, los días de la semana, los números y los títulos *Mr.*, *Mrs.* y *Miss*.

Este listado ha tenido una gran influencia durante muchos años y ha servido como base para crear muchos tipos de material relacionado con la enseñanza del inglés. Las frecuencias que ofrece han sido usadas para considerar la entidad de las palabras que era más importante saber. Sin embargo, hoy en día la antigüedad de los datos, recogidos entre 1938 y 1949 hace que no sea representativo del inglés actual.

- Gougenheim, G.; Michéa, R.; Rivenc, P. y Sauvageot, A. (1956). *L'élaboration du français élémentaire: étude sur l'établissement d'un vocabulaire et d'une grammaire de base*. Paris: Didier. Con una nueva edición revisada en 1964. *L'élaboration du français fondamental: étude sur l'établissement d'un vocabulaire et d'une grammaire de base*. Paris: Didier.
- Gougenheim, G.; Michéa, R.; Rivenc, P. y Sauvageot, A. (1958). *Dictionnaire fondamental de la langue française*. Paris: Didier. (Contiene 3.500 palabras, nueva edición revisada en 1977).

En esta misma década, del lado francés se partía de la frecuencia, o más bien de las deficiencias que esta presentaba, para sentar las bases de lo que más tarde se retomó como Léxico Disponible. Para distinguirlo del frecuente o básico, los autores acuñaron los listados de palabras resultantes como vocabularios elementales o fundamentales. La finalidad de obtener las palabras más frecuentes y más disponibles era la docencia del francés como LE, ya que Francia en esos momentos necesitaba manuales de francés elemental para facilitar el aprendizaje de esta lengua a los inmigrantes y a los habitantes de sus excolonias. Describiremos pormenorizadamente estas obras y su metodología en el capítulo dedicado a la Disponibilidad léxica.

- Müller, Ch. (1964). *Essais de statistique lexicale*. Paris: Bibliothèque française et romaine.⁴¹

En este caso la estadística léxica no tenía fines didácticos, sino que tenía como objetivo el estudio estilístico de autores literarios. La metodología se basaba en el recuento de las unidades léxicas empleadas por determinado autor o en una obra concreta.

- Juilland, A. & Chang-Rodríguez, E. (1964). *Frequency Dictionary of Spanish words*. La Haya: Mouton.

Quizás sea este el diccionario de frecuencias más citado del ámbito hispano. Utilizó textos escritos del español peninsular pertenecientes al drama, narrativa, ensayo, literatura técnica y científica y periodismo de los años comprendidos entre 1920 y 1940. Está basado en un corpus de 500.000 ocurrencias y obtiene un vocabulario básico de 5.000 palabras. Estos autores utilizaron 100.000 ocurrencias en cada uno de los géneros citados, que son los 5 mundos en los que se divide en corpus, por lo que además de indicar la frecuencia absoluta de una palabra, se daban indicaciones de la frecuencia relativa en cada género, y se podían establecer comparaciones de género a género respecto de la importancia de una palabra. Estos 5 mundos o dominios son:

- a) Obras dramáticas
- b) Obras de ficción, novelas e historias breves
- c) Obras en prosa, ensayos, cartas y memorias
- d) Textos periodísticos
- e) Textos técnicos

Excluye el léxico latinoamericano, dialectal y la poesía. Según Rojo (2001), se trata de una obra excelente, cuyos puntos positivos son la lematización, la indicación de la clase gramatical y la resolución de homografías, pero al haber sido publicada hace ya más de cincuenta años, resulta hoy en día obsoleta. Uno de los aspectos que diferencian el vocabulario básico de Juilland y Chang-

⁴¹ Con ediciones posteriores también en otras lenguas: *Essais de Statistique Lexicale* (1964), *Étude de Statistique Lexicale* (1967), *Initiation à la Statistique Linguistique* (1968), *Initiation aux Méthodes de la Statistique Linguistique* (1973), *Principes et Méthodes de Statistique Lexicale* (1977), *Estadística lingüística*. Madrid, Gredos (1968-73).

Rodríguez es el cambio de perspectiva, pues se pasa de tener en cuenta solo la frecuencia simple a considerar también la dispersión o repartición. Esto hace que se desarrolle otro modo de seleccionar y ordenar las palabras, basado en el uso. Esta nueva elaboración técnica se irá perfeccionando, como ya hemos comentado, pero se caracteriza por crear fórmulas matemáticas, más simples o más complejas, para ofrecer datos más fiables. Así se incluye el *rango*, como la dispersión simple o repartición, y el *índice de dispersión compleja* o, más comúnmente, el *índice de uso*. El primero se define como el total de categorías o subconjuntos en que aparece una palabra. Esta fórmula evalúa la repetición de palabras a partir de un conjunto homogéneo. Su ventaja es que en estos casos no se ve afectada por las desviaciones estilísticas de cada categoría, y la desventaja es que necesita un número elevado de subconjuntos para resultar acertada. En cuanto al índice de uso, se introduce para dar cabida a la dispersión, y se perfila con el tiempo a nivel teórico hasta el índice de dispersión compleja o de uso, que se apoya en las desviaciones estilísticas de las categorías. Este índice favorece a las palabras más generales y abstractas, a la vez que, en relación con las categorías gramaticales, disminuyen los nombres en favor de los verbos y los adjetivos. El índice de dispersión permite que las palabras con alta frecuencia pero ligadas a un mundo o ámbito en concreto puedan ser ponderadas, así, las palabras de mayor uso desplazan a las de mayor frecuencia.

Se publicaron también recuentos del mismo tipo en otras lenguas románicas para lo que pretendía ser una colección de léxicos básicos⁴², y su influencia se dejó notar profundamente en otros listados posteriores, como la obra de Morales (1986). A pesar de que hoy en día el corpus se ha quedado obsoleto, sus virtudes en cuanto a proceder metodológico en el tratamiento de los datos y las técnicas que se desarrollaron para medir la frecuencia, la dispersión y el uso, hacen de esta una obra pionera y de gran calidad, que se refleja en el peso que ha tenido en las posteriores. Sus autores fueron perfeccionando las fórmulas para medir el uso desde los primeros léxicos básicos de la colección hasta el último, el del italiano de 1973.

⁴² *Frequency Dictionary of Spanish Words* 1964; *Frequency dictionary of Rumanian words* 1964, *Frequency dictionary of French Words*, 1971, *Frequency dictionary of Italian Words*, 1973.

- Márquez Villegas, L. (1975). *Vocabulario del español hablado*. (Niveles y distribución gramatical). Madrid: Sociedad General Española de Librería.

En su investigación, Márquez Villegas se basa en estimaciones subjetivas de frecuencias y del método absoluto, pues se pide a los participantes que hagan estimaciones de frecuencia a partir de una escala del tipo “empleado frecuentemente/raramente empleado/nunca empleado” o bien “empleado una vez al mes, por semana...”. En la época supuso un estudio innovador, aunque no por su metodología, sino porque se basó en el vocabulario de la lengua hablada, mientras que los diccionarios de frecuencias privilegiaban las fuentes escritas. Las críticas que ha recibido inciden en que no ofrece explicaciones sobre cómo lleva a cabo la elaboración del recuento ni su metodología. Así, por ejemplo, no define lo que entiende por lengua familiar, ni el tipo de palabras que la forman; tampoco menciona el diccionario escolar utilizado en el estudio; los tres niveles están insuficientemente caracterizados e incluso se ha observado que algunas palabras del primero no corresponderían en realidad a un nivel elemental, mientras que hay otras del segundo y tercero que sí deberían estar en el anterior (Izquierdo Gil 2004: 340). Lo que sí se explica es que se basa en las valoraciones subjetivas de 200 alumnos de bachillerato de Madrid. Este grupo tenía que comentar la frecuencia de 20.000 unidades léxicas de un diccionario escolar según 3 categorías: 1. palabras usadas muy frecuentemente en la lengua familiar, 2. palabras usadas bastante en la lengua familiar y 3. palabras conocidas, pero no usadas con mucha frecuencia en la lengua hablada o pertenecientes a áreas semánticas restringidas. El autor comenta que en el primer nivel de un manual deben incluirse 1.000 palabras, más 1.018 en el segundo, 1.198 en el tercero y 1.905 en el cuarto, lo cual supone que el estudiante que complete los cuatro niveles adquiere un total de 5.121 formas léxicas.

- Díaz Castañón, C. (1977). *Vocabulario básico del español y sus aplicaciones a la enseñanza*, Oviedo, ICE.

Este es un trabajo llevado a cabo por un equipo de docentes de la antigua E.G.B. (Educación General Básica) y Bachillerato de Oviedo, centrado en estudiar el léxico del alumnado de estos dos ciclos educativos con el objetivo inicial de

ofrecer datos para describir el léxico real del alumnado de Asturias de esta franja etaria, así como la influencia que el bable puede ejercer en su vocabulario español. La finalidad era publicar una herramienta útil para el profesorado y para los creadores de materiales de lengua española, con la perspectiva de una lengua materna. Se trabajó con 3.150 alumnos —1.890 de E.G.B. y 1.260 de Bachillerato— de centros tanto públicos como privados. Se dividieron en grupos y en 5 franjas de edad para su análisis. Se creó un corpus de 850.000 palabras, en el sentido de palabra de Müller de “unidad de texto”, es decir, palabra gráfica. Se realizaron dos tipos de encuestas, una libre y otra controlada, aunque no se explica la metodología que se siguió en cada caso. La información que se da, en este sentido, es muy escueta. De la encuesta controlada se comenta que se trabajó con 11 centros de interés, por lo que deducimos que se guiarían por la metodología más habitual de los estudios de léxico disponible. Creemos que la encuesta libre fueron conversaciones grabadas que luego se transcribieron o sobre las que luego se extrajo una lista de palabras más usuales, en la línea de los proyectos de léxico fundamental del francés de Gougenheim, Michéa, Rivenc y Sauvageot de los años 50 y 60 y de algunos otros del español (Rojo Sastre 1979) y portugués (Malaca Casteiro 1984; Bacelar do Nascimento, Garcia Marques & Segura da Cruz 1987; Bacelar do Nascimento, Rivenc & Segura da Cruz 1987), que comentaremos más adelante pormenorizadamente. De los datos que sí se nos ofrecen, sabemos que de la encuesta libre se obtienen 170.000 palabras, con una media de 53 por alumno, y de la controlada 680.000. Esta última contó con los siguientes centros de interés: 1. Los animales, 2. El campo, 3. La casa y los muebles, 4. La ciudad, 5. Comidas y bebidas, 6. Los diferentes oficios, 7. La escuela y el material escolar, 8. Juegos y deportes, 9. Los medios de transporte, 10. Las partes del cuerpo y 11. Los vestidos. Sí se indica que para elaborar la lista de las palabras básicas se usaron los índices de frecuencia y reparto por grupos, aunque no se ofrecen más detalles. Para este conjunto de vocabulario se seleccionaron las palabras con frecuencia igual o superior a 20. Se incluyen en la obra el resto de palabras que no entran en estos índices de frecuencia, pues se creyó interesante aportar el resto para observar la influencia del bable sobre el alumnado.

- Morales, A. (1986). *Léxico básico del español de Puerto Rico*. San Juan de Puerto Rico: Academia Puertorriqueña de la Lengua Española.

Esta obra tiene idénticas características que el diccionario de Juilland y Chang-Rodríguez, excepto por el hecho de que los textos son puertorriqueños, de dentro y fuera de la isla, y del período situado entre 1948-1970 y el estudio que acompaña a los listados está escrito en español. Tiene también un interés didáctico y solo recoge lengua escrita. En cuanto al contenido, ofrece un listado principal de 4.456 palabras que representan 486.878 ocurrencias (un 97,37% del total del corpus de 500.000 palabras) y un apéndice de 295 palabras con índice de uso mayor de 3,08 y menor de 3,55. La obra ofrece un listado principal, lo que es el cuerpo del diccionario propiamente, en la segunda parte del libro. Las palabras están ordenadas alfabéticamente, con lo que se facilita la localización de cada palabra en concreto, pero dificulta la consulta por gradación de uso o frecuencia, de utilidad fundamental dado el tipo de diccionario de que se trata. Cada entrada da información de la clase de palabra y de los índices de uso, frecuencia y dispersión, e incluye desglosadas las variantes morfológicas con sus propios valores.

Al igual que Juilland y Chang-Rodríguez, también Morales subdividió el corpus en los mismos 5 mundos, y su concepto de palabras y las fórmulas para medir la frecuencia y la dispersión, y por lo tanto el uso son también similares. En ambas obras se dedican bastantes líneas a explicar estos criterios, que Juilland y Chang-Rodríguez perfeccionaron con respecto a los anteriores e incluso a obras posteriores, estableciendo las bases de las fórmulas y los índices con los que se obtienen los resultados con una base más científica y acertada. Morales selecciona 500.000 palabras, 100.000 por cada mundo léxico, siguiendo la división del conjunto de textos en cinco géneros tradicionales: drama (84 up⁴³), narrativa (180 up), ensayo, (187 up) literatura técnica (315 up) y periodismo (195 up). El texto deja de presentarse como un bloque indiviso para pasar a ser un conjunto de corpus diferentes que puedan medir la distribución de cada palabra en ellos y estabilizar, de ese modo, la frecuencia. La selección de oraciones de los textos se realiza *at random* (al azar): del mayor número posible

⁴³ “Unidades Publicadas” o volúmenes, artículos en el caso de la miscelánea. El número se selecciona siguiendo el volumen real de cada mundo en la producción puertorriqueña.

de páginas, y en el mayor número de obras. Elige el material en función de la clase de texto (editorial, noticias, deportes) y con numeración particular por materias, pero recibe tratamiento especial el mundo periodístico. En cuanto a la metodología, la unidad de segmentación es la palabra. Dentro de este criterio, frente a las meras listas de palabras (que parten del concepto de “palabra” como “palabra gráfica”), se opta por listas lematizadas, tomando como unidad la unidad léxica. De esta unidad léxica forman parte las formas flexivas. Esta es, en teoría, la opción favorecida mayoritariamente, pero los problemas técnicos dificultan a veces su aplicación. Se añade la posibilidad de tener en cuenta, además, la secuencia de formas que funcionan como un todo (por ejemplo: *casa de muñecas, he comido, sin embargo*, etc.), y se dice que “la metodología que incluya esta última posibilidad y a la vez elabore una lista lematizada constituirá el desiderátum de la léxico-estadística”. La autora explica en los criterios de elaboración que se guió por la intención de imprimir a la obra un carácter de sincronía, exhaustividad y homogeneidad. La sincronía, porque se eligen aquellos textos que representen una etapa dada de la vida de la lengua, lo más actual posible; de exhaustividad, porque se incluye todo lo publicado en las fechas elegidas, sin que medie valoración subjetiva alguna. Se deja fuera lo no impreso (trabajos escolares, cartas, etc., a los que sí se daba cabida en algunos recuentos anteriores, por ejemplo Rodríguez-Bou); y, homogeneidad, porque se eliminan las obras que representan lengua estereotipada o artificial (literatura infantil, devocionarios, textos gramaticales, catálogos, lengua dialectal). La muestra representativa comprende un 15% de todas las obras obtenidas para la época seleccionada.

Como ya hemos indicado, se lematizan las formas, siguiendo un criterio de no ambigüedad, que según la autora “no ha sido nunca alcanzado por ninguno de los diccionarios publicados hasta el presente”. Para ello se deben combatir problemas como la polisemia y la homografía, la división y agrupación de cadenas gráficas y la asignación de palabras a clases. Resultan especialmente problemáticas las locuciones preposicionales, adverbiales y conjuntivas, así como las perífrasis verbales, verbos reflexivos, verbos que rigen preposición, verbos compuestos, etc. Los resultados obtenidos se presentan desglosados en unidades gráficas, lo cual permite hacer estudios posteriores sobre los

compuestos o las perífrasis, y se establecen dos niveles, uno en el que se diferencian las clases formales (para lo que hay que resolver la homonimia sintáctica en casos de palabras muy frecuentes como determinantes, posesivos, diferentes tipos de *se*, etc.) y otro segundo nivel, el de las categorías (accidentes gramaticales), para lo cual se tienen que resolver las homonimias morfológicas (en cada entrada léxica se ofrecen las diferentes formas con sus respectivos porcentajes de frecuencia).

Siguiendo la metodología de Juilland y Chang-Rodríguez en cuanto a las fórmulas para medir la frecuencia, uso y dispersión, ofrece en cada palabra los tres datos por separado. Para desarrollar esta fórmula de dispersión, a su vez se basó en los cálculos de Müller sobre el promedio (ratio) del “desvío estándar a la media”, es decir, los desvíos de la frecuencia de cada término en cada subcategoría con respecto a la frecuencia teórica del término. Ofrece el coeficiente de variación de cada palabra en la muestra, independiente de su frecuencia. Esta adaptación, que desarrollan Juilland y Chang-Rodríguez en el diccionario italiano, establece una comparación entre las subfrecuencias de cada una de las categorías en que se ha dividido el texto y su frecuencia teórica. La fórmula para hallar la Dispersión es la siguiente:

$$D = \frac{I - \sqrt{n x^2_i} - T^2}{2T}$$

Donde $T^2 = (\sum x_i)^2 =$ es la suma al cuadrado de la frecuencia total de cada palabra y $x^2 =$ es el producto del número de mundos por la suma de los cuadrados reales de las frecuencias de la palabra en cada mundo. Posteriormente se hace otro cálculo sobre el número de categorías consideradas que permite obtener un índice entre 0 (dispersión mínima) y 1 (dispersión óptima). Luego se aplica otra fórmula, la del uso, que pondera la dispersión por su frecuencia, con la que se hace la selección final:

$$U = F \times D$$

La ventaja de esta fórmula es que da máxima valoración a los repartos equitativos de frecuencia frente a los que tienen mayor número de ocurrencias

pero distribución desigual. A igual frecuencia, el uso estará tanto más cercano a la frecuencia cuanto más regular sea su repartición.

- Alameda, J. R. & Cuetos, F. (1995). *Léxico de frecuencia de las unidades lingüísticas del castellano*. Oviedo: Universidad de Oviedo.

Los autores elaboran este diccionario pensando no solo en docentes de español como lengua materna, sino también en investigadores, expertos en recuperación del lenguaje y lingüística clínica. Hay que tener en cuenta que la frecuencia léxica es la variable de mayor influencia en el procesamiento del lenguaje, en la comprensión y la producción, especialmente en el caso de afasias, en las que la incapacidad de usar palabras poco frecuentes es el principal síntoma. Se parte de un corpus de 606 textos de distintos géneros públicos en los últimos 15 años antes de 1995. El recuento ofrece las frecuencias de diferentes unidades lingüísticas del español escrito: 103.335 oraciones, 2.000.000 de palabras, 3.930.954 sílabas, 4.616.502 bigramas y 9.233.004 letras. También recoge, aunque en escasa medida, siglas y abreviaturas. La información que se ofrece es la de frecuencia, sin tener en cuenta las fórmulas para medir la dispersión y el uso, por lo que no incluye tampoco la frecuencia relativa. Otro inconveniente es que se elaboraron las listas sin lematizar, es decir, se trabaja con formas exactas, no con lemas.

En la selección del corpus pretende plasmar un equilibrio entre la unidad y la diversidad del idioma. De esta forma, excluye textos dialectales o traducidos, se incluye literatura hispanoamericana, aunque en un porcentaje del 10% y con presencia mínima de elementos indígenas. Se excluyen la literatura infantil, la poesía y el teatro. Sí incluye estilos de lengua escrita: novela 50%, prensa 25%, ensayo 15%, divulgación científico-técnica 10%. La metodología que sigue es muy diferente a la de otros léxicos de frecuencias: segmenta y cuenta las sílabas de modo automático con un programa informático con reglas para la segmentación cuando los patrones ortográficos coinciden con puntos de corte de sílaba castellana, según criterios fonológicos; no se indica la categoría gramatical ni tampoco se diferencian homónimos ni palabras de una misma familia o formas verbales; se mantiene la frecuencia particular de femeninos, plurales, formas

verbales, etc.

Los resultados se ofrecen en dos volúmenes. El primero contiene listas ordenadas alfabéticamente en las que se indica el número y porcentaje de apariciones en el corpus. El segundo volumen contiene diferentes listas, clasificando las unidades por orden de frecuencia y por orden alfabético. Para cada bigrama y cada sílaba se ofrece un recuento de *types* (vocablos) y *tokens* (todas las formas). En el caso de los bigramas y las sílabas, se presentan columnas para las frecuencias en cada posición que ocupa dentro de la palabra y en función de su longitud, del tipo de palabras, etc.

Las conclusiones que se obtienen de estos recuentos son interesantes. Por ejemplo, a partir de ellos sabemos que el porcentaje de mayor frecuencia de las vocales frente a las consonantes es de 46,55% y las letras más frecuentes lo son también en inglés y francés. También hay coincidencia en los primeros puestos de frecuencia entre las palabras funcionales del español, inglés y francés, y en menor medida en las palabras léxicas. Los resultados de los primeros puestos también coinciden con los de Juilland y Chang-Rodríguez. Se confirma, así mismo, algo que había recogido ya Crystal (1987): que hay una relación entre la longitud y la frecuencia, en la medida en que las palabras de mayor uso suelen ser las más cortas.

- Justicia, F. (1995). *El desarrollo del vocabulario. Diccionario de frecuencias. (niños de 6 a 13 años de Andalucía oriental)*. Granada: Universidad de Granada.

Esta obra se presenta como un instrumento científico y pedagógico para la investigación en el ámbito de la psicolingüística, la lectoescritura, la docencia, etc. Se trata de un recuento que estudia el vocabulario de un segmento de la población muy concreto, por lo que esto condiciona el contenido y la estructura, y los resultados están directamente ligados a la muestra y no son extrapolables. En ese sentido, se intenta confeccionar una muestra que alcance el límite de palabras considerado vocabulario promedio para la edad de los sujetos y obtener una frecuencia media del conjunto que permita considerarlo usual.

Cuenta con un corpus, recogido entre 1985 y 1988, basado en producciones escritas de 3.402 niños —organizados en tres niveles por franjas

etarias de 6 a 7, 8 a 10, 11 a 13 años— de procedencia urbana y rural de Andalucía oriental (las provincias de Almería, Granada, Jaén y Málaga). Tiene más de medio millón de ocurrencias, con lo que obtiene un vocabulario básico de 8.937 palabras. Según diversos autores, la amplitud media del vocabulario del niño en esas edades se sitúa en torno a las 8.500 palabras, cifra tomada como referencia de la cantidad que se debía estimar. Los datos se recogieron mediante dos pruebas: redacción de tema libre y test cronometrado de asociación libre de palabras de 4 minutos. De la primera se obtienen las $\frac{3}{4}$ partes del vocabulario.

Las tablas con los resultados se ofrecen por ciclos educativos, en columnas de *types* y *tokens* y en dos versiones convencionales en este tipo de estudios: por orden alfabético y por orden de frecuencia. Una información interesante que se añade son los “vocabularios diferenciales”, que registran los vocablos que aparecen exclusivamente en cada nivel y no se repiten en los posteriores, e ilustran, por tanto, el ritmo de crecimiento en progresión.

Se registra un incremento progresivo del vocabulario con la edad, que no es efecto exclusivo de la escolarización, si bien esta lo intensifica. El vocabulario del niño al finalizar el ciclo medio (5º curso) es más del doble del vocabulario en el ciclo inicial; al finalizar el 8º curso lo supera casi 6 veces. Hasta el ciclo superior, la producción escrita de las niñas es superior a la de los niños, reflejo de sus diferentes desarrollos. La clase social y el ambiente urbano también se dejan notar.

- Benítez Pérez, P.; Hernández Cabrera, C. E. & Samper Padilla, J. A. (1995). Léxicos básicos de España (LEBAES) y de Canarias (LEBAICan). Proyectos de Investigación. *REALE*, 3, 9-18.

Como ya sabemos, en este caso no estamos ante una obra ni recuento léxico publicado, sino que se trata de un artículo en el que se da cuenta de un proyecto de investigación que se pretende llevar a cabo en el ámbito español, por un lado en la península y por otro en las Islas Canarias. Estos proyectos de investigación se llamaban LEBAES (*Léxico Básico de España*) y el LEBAICan (*Léxico básico de Gran Canaria*) y no llegaron a desarrollarse hasta el punto de salir a la luz algún resultado. Ambos recuentos de frecuencias seguían la misma metodología, que bebía de la seguida en los trabajos de Juilland y Traversa

(1973) y de Morales (1986). Realmente se puede decir que pretendían llevar a cabo el mismo trabajo que había hecho Morales para Puerto Rico pero en el ámbito del español de España y con textos no obsoletos, como ocurre con el diccionario de Juilland y Chang-Rodríguez, ya que se preveía seleccionar textos escritos de entre 1970 y 1994. En este sentido, este proyecto no supone ningún avance metodológico más allá de los que décadas antes habían llevado a cabo los dos trabajos mencionados. Tampoco están entre sus virtudes cuestiones como la de que la unidad mínima de segmentación tiene que superar la palabra gráfica y acceder a unidades léxicas simples y compuestas (colocaciones, locuciones, etc.) o también la inclusión de otros registros, otros tipos de textos además de los escritos y formales, o bien una concepción panhispánica del español. En ninguna de estas cuestiones pues, se daban respuestas actuales, sino clásicas: se recurría a la idea de seleccionar textos escritos de carácter homogéneo (ni orales, ni de registros diferentes, pues se copian los 5 mundos de Juilland y el número de palabras de la muestra), se lematiza el corpus (algo que hoy en día parece evidente como ventaja) y se parte de los mismos criterios en cuanto a la segmentación y clasificación de los elementos.

- Molero Huertas, J. (1999): *El vocabulario usual de nuestros alumnos*. Granada: Consejería de Educación y Ciencia de la Junta de Andalucía.

Sigue las pautas establecidas por el diccionario de Justicia (1995). Se trabajó con 932 alumnos de 14 y 15 años de Granada en una investigación llevada a cabo en los años 90. Se recopilaron 412.433 palabras para obtener 9.887 vocablos procedentes de 1864 composiciones escritas.

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

Esta obra está basada en textos orales —54 horas y media de grabación— correspondientes a 5 tipos discursivos: conversaciones cara a cara, tanto libres como semidirigidas (76%), bidireccional a distancia libre (17%), no libre (4%) y unidireccional (3%). Participan 291 sujetos de la ciudad de Málaga y se tienen en cuenta las variantes sociológicas de sexo, edad y educación. El corpus analizado consta también de 500.000 palabras para llegar a un vocabulario básico de 5.228

lemas.

El objetivo de este trabajo es determinar el español hablado en un lugar concreto (Málaga) por grupos sociales diversos y en diferentes contextos comunicativos (usando variables sociolingüísticas y situacionales) y establecer sus diferencias con índices estadísticos. Con esto se pretende dar cabida a la modalidad oral en este tipo de estudios, pues casi todos los trabajos anteriores se habían basado en la modalidad escrita, dándole preferencia como modalidad estándar, formal, cuidada, culta, etc. frente a la oralidad, ejemplo de todo lo contrario. El propio autor indica que “Existen recuentos que presentan la contradicción de pretender enseñar a hablar con materiales obtenidos de textos escritos” (Ávila Muñoz 1999). Se intenta observar el uso que los hablantes hacen del sistema léxico en esta modalidad para determinar subsistemas que pudieran ser considerados variedades sociolingüísticas.

El procedimiento para el tratamiento del corpus sigue las pautas de Juilland y Chang-Rodríguez y de Morales, distinguiendo así *types* y *tokens* y lematizando las palabras. Sigue también las fórmulas de Juilland de 1973 que explicamos más arriba para tener en cuenta el *índice de uso*, con el que organiza los resultados. No obstante, a diferencia de los recuentos anteriores no presenta el vocabulario en subconjuntos en función de su mayor uso, pues eso puede deducirse en la información que ofrece. En este sentido, sigue los parámetros del diccionario de Juilland y Chang-Rodríguez al presentar cada vocablo con sus datos sobre uso, frecuencia y dispersión.

La novedad y salto cualitativo que presenta este diccionario con respecto a los anteriores es su concepto de palabra, pues se prescinde de la dimensión gráfica y se pasa a entenderla como la unidad semántica mínima dotada de autonomía, de operatividad comunicativa. Así mismo, la consideración de unidad pluriverbal como unidad léxica aporta el matiz más novedoso. El criterio a seguir es mantener la unidad cuando no se pueda prescindir de alguno de sus componentes sin alterar el significado y no pueda introducirse ningún otro elemento entre las partes que la forman. Así, se registran locuciones adverbiales, preposicionales y conjuntivas, perífrasis verbales, verbos de uso preposicional, palabras compuestas, etc., pero se separan los verbos enclíticos, las formas contractas, las formas del pronombre, etc.

Entre las conclusiones que presenta la obra están las siguientes: las palabras de más uso son poco complejas fonéticamente (más de la cuarta parte monosílabos); en su mayoría los resultados coinciden con los de otros léxicos de frecuencia, como el de Morales (1986); hay una tendencia a la economía lingüística y la repetición, aunque se emplea menos vocabulario cuando se habla, sobre todo en conversaciones y diálogos, no tanto en los monólogos.

También se corrobora lo apuntado por otros estudios sobre las listas de frecuencia, es decir, que se obtienen mayoritariamente términos de función gramatical, con los que no es posible elaborar frases ni construir expresiones con contenido lógico, ya que faltan los elementos que expresan objetos y acciones concretas, que existen pero no son frecuentes a no ser que se aborden temas muy concretos. Entre las unidades más usadas está un reducido número de verbos, escasos sustantivos y de carácter muy general, adverbios, preposiciones, conjunciones, artículos, pronombres y apreciativos que sirven como “comodín”.

- Sebastián, N.; Cuetos, F.; Martí, M. A. & Carreiras, M. F (2000). *LEXESP: Léxico informatizado del español* [CD-ROM]. Barcelona: Ediciones de la Universitat de Barcelona.

Este diccionario, que se vende en formato de CD con un libro anexo en el que se describe sus características y manejo, se presenta como una herramienta útil no solo para lingüistas, sino también para educadores, psicólogos, investigadores e informáticos. Su contenido, la base de datos, es una herramienta codificada que permite recuperar información distribucional y estadística sobre diferentes parámetros lingüísticos y de uso. La base de datos se divide en dos partes: una que cuenta con un corpus lematizado y permite búsquedas por frecuencias de palabras, lemas o categorías, así como la frecuencia de apariciones y coapariciones, lemas y categorías de palabras en contexto; la otra tiene listados de términos indexados previamente para la creación de estudios que permitan buscar ciertas características de las palabras, además de tres estudios ya elaborados que pueden adaptarse a las necesidades del usuario. Para trabajar con este diccionario es necesario instalar un programa, llamado CORCO, que permite buscar las formas en un contexto de longitud y características ajustables.

El corpus se constituyó sobre todo con textos de procedencia europea,

pues apenas se introducen autores hispanoamericanos en un 10%, y se dejaron fuera la literatura infantil, la poesía, el teatro y las traducciones. Los textos seleccionados pertenecen al periodo de 1978 a 1995, en las categorías de narrativa, divulgación científica, ensayo, prensa, prensa deportiva y semanarios. Entre ellas prima sobre las otras la narrativa, con un 40% de representatividad.

En la descripción de esta obra, una de las ventajas que se destaca frente a los anteriores diccionarios en papel es que mejora en rapidez frente a las búsquedas manuales. Además de lentas, se comenta sobre estas que ofrecían la información fragmentada y dispersa. Esta obra, sin embargo, pretende ofrecer un conjunto de posibilidades de búsqueda mayor y más ágil. En su concepción está el fin de ofrecer con un listado lematizado y con información de frecuencia en modo informático, de forma semejante a como lo hacen los corpus actuales disponibles en línea, como el www.corpusdelespanol.org de Davies o los de la RAE. No obstante, el manejo de esta obra no es tan simple e intuitivo como las precedentes, ni está en línea para su consulta. El formato CD parece más propio de tiempos pasados que de futuros, teniendo en cuenta que en la actualidad ya se prefieren el formato mp3 o la información compartida en la Red, e incluso hay dispositivos que no cuentan con lector de CD. Por otra parte, visualmente su apariencia parece propia de los programas creados en los años 90 (la publicación es del año 2000), pues este tipo de cuestiones han evolucionado mucho desde esa década. Su manejo, en el que se usan códigos de búsqueda, también lo hacen un tanto confuso, mientras que este tipo de corpus han evolucionado hacia un sistema de búsqueda más simple, a pesar de que ofrezcan diferentes posibilidades. Sin duda, sería una herramienta mucho más útil y flexible si se encontrara alojada en la Red. Su acceso sería más fácil, a pesar de que pudiera cobrarse igualmente por su uso, y le permitiría adaptar su apariencia y opciones a los avances tecnológicos cuando fuera necesario, sin tener que publicar una nueva edición en papel y CD. Con todo, la idea de poder hacer búsquedas de las frecuencias de las palabras en un formato electrónico, que tengan en cuenta la dispersión y el uso como lo había definido Juilland, y que además permita estudiar las coapariciones, nos parece altamente interesante. Lo sería aún más si ofreciera esta opción de frecuencia de uso de las colocaciones, locuciones, etc. y si partiera de un corpus mayor, otro de sus puntos débiles.

- Terrádez Gurrea, M. (2001). Frecuencias léxicas del español coloquial: análisis cuantitativo y cualitativo. *Cuadernos de Filología, Anejo XLI*. Valencia: Universitat de València.

El autor estudia un corpus que incluye un total de 100.000 palabras procedentes de 25 conversaciones coloquiales grabadas en Valencia y su área metropolitana durante aproximadamente 8 años (de 1991 a 1999) por el grupo Val.Es.Co. Las conversaciones tratan de temas variados (familia, trabajo, tabaco, deporte, amigos, salidas, estudios, política, relaciones amorosas, enfermedad, comida, carné de conducir...). El autor del trabajo ha tenido en cuenta las variables sexo, edad y nivel sociocultural. La unidad de estudio de frecuencias va desde las letras hasta las palabras. No ofrece los listados, sino que se dedica a hacer un análisis descriptivo de tipo cuantitativo y cualitativo de los datos y muestra, en tablas, las 30 palabras más frecuentes por categoría.

Esta obra resulta interesante porque no hay demasiados estudios de frecuencias del español que se centren en describir la modalidad oral, sobre todo en la fecha en la que fue publicado. Además, se incluyen varios capítulos en la fundamentación teórica dedicados a describir y valorar las aportaciones de los corpus lingüísticos y la lingüística del corpus, la lingüística computacional y la lexicometría.

- Martínez J. A. & García, E. (2004): *Frecuencias del castellano escrito en niños de 6 a 12 años*. Salamanca: Servicio de Publicaciones de la Universidad Pontificia de Salamanca.

En realidad se trata de un diccionario que mide la frecuencia de las palabras en textos leídos por niños de 6 a 12 años. Para ello se seleccionaron 204 niños de 17 centros públicos y privados de Salamanca y se trabajó con sus manuales y textos de lectura. Con ello se ofrecen varios diccionarios de frecuencias, por cada curso y en conjunto, así como de palabras, sílabas, bigramas y trigramas. Su utilidad se defiende para diferentes ámbitos, como son investigadores del ámbito del lenguaje, la lectoescritura, la psicometría de ámbito clínico, trastornos evolutivos, etc.

En este diccionario de frecuencias de entradas léxicas leídas por escolares en cada uno de los seis cursos de primaria se entiende por palabra el conjunto de caracteres comprendidos entre espacios. Además, solo se incluyen las que constituyen entradas de la edición informatizada del DRAE de 1998. Con todo, por su abundancia, se respetan algunos vulgarismos, apócope y formas familiares como “profe”, “estao”, etc. Se trabaja con frecuencias *type* —palabras distintas— y *token* —total de apariciones—. Este diccionario se vende en soporte electrónico para facilitar su manejo en un CD multiplataforma.

Los resultados se estratifican por intervalos de 10.000 palabras (11 segmentos). Las primeras 10.000 tienen frecuencia media de 250 apariciones en disminución progresiva. Las palabras más frecuentes (palabras funcionales) son las mismas de Alameda y Cuetos (1995) y con una frecuencia similar. Las diferencias aparecen en las palabras léxicas. En cuanto a Justicia (1995), también los términos de función coinciden, si bien no en la misma posición. La comparación de diccionarios basados en muestras escritas y exposición al texto impreso muestra que no necesariamente coinciden las frecuencias relativas. Palabras que en Justicia se encuentran entre el 10% de las menos frecuentes (con una sola aparición), aparecen aquí entre las 3.500 primeras, y sucede al revés con otras.

- Almela, R. *et al.* (2005). *Frecuencias del español. Diccionario y estudios léxicos y morfológicos*. Madrid: Universitas.

Este diccionario se elaboró a partir del corpus *CUMBRE* (Sánchez, Sarmiento, Cantos & Simón 1995), un trabajo de la editorial SGEL cuyo director fue el profesor Aquilino Sánchez Pérez. Está basado en 20.000.000 de palabras del español escrito y oral de España e Hispanoamérica. Cuenta con más de 15.000 textos, cuya distribución es: España 65% e Hispanoamérica 35%. En la parte de España: 70% escrito y 30% oral (la mitad pertenece a muestras grabadas de la radio y televisión y la otra mitad a grabaciones de la vida diaria). En la parte de Hispanoamérica: 60% escrito y 40% oral. Las muestras orales se recogieron entre 1993-1994 y las escritas datan de la segunda mitad del siglo XX. El corpus se compone mayoritariamente de muestras de España, entre otras razones,

porque el español peninsular se percibe por los hablantes como una variedad de prestigio y porque las publicaciones impresas de autores españoles son más abundantes (Sánchez 1995: 26-27).

Se presenta como una obra de utilidad para lingüistas, autores de manuales, profesores de español, etc. El diccionario ofrece un subcorpus etiquetado de 2.000.000 de palabras. El Capítulo 1 ofrece las 10.000 formas más frecuentes ordenadas por frecuencia. Se analiza también la frecuencia de los significados. En el Capítulo 2 tenemos los 5.000 lemas más frecuentes y en el Capítulo 3 los encontramos por categorías morfológicas. En el Capítulo 4 se combina cuantificación y análisis para los afijos. En el 5 y el 6, se tratan las colocaciones o agrupaciones léxicas, a modo de muestra y propuesta de modelo: las 500 combinaciones —de dos y tres palabras— más comunes. Se añade, además, un glosario de términos utilizados con estadísticas.

Algunos aspectos negativos son que se limita a la propuesta de datos, sin valoración, salvo en contadas ocasiones y con finalidad ilustrativa (semántica léxica). Así, en la lista general de frecuencias, las entradas no aparecen lematizadas: tanto masculinos y femeninos, como singular y plural, los tiempos y personas de los verbos, los artículos, y también afijos, aparecen todos por separado. En los Capítulos 2 y 3 sí que se utilizan entradas lematizadas sobre una versión etiquetada de *CUMBRE*, de 2 millones de palabras. Sustantivos y adjetivos en masculino singular (aunque aparece *tía*), verbos en infinitivo, aunque aparecen como entradas independientes las formas de artículos y pronombres (*el, lo, les -no los femeninos-; te, ti, os*), las amalgamas (*al, del*), interjecciones, nombres propios (*Bogotá, Pacífico*).

En las cifras no se diferencian las formas homófonas ni homógrafas derivadas de lexemas distintos, por ejemplo, el sustantivo *vino* y la forma verbal *vino* aparecen tipificadas como una misma forma lingüística, por lo que se computa la frecuencia total de *vino* como el resultante de todas las apariciones del sustantivo y el verbo indistintamente.

La fórmula de frecuencia que se aplica es la frecuencia relativa/millón (índice ponderado) que indica las frecuencias a un millón de palabras, lo que permite comparar los datos con otros corpus de otros tamaños o diseños. Este índice se obtiene de dividir la frecuencia total por el total de formas del corpus, y

multiplicar el resultado por un millón. Con estos cálculos se establecen bandas de frecuencia a modo de tramos de frecuencia previamente definidos, del 1 al 5: frecuencia baja (1) hasta 3 ocurrencias/millón; moderada (2) de 4 a 10/millón; notable (3) de 11 a 25/millón; alta (4) de 26 a 75/millón; muy alta (5) de más de 75 ocurrencias/millón. Sobre la frecuencia, no se aplican las fórmulas que tienen en cuenta dispersión o repartición, por lo que solo se estudia la frecuencia y no el uso.

Los aspectos que más positivos nos parecen de este diccionario son que, en cuanto a los significados, se presenta una posible distribución de acepciones según frecuencia, para un número limitado de ocurrencias en cada palabra. Altamente útil, y poco común en la mayor parte de los diccionarios, es la información sobre la frecuencia de las combinaciones de dos palabras. Todas están catalogadas dentro de las muy frecuentes, y en buena medida automatizadas. Obtienen gran dispersión del rango de frecuencia, con una estabilización relativa a partir de la combinación 300. Se registran 27 construcciones sintácticas distintas, entre las que destaca de forma notable la combinación prep. + art. (35% *de la*) seguida de conjunción más artículo (8%, *que el, y la*), conjunción + prep. (5% *y en, y de*), conj. + pronombre (4% *que las, y se*). En las combinaciones de tres palabras se observa un descenso notable de uso respecto de las anteriores: las combinaciones de tres más usadas tienen casi la misma frecuencia que las menos usadas de dos. Aumentan las combinaciones sintagmáticas (42 estructuras), de las que las más frecuentes son prep. + art. + sust. (17% *de la vida, en el caso*); art. + sust. + prep. (11% *el caso de, una serie de*); prep. + art. + pron. (7% *en el que*). Se ofrece también información sobre las coocurrencias, aunque se analizan solo algunos ejemplos concretos a modo de propuesta (*hombre/mujer*, con todo tipo de palabras, relación con las distintas acepciones, etc.).

- Davies, M. (2006). *A Frequency Dictionary of Spanish: core vocabulary for learners*. New York: Routledge.

Este diccionario, cuyas 297 páginas se pueden adquirir en papel o en formato de libro electrónico en pdf., se ha elaborado a partir del corpus en línea

que el propio autor creó y que se encuentra disponible para consulta abierta en el sitio www.corpusdelespanol.org. Este corpus contiene muestras de lengua oral y textos escritos, de ficción y de no ficción tanto de España como de Latinoamérica. Hasta el año 2016 contaba con 20.000.000 de palabras, pero en esa fecha fue actualizado y mejorado.

A partir del corpus se elaboró este diccionario del español básico, que contiene las 5.000 palabras de uso más frecuente. Cada entrada nos ofrece el rango de frecuencia en la lista creada, su traducción al inglés, una frase de ejemplo con esa palabra y a qué tipo de registro pertenece o en cuál se usa más. En su propio título y desde la portadilla en la que se nos presenta de modo conciso la obra, ya se anuncia que su objetivo es ser una herramienta útil para estudiantes extranjeros de español, así como para el profesorado de ELE, en concreto a la hora de aumentar su vocabulario. No obstante, la obra está escrita en inglés, pues su autor es profesor en el Departamento de Lingüística en Brigham Young University en Utah (EE. UU.).

Nos ofrece 3 listas diferentes, es decir, nos muestra las 5.000 palabras con mayor frecuencia de uso de 3 modos diferentes: el primero es una lista con las palabras por orden de frecuencia, en segundo lugar, por orden alfabético y, en tercer lugar, por categorías. En esta última relación se distribuyen las palabras en funcionales (artículos, conjunciones, interjecciones, numerales, preposiciones y pronombres) y léxicas (adjetivos, adverbios, nombres y verbos). Otra posibilidad de búsqueda o más bien de selección de un vocabulario elemental, son los temas. Davies lo llama vocabulario temático y, a pesar de que su diccionario no tiene una organización semántica, permite estudiar un número reducido de palabras dentro de un mismo campo temático. Para reunirlos, complementa las palabras que están dentro de las 5.000 más frecuentes con otras nuevas de cada tema. Presenta los temas en cuadros destacados salpicados en las páginas del diccionario, y, al lado de cada término, ofrece también información de su frecuencia. Los va dosificando por la lista de las 5.000 palabras más frecuentes de modo que, cuando aparece una palabra frecuente que pertenece a uno de estos temas, aprovecha para introducir en esa página el cuadro correspondiente. Los temas que distingue son los siguientes:

1. Animals, 2. Body, 3. Food, 4. Clothing, 5. Transportation, 6. Family, 7. Materials, 8. Time, 9. Sports, 10. Natural features and plants, 11. Weather, 12. Professions, 13. Creating nouns, 14. Diminutives and superlatives, 15. Nouns – differences across registers, 16. Colors, 17. Opposites: frequent pairs, 18. Nationalities and place adjectives, 19. Adjectives with *ser/estar*, 20. Adjectives of emotion, 21. Adjectives – differences across registers, 22. Verbs of movement, 23. Verbs of communication, 24. Use of the “reflexive marker” *se*, 25. Preterit/imperfect, 26. Subjunctive triggers, 27. Verbs – differences across registers, 28. Adverbs – differences across registers, 29. New words since the 1800s, 30. Word length (Zipf’s Law). [Davies 2005]

En este último caso (30. *Word length*), recoge ejemplos de una tendencia en las palabras más frecuentes, que toma el nombre de Ley de Zipf por ser este autor su descubridor: las palabras más comunes de la lengua son más breves que el resto. Esto se puede visualizar en un cuadro en forma de triángulo con ejemplos de las más breves en la parte superior y las más largas en la inferior. Se aportan los datos correspondientes para mostrar cómo se produce esto también en español. Así, sabemos que de los 20 millones de palabras que recoge, 5.4 millones tienen solo dos letras, 1.3 millones tienen ocho y 74 mil palabras están constituidas por 14 letras, datos suficientemente ilustrativos.

El diccionario se presenta con un Prefacio, escrito por los editores (MacEnery & Rayson), y una Introducción del autor. Ambos textos nos ofrecen una visión general sobre los precedentes y sobre cuestiones relacionadas con la frecuencia como método de selección léxica para la docencia de lenguas extranjeras. En concreto, en apenas página y media del prefacio se comentan las virtudes de usar los datos de la frecuencia para diseñar los currículos léxicos académicos, tomando la cobertura textual como uno de los argumentos. En la Introducción, en 10 páginas se explican de modo conciso, pero con precisión, diferentes cuestiones a modo de presentación y descripción del tipo de obra ante la que nos encontramos. En un primer momento se hace un repaso crítico de varios de los diccionarios de frecuencias anteriores más relevantes, comentando sus aportaciones, pero sobre todo los defectos que hacen del español una lengua con escasez de obras de este tipo, o por lo menos obras actuales, generales, completas, representativas y realmente prácticas para aplicarlas a la docencia y convertirse en una referencia, tanto para el profesorado como para el alumnado. La obra de Davies, con quien compartimos estas críticas e impresiones, pretende erigirse en diccionario de referencia.

La verdad es que el ámbito hispánico carece de aportaciones actuales que no se circunscriban a una zona concreta y que sirvan para este propósito: las obras clásicas resultan obsoletas, otras más actuales no presentan rigor metodológico, bien en la selección del corpus, bien en su tratamiento posterior (algunas no lematizan las palabras, otras no elaboran una selección representativa, otras no proceden con las fórmulas adecuadas, otras son de acceso limitado, etc.). No obstante, creemos que destaca entre ellas el diccionario de Juilland y Chang-Rodríguez por sus avances y rigor metodológicos, a pesar de estar ya obsoleto y de que solo recogía textos escritos y de España.

En el siguiente apartado tras la Introducción, Davies se dedica a explicar la composición y organización del corpus en el que se basó para llegar a su lista de léxico básico. En primer lugar el corpus pretende ser representativo y suficientemente completo en cantidad de textos, los cuales provienen del siglo XX y están en el *Corpus del Español* (www.corpusdelespanhol.org), que recoge textos desde el siglo XIII al XX. En tipo de registros también pretende ser plural, pues tiene dos tercios de textos de registro escrito y un tercio del oral. La oralidad procede de transcripciones contenidas en el corpus *Habla Culta* (también creado por Davies, está compuesto de muestras de once países diferentes) y del *Corpus Oral de Referencia*, que recoge textos de diverso tipo de procedencia únicamente española⁴⁴. En cuanto a los textos escritos, se clasifican en dos tipos, literarios y no literarios (periódicos, ensayos, enciclopedias, cartas, textos humanísticos, etc.). La selección de estos textos se rigió por el porcentaje de 43% de España y 57% de Latinoamérica. Todos los textos son de 1970-2000, y la mayoría de la década de los 90.

Además, la lista de 5.000 palabras se creó anotando el corpus y lematizándolo. El autor explica brevemente en qué consiste este proceso, que puede resultar un tanto oscuro para personas ajenas a la lingüística de corpus o

⁴⁴ Se trata de una base de datos textual (corpus de lengua hablada) a partir de la transliteración de textos grabados en cintas de audio del registro oral (1.100.000 de palabras transliteradas en soporte informático). Dirigido por Francisco Marcos Marín, comenzó a elaborarse en enero de 1991, finalizándose en febrero de 1992. Se elaboró en la Universidad Autónoma de Madrid con una ayuda de IBM España dentro del programa de cooperación entre IBM España y la Cátedra de Lingüística General de la UAM (<http://www.llf.uam.es/ESP/Info%20Corlec.html>).

la lexicografía. De ahí que comente que se llega a un lexicón de 400.000 formas, a los que después se tuvo que asignar una categoría y a un lema, y que el lema hay que entenderlo como una “palabra base” o una “entrada de diccionario”. Este proceso se prolongó durante dos años. Las ventajas de la lematización frente a los corpus no lematizados son varias, explica Davies. Por una parte, pensemos por ejemplo en una forma verbal: aunque tengamos datos separados de frecuencia de cada forma (persona o tiempo verbal, por ejemplo), así podremos saber la frecuencia de todas las formas de ese verbo en conjunto. Este proceso se lleva a cabo con diferentes programas de ordenador, una tecnología que hoy en día está ya bastante extendida y popularizada entre los especialistas. Por otra parte, se debe decidir qué categorías gramaticales se consideran y a qué se da cabida en cada una de ellas. En este sentido, Davies habla de nombres, adjetivos, participios de pasado, determinantes, pronombres y adverbios. Sin embargo, a la hora de asignar palabras a categorías, surgen muchas dudas y problemas para distinguir los casos dudosos y difusos. La lingüística del corpus pretende ser lo más clara y rigurosa en este proceso, pues de esas decisiones deriva la composición de la lista final lematizada.

Las decisiones de Davies pretenden simplificar estos problemas en la medida en que su lista será de aplicación a la enseñanza y el aprendizaje de una lengua extranjera, ámbito en el que los matices y distinciones sintácticas o gramaticales nunca van a ser de gran envergadura. Todo lo contrario, en la docencia de ELE, la información morfológica y sintáctica que se transmita debe ser funcional y práctica para que el alumnado use correctamente las palabras sin cometer errores, aunque esté usando dos categorías sintácticas diferentes; es decir, una palabra homófona o homógrafa. Es en esta línea en la que opera Davies a la hora de etiquetar su corpus, por ejemplo al distinguir entre algunos sustantivos y adjetivos, o participios y adjetivos, o adjetivos y determinantes, etc. Evidentemente, si el resultado pretendiese ser de aplicación en ámbitos lingüísticos y especializados de la lexicografía, esto podría resultar en un proceso falto del rigor que se debe perseguir en todo trabajo científico. En cambio, puesto que la aplicación de este listado es práctica y en un ámbito muy concreto, se podrán tomar decisiones de este tipo si lo que se consigue es una mayor

operatividad y realmente ayuda a ser más eficaz con menos complejidad. Quizás lo que resulta relevante en un caso y otro cambia bastante, y en la enseñanza de la lengua a extranjeros, sobre todo en los primeros niveles, distinguir la categoría de algunas palabras no parece ser, en opinión de Davies, lo más importante si no se refleja en una diferencia de uso.

Algunos ejemplos de esto son las palabras para expresar las nacionalidades, las religiones, palabras como “rico”, etc., que pueden comportarse como adjetivos o sustantivos dependiendo de la función que desempeñen. Se optó por clasificarlos como adjetivos, porque se creyó que fácilmente el alumno sabrá aplicar la información a los casos en que desempeña una función nominal si conoce los usos de estas palabras, su funcionamiento y sus combinaciones. Los participios y los adjetivos se revisaron manualmente para valorar los adjetivos ambiguos y los que tenían un sentido marcado de agentes se recolocaron bajo el lema de participio. Con los determinantes, pronombres, adjetivos y adverbios, Davies establece aquí otra fina línea de unión entre ellos. Así, en los casos de “varios”, “algunos”, “cuyos” no ve pertinente la distinción de la lingüística tradicional entre determinantes y adjetivos, pues muchas veces los aprendices no consiguen distinguir en su uso los matices mínimos. Ocurre lo mismo entre algunos adjetivos y adverbios, o entre adjetivo y pronombres como “todos”. La conclusión para Davies fue hacer confluir estas categorías en la de adjetivo por el motivo ya mencionado, esto es, el aprendiz puede fácilmente aplicar el mismo significado a los diferentes contextos. En el caso de los pronombres átonos y tónicos, los registra por separado, al contrario que Juilland y Chang-Rodríguez, que unificaban bajo una entrada “yo, me, nos” o también “nosotros, le/les, se”.

Por otra parte, Davies explica también cómo se resolvió la cuestión de seleccionar aquellas palabras que debían aparecer en la lista de las 5.000 más frecuentes, pues hay que tener en cuenta qué factores se miden como “frecuencia”, de acuerdo con lo que hemos venido comentando en otros vocabularios. En Juilland y Chang-Rodríguez y en Morales, por ejemplo, eran importantes las nuevas fórmulas usadas para medir no sólo el número de apariciones sino también el peso que la palabra tenía en los diferentes dominios

en los que se dividía el corpus. Davies, al igual que ellos, intenta generar una fórmula que le ayude a obtener este rango de aparición o dispersión de cada vocablo, aplicándolas a las secciones que estableció, que, como ya hemos dicho, distinguían tipo de texto, registro, escritura o oralidad, etc. Por ello, crea sus propias fórmulas, que aplica al corpus teniendo en cuenta el criterio de dispersión. Con ello mide el rango y no solo la frecuencia, pues ese será el vocabulario más usual, el que aparece más número de veces en los distintos tipos de mundos.

- García Surrallés, C. (2012). *Léxico básico de la modalidad andaluza del español*, Cádiz, Universidad de Cádiz.

Este léxico básico pretende ser un instrumento útil para la enseñanza-aprendizaje del vocabulario de forma jerarquizada y la didáctica de cuestiones léxico-semánticas y ortográficas de la lengua española. Se orienta a los maestros de esta disciplina y a la elaboración de libros de texto. Se destina principalmente al aprendizaje de la lengua por hablantes nativos, aunque también se preveía su uso en la enseñanza a extranjeros.

Esta obra sigue las características de las de Juilland y Chang-Rodríguez y Morales. Se reconoce en la presentación que la idea de su elaboración proviene de un encuentro científico con López Morales en la Universidad de Huelva. La selección del material procede de muestras de lengua escrita, de primeras ediciones publicadas en Andalucía de autores nacidos en esa misma Comunidad. Las categorías que se establecen son: textos dramáticos, narrativos, ensayísticos y técnico-científicos. El periodo de tiempo al que se limita la elección va de 1973 a 1993, con lo que se mantiene la sincronía. El número de palabras que se incluye en cada universo es de 100.000, pero la escasez de textos dramáticos hizo que en ese caso se limite a 90.000. Se intenta, con estas cifras, que sea representativo, además de que son las mismas con las que trabajaron los autores antes citados. Se descarta en este caso la prensa escrita.

A la hora de analizar los datos, el concepto de palabra con el que se trabaja es el de “palabra gráfica”, es decir, el segmento gráfico comprendido entre dos espacios en blanco. Las fórmulas y los índices que se usaron fueron los que escogió Morales en su *Léxico básico de Puerto Rico*, que a su vez, recordemos,

emplea las fórmulas de Juilland y Chang-Rodríguez. Así, se presenta el vocabulario con los datos sobre la frecuencia, el uso y la dispersión, pero se ordena teniendo en cuenta la primera. Al igual que aquellos, también lematiza las formas.

Así, se cuentan 354.614 palabras, las cuales arrojan 18.891 lemas . El resultado final es un léxico básico de 4.339 vocablos. Contiene tres listas diferentes y otra más en anexo. La primera organiza las 1.000 primeras palabras por orden de frecuencia decreciente, la segunda lista son las 2.000 siguientes y el tercer listado recoge el resto hasta el total. Con esto se pretende ofrecer un primer número de palabras con las que se consigue una cobertura textual del 85%, otras con las que se consigue un 10% más, y el resto. La lista del anexo ofrece todo el vocabulario por orden alfabético.

- Listado de frecuencias de *CORPES XXI* y *CREA*. Real Academia Española.

Para observar la frecuencia de las palabras también podemos recurrir a los corpus de frecuencias que nos ofrece la RAE, y que precisamente por ser creados por esta institución son una herramienta muy potente y fiable, ya que los avalan sus criterios a nivel cualitativo y cuantitativo. Por ejemplo, el *CREA* proporciona varias listas para poder consultar o descargar: las lista de las 1.000 formas más frecuentes, de las 5.000 y de las 10.000. La información que se ofrece de las palabras es la frecuencia absoluta de esas unidades, es decir, el número total de veces que ese vocablo aparece en el corpus, y la frecuencia normalizada, que se encarga de valorar la frecuencia entre un conjunto limitado (en este caso de un millón) de unidades para poder comparar con otros listados. No obstante, el problema de estos listados está en que no valoran la dispersión de los vocablos entre los diferentes conjuntos, y por lo tanto no es posible obtener el índice de uso de cada unidad. El *CORPES XXI* todavía no ofrece listados de las palabras más frecuentes, pero añade información de la frecuencia absoluta y normalizada de cada lema y forma que se busque, y además informa de su distribución en cada país, zona y período temporal. En conversación privada con uno de los responsables de la RAE, hemos podido saber que se prevé la publicación de un listado básico de palabras a partir de este corpus, en el que sí se considerará el

índice de uso.

5.3. Léxicos frecuentes y léxicos básicos del español: obras escogidas

Después de haber valorado las opciones con las que contamos en español para obtener el léxico básico, acotamos nuestra búsqueda a un grupo de obras que consideramos más adecuadas. Son las siguientes: García Hoz (1953), Juilland & Chang-Rodríguez (1964), Díaz Castañón (1977), Morales (1986), Alameda & Cuetos (1995), Ávila Muñoz (1999), Sebastián, Cuetos, Martí & Carreiras (2000), Almela *et al.* (2005) y Davies (2006).

Sin embargo, como explicamos en el apartado sobre nuestra metodología y discusión, decidimos centrarnos en la obra de Davies (2006) para obtener el vocabulario básico seleccionado.

Capítulo 6. Disponibilidad léxica: historia, metodología y aplicaciones

Para llevar a cabo la selección del léxico disponible con el que completar el vocabulario obtenido de los corpus de frecuencia, en un primer momento hemos estudiado las diferentes aportaciones que esta línea de trabajo nos ha ofrecido desde su nacimiento en los años 50 hasta el momento actual. Así, hemos podido tener una visión global de los objetivos, metodología y resultados desde sus inicios, entender sus evoluciones y avances, sus diferentes intereses, etc. Conocer los trabajos de disponibilidad léxica en su conjunto, aunque desde una perspectiva externa a ellos, nos ha dado una visión más objetiva para evaluar cuáles han sido sus aportaciones, sus puntos fuertes y sus posibles aplicaciones, así como también sus puntos débiles y aquello que aún debe perfeccionar para conseguir mejores resultados para nuestra meta, cuyo fin es la selección de las palabras más disponibles para la enseñanza. En este sentido, haremos una revisión y descripción general de esta línea de investigación, centrándonos en algunos aspectos que para nuestro estudio son más relevantes, sobre todo en su metodología, y, por último, explicaremos cuáles han sido nuestras elecciones de entre todos los trabajos sobre léxico disponible que existen en este momento, así como los criterios que hemos seguido para quedarnos con la nómina de palabras disponibles que consideramos más útil y completa.

6. 1. Orígenes y definición

No consideramos necesario dedicar muchas líneas a explicar con profundidad la historia de la disponibilidad léxica. Existen, para ello, algunos artículos que reúnen los conceptos básicos de lo que se denomina el Proyecto

Panhispánico de Estudio de la Disponibilidad Léxica (PPHDL) y explican en qué consisten estos trabajos de investigación y sus aportaciones. Recurriremos a ellos para esta presentación de la disciplina. En lo que respecta al ámbito hispánico, la bibliografía que se encuentra es ya numerosa. Son muy abundantes los diccionarios de léxico disponible de diferentes países y Comunidades Autónomas, sobre todo memorias de máster o tesis doctorales que se defendieron desde finales de los 90 hasta la actualidad. En este sentido, la disponibilidad léxica ofrece unas pautas a seguir ya establecidas, bien definidas en el Proyecto Panhispánico, que resultan atractivas para quienes quieren realizar un trabajo que deba llevarse a cabo en un período temporal breve.

Para recopilar algunos conceptos básicos sobre disponibilidad léxica recurrentes en la bibliografía sobre este tema, podemos, por ejemplo, acudir a *Dispolex*⁴⁵. Recordemos que, según se nos indica en la propia página, este sitio pretende servir como punto de encuentro, apoyo de información y herramientas, así como base de datos para todos aquellos que con su trabajo sobre disponibilidad contribuyan de alguna manera al Proyecto Panhispánico. Pues bien, según este referente en línea sobre disponibilidad, al hablar de disponibilidad debemos entender que es:

el conjunto de palabras que los hablantes tienen en el lexicón mental y cuyo uso está condicionado por el tema concreto de la comunicación. Lo que se pretende es descubrir qué palabras sería capaz de usar un hablante en determinados temas de comunicación. Se diferencia del léxico básico en que este lo componen las palabras más frecuentes de una lengua con independencia del tema tratado. [www.dispolex.com]

Y continúa explicando el tipo de unidades que componen mayoritariamente ambos tipos de vocabulario:

Mientras que el léxico básico está formado en su mayoría por palabras gramaticales —y las que no lo son pertenecen, en orden decreciente de frecuencia, a verbos, adjetivos y sustantivos de significado general—, en el léxico disponible abundan los sustantivos que aluden a realidades concretas. [www.dispolex.com]

⁴⁵ <http://www.dispolex.com/>

Por otra parte, también sabemos que el léxico disponible es un léxico potencial, no actualizado, a diferencia de la frecuencia, que solo trabaja con léxico actualizado. Recordemos el motivo por el que se había recurrido al léxico disponible: “La suma de ambos tipos de léxico constituye el léxico fundamental de una lengua” (www.dispalex.com).

La primera definición de palabras disponibles —frente al concepto de palabras frecuentes— nos la había ofrecido Michéa:

Un mot disponible est un mot qui, sans être particulièrement fréquent, est cependant toujours prêt à être employé et se présente immédiatement et naturellement à l’esprit au moment où l’on en a besoin. C’est un mot qui, faisant partie d’associations d’idées usuelles, existe en puissance chez le sujet parlant, dès que ces associations entrent en jeu. [Michéa 1953: 340]

La segunda corresponde a López Morales, el promotor y director del Proyecto Panhispánico:

La disponibilidad léxica pasó a entenderse como el caudal léxico utilizable en una situación comunicativa dada, solo que si esa no se producía, ciertas unidades léxicas no se realizaban. Lo que comenzó entonces a vislumbrarse era que la lengua —toda lengua— poseía un amplio conjunto de palabras de contenido semántico muy concreto, que únicamente manejaba si lo permitía el tema del discurso. O dicho de otra forma: existe en el lexicón mental una serie de términos que no se actualizan a menos que se necesiten para comunicar una información muy específica. [López Morales 1995: 245]

No obstante, es escasa la bibliografía que se centre específicamente en los inicios de esta corriente, nacida en el ámbito de la lengua francesa. En la mayor parte de los casos simplemente se citan algunos de los primeros trabajos y autores, en otros se dan pinceladas de lo que pretendían hacer. Desarrollaremos estas cuestiones sobre los orígenes de los trabajos de disponibilidad en el capítulo correspondiente al léxico fundamental. En un principio hablar de léxico disponible suponía una forma de complementar el léxico frecuente y no una vía independiente de trabajo, como más tarde se ha hecho en los trabajos sobre el español, de ahí que nos tengamos que situar en describir los inicios y la historia de los vocabularios fundamentales.

Se suele resumir la trayectoria de la disponibilidad léxica con sus orígenes en Francia en la década de los 50, con un proyecto propuesto por la Unesco del que se encargó el *Centre d’Étude du français élémentaire*. El objetivo último era

difundir y promocionar el francés a nivel internacional, sobre todo pensando en las antiguas colonias francesas y en los inmigrantes que recibía el país en aquel momento, mediante la difusión de materiales didácticos cuyo método y lenguaje fuesen lo más eficientes y útiles posibles. Aunque el fin en cuanto al vocabulario era encontrar un glosario reducido con las palabras más usuales y básicas para así hacer más accesible la lengua. El proyecto, que contaba entre sus participantes con Gougenheim (director), Rivenc (subdirector), Michéa y Sauveageot, era ambicioso y pretendía jerarquizar el vocabulario con criterios lexicométricos, científicos y eficientes. Por aquel entonces, para elaborar estos diccionarios se recurría a corpus de frecuencia, sobre todo en el ámbito anglosajón, pionero en este terreno aún hoy día. La elaboración del *Français élémentaire* se inició atendiendo exclusivamente a este factor, partiendo así del índice de “frecuencia” para cuantificar los datos léxicos. Sin embargo, pronto se vio como insuficiente, ya que términos muy conocidos y usados por la comunidad idiomática francesa, o bien no se actualizaban en las listas, o bien alcanzaban posiciones tan bajas que impedía incluirlas en el recuento léxico definitivo. Los ejemplos de palabras que se consideran cotidianas pero no aparecen entre las más frecuentes son múltiples. Un caso significativo al que Gougenheim, Michéa & Sauvageot (1964: 138-139) recurrieron para ilustrar este tipo de vocablos que son altamente conocidos por los hablantes pero pueden ser poco usadas fuera de un contexto concreto fue la palabra “fourchette” (*tenedor*).

Como ya hemos mencionado con anterioridad en esta tesis, este ejemplo lo rescata Paredes García (2015: 4), quien apunta que en la lista de frecuencias del *CREA* ocupa la posición 17.942, lo que evidencia que estaría lejos de introducirse en la enseñanza del español si el único criterio de referencia fuera el de la frecuencia. Los investigadores franceses decidieron no descartar esas palabras poco frecuentes, pues reconocieron su evidente popularidad entre los hablantes de la comunidad, de ahí que se preocuparan en analizar de modo más riguroso estas unidades. Hasta que Michéa confirmó que estos tipos léxicos se actualizaban en contextos determinados y en áreas semánticas específicas, por lo que su propuesta fue estimular a los informantes mediante núcleos temáticos concretos, que llamaron “centros de interés”, con los que conseguir hacer

efectivas estas voces, a las que denominaron “disponibles”. Así pues, debemos el descubrimiento de la disponibilidad léxica y los presupuestos iniciales sobre los que se basa esta fructífera línea de investigación al artículo de Michéa (1953), en el cual delimita estos conceptos y también los de léxico “temático” y “atemático”. El conjunto del léxico temático posee un contenido semántico específico y solo se actualizan en el discurso si surgen determinados temas en la conversación. Recogen sobre todo sustantivos, y de forma decreciente adjetivos y verbos, y se caracterizan por una marcada inestabilidad estadística. Son las palabras disponibles.

Antes de esta formulación, se habían hecho algunos experimentos preliminares que dieron pistas para sentar las bases del léxico disponible. Michéa trabajó con algunos de sus estudiantes del Lyceé de Périgueux, a quienes les pidió que anotasen las veinte palabras que más les hubiesen sorprendido después de leer un fragmento de *Tonnelier de Nuremberg*. Como hoy podemos deducir, los resultados ofrecieron palabras con índices frecuenciales muy bajos, como también habían observado en otra prueba efectuada en dos liceos diferentes de París, donde se invitaba al alumnado a apuntar los veinte primeros vocablos que les viniesen a la mente cuando se imaginaban en una estación de tren justo antes de realizar un viaje. Las palabras anotadas eran mayoritariamente sustantivos y solo tres verbos de carácter general, con una frecuencia muy baja. Ambos experimentos hicieron ver a este investigador la complementariedad de los métodos de la frecuencia y disponibilidad para conocer realmente el vocabulario de una comunidad de habla, es decir, su léxico fundamental. Se habló de léxico “disponible” porque remite a palabras del lexicón mental y listas, efectivas en el discurso si se trata un determinado tema. Para ello se necesitaba unas áreas semánticas que hiciesen posible la actualización. Con este objetivo, los miembros del equipo galo recurrieron a un modelo de prueba asociativa, usual en psicolingüística, que ofrecía una serie de estímulos verbales ante los que los informantes reaccionaban actualizando términos. Estos estímulos verbales se agruparon en campos semánticos llamados “centros de interés”, que pretendían ser lo más universales y genéricos posible. Para la elaboración del *Français Fondamental* se recurrió a los mismos centros de

interés (15) que hasta la actualidad utilizan, con pocas modificaciones, los investigadores que trabajan en el PPHDL:

01. les parties du corps,
02. les vêtements (peu importe que ce soient des vêtements d'homme ou de femme),
03. la maison (mais pas les meubles),
04. les meubles de la maison,
05. les aliments et boissons des repas (à tous les repas de la journée),
06. les objets placés sur la table et dont on sert à tous les repas de la journée,
07. la cuisine, ses meubles et les utensiles qui s'y trouvent,
08. l'école, ses meubles et son matériel scolaire,
09. le chauffage et l'éclairage,
09. la ville,
10. le village ou le bourg,
11. les moyens de transport,
12. les travaux des champs et du jardinage,
13. les animaux,
14. les jeux et les distractions, y
15. les métiers (les différents métiers et non pas les noms qui se rapportent à un seul métier).

6. 2. Panorámica de los trabajos de disponibilidad léxica

Si bien con ligeras variaciones según los objetivos de cada estudio, la línea de trabajo marcada por estos lexicógrafos franceses será la que impere en los análisis posteriores, entre los que se suele nombrar a Mackey (1971) y Azurmendi (1983), investigadores sobre los que volveremos más adelante. En 1969 ya se había realizado un análisis de disponibilidad léxica en el que se intentaba demostrar la incidencia de la variable denominada "inteligencia". Dimitrijévich pasó las pruebas de disponibilidad a un grupo de 185 estudiantes escoceses de enseñanza secundaria de dos escuelas distintas, *Junior Secondary School* y *Senior Secondary School*. La escuela *Senior* obtuvo una media de 5,51 más de vocablos diferentes en cada Centro de Interés que el *Junior Secondary School*, por lo que los resultados permitieron corroborar la hipótesis de partida sobre la variable "inteligencia". El cambio que introdujo con respecto a estudios anteriores fue emplear listas abiertas y cinco minutos de límite para responder a cada campo, frente a las investigaciones francesas, que trabajaban con listas cerradas e imponían un máximo de veinte palabras para cada Centro de Interés

(Gougenheim *et al.* 1956, 1964). Las listas abiertas son las que imperan aún hoy en la tradición española, aunque se haya rebajado el tiempo dispensado para las respuestas, generalmente a 2 minutos.

Posteriormente, la investigación de Mackey (1971) sobre la disponibilidad léxica francocanadiense intenta la comparación entre dos comunidades de habla pertenecientes a una misma comunidad idiomática, por un lado, el francés de Canadá y, por otro, el francés de Francia. Para ello recoge el léxico disponible en francés de un sociolecto infantil canadiense y lo coteja con el trabajo galo. Para poder comparar ambos resultados Mackey se mantuvo fiel a las pautas metodológicas diseñadas por el proyecto pionero, ya que únicamente de este modo conseguiría unos resultados científicamente válidos. En Canadá se comienza con la disponibilidad léxica entre 1958-1959, bajo la supervisión del “Círculo lingüístico de Montreal”. Este determina que las pautas metodológicas generales a seguir por todas las universidades partícipes en el proyecto son: adoptar el modelo de encuesta elaborado para el *Français Fondamental*, añadir seis centros de interés a los 16 tradicionales (17. La Iglesia y la vida parroquial, 18. Los deportes y el equipamiento deportivo, 19. Herramientas, 20. El automóvil y sus partes, 21. Actos corrientes y 22. Negocios), asumir como sociolecto específico de la investigación a escolares de 9 a 12 años (como inicialmente habían hecho los franceses) y, por último, añadir a la prueba algunas cuestiones referentes a la edad, lecturas, etc. Estas primeras investigaciones de léxico disponible llevadas a cabo en Canadá seguían trabajando con la frecuencia relativa como parámetro exclusivo para obtener el índice de disponibilidad. Pero muestran un avance relevante en otros aspectos, pues ofrecen interesantes conclusiones de comparación intercomunitaria y etnolingüística derivadas de sus cotejos. Esta nueva vertiente será continuada en obras posteriores, como la promovida por el “Centre International de Recherches sur le Bilinguisme de la Université Laval” de Quebec, que compara el francés y el basaa de la población infantil camerunense, obra de Njock (1979). El estudio de Njock, a pesar de ser bastante desconocido, tuvo en cuenta por primera vez el ‘rango asociativo’, es decir, el puesto que ocupa cada unidad léxica en los listados de los informantes.

Los principios de la disponibilidad en español se encuentran en el ámbito hispanoamericano, concretamente en Puerto Rico, donde López Morales (1973,

1978, 1979) lleva a cabo los primeros estudios en esta línea de investigación. Recordemos que por aquel entonces el único parámetro con el que se trabajaba era la frecuencia. Antes de mejorar la fórmula para hallar el índice de disponibilidad teniendo en cuenta la frecuencia y la posición en las listas se realizaron varias investigaciones, como las obras de Mena Osorio (1986), Justo Hernández (1987), Cañizal Arévalo (1987) y Echeverría *et al.* (1987), Valencia y Echeverría (1998), Román (1985), en Hispanoamérica, y Azurmendi (1983), en España. Esta última es la obra pionera de estos estudios en nuestro país. Azurmendi se sirve de la disponibilidad léxica para explicar la situación de bilingüismo euskera-español en la comarca de San Sebastián desde un punto de vista sociolingüístico. No obstante ya existía en español la contribución de Müller de 1968, en la que critica la metodología utilizada hasta entonces y plantea la necesidad de contar con el orden de aparición de los vocablos en las listas para obtener el índice de disponibilidad, cuestión que aún tardó unos cuantos años más en resolverse, como veremos.

Asimismo, existen algunos artículos de lectura básica para quien se acerca por primera vez a la disponibilidad léxica en español y al Proyecto Panhispánico. Resultan ya clásicos dos de ellos, las aproximaciones de López Morales (1995-6), Samper Padilla, Bellón Hernández y Samper Hernández (2003) y las de Bartol Hernández (2006), en las que se ofrece una visión general sobre el tema en cuanto a definición, metodología, utilidades, situación de los diferentes proyectos, etc. El primero es el de López Morales, que nos introduce de forma sucinta en el tema, desde el origen del concepto de léxico disponible hasta los trabajos posteriores que se realizaron en Canadá (Mackey 1971), Puerto Rico (López Morales 1973, 1978, 1979) y España (Azurmendi 1983). Se ofrece una presentación general sobre el motivo de buscar este tipo de vocabulario frente al frecuente, así como la metodología usada, tanto en la puesta en práctica de las encuestas como en la interpretación de los datos obtenidos, desde sus inicios hasta la fecha de publicación del artículo. Se explica cómo surgen los centros de interés y en qué consisten las pruebas de asociación léxicas, haciendo una retrospectiva desde el primer artículo de Michéa (1953). Aborda la cuestión metodológica también de una forma panorámica a la hora de estudiar los resultados, es decir, retoma el concepto de disponibilidad desde su origen, en el

que era una mera medida de frecuencia, hasta el momento en el que se entiende el índice de disponibilidad como una medida que une criterios de frecuencia y orden, “final matemático de un camino axiomático muy riguroso”, en palabras del propio López Morales.

6.2.1. Cuestiones metodológicas: la fórmula matemática y los índices de disponibilidad léxica

Hasta el momento actual se han desarrollado diferentes fórmulas matemáticas para intentar interpretar la disponibilidad en los resultados del modo más fiable posible. Aunque es conocida por estos años la aportación de Müller a la disponibilidad, no se llega a concretar este avance metodológico hasta 1983, cuando la cooperación entre el matemático Lorán y el lingüista puertorriqueño López Morales da origen a una fórmula capaz de calcular el índice de disponibilidad léxica en base a dos parámetros: la frecuencia y la posición ocupada por las unidades léxicas en los listados del cuestionario. Con ello se intentaba tener en cuenta no solo la aparición de las palabras en las diferentes encuestas, sino también el orden en que los informantes las escribían. Para ello, las primeras fórmulas que se diseñaron procuraban ponderar la frecuencia de las palabras con los lugares que estas ocuparan en las listas. A las primeras propuestas de Lorán y López Morales (1983) y de Lorán (1987), siguieron las de López Chávez y Strassburger (1987). En el primer caso se introdujo un coeficiente con valor inferior a la unidad (valor λ) que tenía que multiplicarse por la frecuencia alcanzada por la palabra. El problema que se observó al aplicar esta fórmula es que perdía su poder discriminatorio a partir del lugar vigésimo tercero. Las aportaciones anteriores a 1983 todavía se basan en el mero criterio de frecuencia para efectuar el cómputo, a pesar de que los trabajos se realizan en una época próxima a la creación de la fórmula de Lorán-López Morales.

La segunda fórmula, la de López Chávez y Strassburger, era más compleja, y discernía con un análisis de vectores varios datos:

- a) la frecuencia relativa que obtiene cada palabra en cada posición de las listas,

- b) la frecuencia absoluta de la palabras, es decir, la suma de las diferentes frecuencias alcanzadas en cada posición
- c) el número de informantes de la prueba,
- d) el número de posición que alcanzaba en el centro de interés estudiado,
- e) las posiciones en que fue dicha cada palabra.

Esta fórmula consigue ponderar sin distorsión la frecuencia y la posición de cada palabra, midiendo su disponibilidad con una variable entre 1 a 0.1, independientemente del número de encuestados, de la extensión de los listados producidos por cada hablante y de la frecuencia de aparición del vocablo. La aplicación del factor e), elevado a un exponente complejo, es la responsable de la auténtica ponderación de la posición y lo que permite combinar de manera conveniente la frecuencia y la posición de cada unidad.

Actualmente, la fórmula de disponibilidad léxica usada es la de López Chávez y Strassburger (1991a y 1991b):

$$D(P_j) = \sum_{i=1}^n e^{-2.3 \left(\frac{i-1}{n-1} \right)} \frac{f_{ji}}{I_1}$$

Donde:

n = máxima posición alcanzada en el centro de interés en esta encuesta.

i = número de la posición en cuestión.

j = índice de la palabra tratada.

e = número de Euler, o constante de Napier (2,718281828459045...)

f_{ji} = frecuencia absoluta de la palabra j en la posición i .

I_1 = número de informantes que participaron en la encuesta.

$D(P_j)$ = disponibilidad de la palabra j .

Esta es la fórmula que se usa en el programa informático para el tratamiento del léxico disponible desarrollado para Windows, denominado *LexiDisp*⁴⁶. Este surge cuando López Morales, siendo presidente de la Asociación de Lingüística y Filología de la América Latina (ALFAL), encarga a la Universidad de Alcalá la elaboración de un programa informático para el cálculo de la

⁴⁶ Para saber más sobre el tema se pueden consultar F. Moreno Fernández, J. E. Moreno Fernández & García de las Heras (1995) y López Morales (2014).

disponibilidad léxica, con el fin de ponerlo a disposición de la comunidad investigadora internacional. En 1996, ya iniciadas las tareas de programación, el Instituto Cervantes dio su apoyo al proyecto. De esta forma comenzó la labor de diseño y realización del programa *LexiDisp*. De desarrollarlo se encargaron F. Moreno Fernández, J. E. Moreno Fernández y García de las Heras (1995), de la Universidad de Alcalá. Más tarde, el programa accesible a través del sitio *Dispolex*, permitió incluir más variables.

El concepto clave es el de *índice de disponibilidad*, que fija la disponibilidad máxima en 1 y la mínima en 0 y es el dato estadístico que los listados de léxico disponible muestran al lado de cada vocablo. Este índice permite jerarquizar las unidades y establecer comparaciones de diverso tipo. A continuación del índice de disponibilidad se suele ofrecer el porcentaje de aparición y la frecuencia absoluta, así como los porcentajes de frecuencia relativa y acumulada. Con el porcentaje de aparición conocemos el número de individuos de toda la muestra que han actualizado cada palabra. Toda esta información es la que nos ofrece el programa *LexiDisp*.

Otro análisis estadístico habitual es el que aporta el *índice de cohesión léxica*, creado por Echeverría (1991) y su equipo chileno. Con él medimos si un centro de interés es abierto/difuso o cerrado/compacto, es decir, en qué medida dentro de él las respuestas de los entrevistados son coincidentes. Se calcula dividiendo el promedio de respuestas de cada sujeto entre el número de palabras diferentes obtenidas en el centro en cuestión, y también se expresa numéricamente como (1) la mayor cohesión y más cerrado o bien (0.1) la menor cohesión pero más abierto. Según Echeverría, los centros de interés más compactos resultan más relevantes para la psicolingüística, en la medida en que ponen de manifiesto relaciones estables entre las unidades léxicas en función de su orden de aparición en las listas.

Más cálculos que se pueden realizar en los estudios sobre disponibilidad son la *densidad léxica* o el *índice de disponibilidad individual*. Para ambos se relaciona la cantidad de palabras emitidas con la cantidad de vocablos (es decir,

de palabras diferentes). Así podemos medir la densidad de respuestas diferentes en cada Centro de Interés o bien por cada informante⁴⁷.

El índice de disponibilidad no predice un comportamiento individual, sino social, de ahí que resulte útil cuando se pretende determinar la norma de la comunidad o para estudios contrastivos entre variedades sociales o geográficas. Sus aplicaciones en este sentido han sido considerables. Pero si combinamos este dato con otros valores, podremos también inferir características de comportamientos individuales. Para ello, López Chávez y Strassburger (1991a) establecen el índice de disponibilidad léxica individual (IDLI) con la finalidad de medir la participación de cada individuo en el resultado global de la muestra. Parten de la hipótesis de que si un informante actualiza los términos de mayor índice de disponibilidad léxica probablemente tendrá un índice individual de disponibilidad léxica más elevado.

$$D(S_i) = \frac{1}{k} \sum_{j=1}^n d(P_{ij}) \cdot e^{-2.3 \left(\frac{j-1}{n-1} \right)}$$

Donde:

$D(S_i)$ = disponibilidad del sujeto i

$d(P_{ij})$ = disponibilidad de la palabra respondida por el sujeto i en la posición j

n = máximo número de respuestas del centro de interés en cuestión

k = constante para ajustar las calificaciones

Más tarde, Ávila Muñoz y Villena Ponsoda (2010: 66-81) consideran que el hecho de que los sujetos pertenezcan a una misma comunidad de habla hace que compartan también una buena parte del léxico, pero para ellos la actualización de vocablos de bajo índice de disponibilidad es un indicador de mayor capacidad léxica del sujeto, teniendo en cuenta el supuesto de que la facilidad de acceso a los vocablos menos disponibles presupone el acceso a los más disponibles, aunque estos no se actualicen. Apuestan entonces por el uso de

⁴⁷ Para ampliar información sobre estas y otras cuestiones metodológicas se pueden consultar los artículos siguientes: Borrego Nieto (2004), Hernández Muñoz & Borrego Nieto (2004), López Chávez & Strassburger Frías (1991), Moreno Fernández (2012), Paredes García (2012), Jiménez Catalán (2017).

un nuevo índice, denominado índice de descentralización léxica, que determina la capacidad léxica individual tomando en cuenta las palabras que un hablante no comparte con los demás en relación con las compartidas colectivamente. Podría también explicarse como un valor que mide la descentralización o “especificidad” de la producción léxica de cada informante respecto al conjunto de datos “compatibles” o prototípicos de un centro de interés.

Las aportaciones en relación a las fórmulas matemáticas a los índices resultantes han ido avanzando todavía más, pues han surgido nuevas modificaciones y revisiones, como el índice de compatibilidad léxica de un vocablo (COMP) de Callealta Barroso y Gallego Gallego (2016). En un reciente artículo de estos autores puede leerse un resumen de todos estos índices con abundantes explicaciones y matizaciones matemáticas, así como otras aportaciones hechas por ellos mismos. También Paredes García (2012) y Jiménez Catalán (2017) se han detenido recientemente en estudiar los avances metodológicos de la disponibilidad léxica a lo largo de los años. En estos estudios se ofrecen explicaciones relativas a la evolución de las fórmulas e índices, así como a las variaciones de los centros de interés.

6.2.2. EL Proyecto Panhispánico de Estudio de la Disponibilidad Léxica (PPHDL)

A lo largo de las décadas final del siglo XX e inicial del siglo XXI se ha venido produciendo una gran eclosión de contribuciones realizadas a esta línea de investigación en variedad de países y regiones. Como observa Bartol Hernández (2006), la mayor parte de los trabajos toman el Proyecto Panhispánico como referencia, bien para seguirlo por completo y así poder establecer comparaciones, bien porque no se ajustan a él. Entre los segundos, están aquellos que contienen alguna variación en la metodología y otro grupo más significativo, que poco a poco se ha ido incrementando en las fechas más

recientes, que es el que hace las encuestas a estudiantes de español como lengua extranjera⁴⁸.

Recordemos que, tal y como resume Bartol Hernández (2006, 2007), los puntos básicos que tienen en común los trabajos de disponibilidad léxica que se alinean con el PPHDL son:

- Trabajan con, al menos, los 16 centros de interés clásicos.
- Usan como informantes a alumnado del último curso antes de los estudios universitarios.
- Encuestas escritas y con duración de 2 minutos para cada campo léxico.
- Las variables sociales que se tienen en cuenta son: sexo, nivel sociocultural, titularidad del centro de estudios y su ubicación.
- Se siguen las normas de edición establecidas por Samper Padilla (1998a), con algunas modificaciones.
- Usan los mismos programas informáticos para calcular el índice de disponibilidad: *Lexidisp* y el del proyecto *Dispolex*. Aunque diferentes, ambos emplean las mismas fórmulas matemáticas, las de López Chávez y Strassburger (1987, 1991a, 1991b, 2000).

Debido en gran medida a la existencia del PPHDL⁴⁹, en general se ha mantenido la misma metodología, para permitir los cotejos entre sintopías y la unión de los resultados parciales al conjunto global formado por los países donde se habla español. A su vez, también podemos decir que los primeros listados de disponibilidad léxica en español se identifican en líneas generales con las características de los estudios llevados a cabo en Francia. Con posterioridad se han ido aplicando los avances en las fórmulas para hallar los índices de disponibilidad y otros, así como en la aplicación de las técnicas informáticas. Y aunque muchos autores hayan procurado una mayor diversidad en las respuestas incorporando nuevos centros de interés, a veces más específicos para

⁴⁸ Junto al de Bartol Hernández (2006), otros estudios dedicados a presentar el estado de la cuestión en las investigaciones sobre la disponibilidad léxica son Carcedo González (1998b) y Ávila Muñoz & Sánchez Sáez (2010).

⁴⁹ Para saber más sobre este Proyecto: Samper Padilla & Samper Hernández (2007), Samper Padilla (2009), Samper Padilla, Bellón Hernández & Samper Hernández (2003).

adaptarse a las características de la zona y, en otras ocasiones, de diferente tipología para intentar recopilar un mayor número de tipos léxicos diferentes de los sustantivos, no se llegó a afrontar una revisión profunda sobre los centros de interés u otras cuestiones metodológicas.

Recientemente Tomé Cornejo (2015) plantea, sin embargo, una modificación profunda sobre la naturaleza, funcionamiento y resultados obtenidos de los estudios de léxico disponible, ofreciendo una alternativa y una reforma integral de los centros de interés usados desde los inicios. Se trata de su trabajo de tesis, que bebe de la experiencia de la psicología cognitiva con las encuestas de asociaciones léxicas basadas en el estímulo respuesta, y supone un salto cualitativo en la metodología de la disponibilidad léxica. Como ya hemos expuesto, esta autora reelabora una lista con 47 centros de interés y los reparte en cuatro cuestionarios, cada uno de ellos con dos modalidades. En esta distribución procuró que los estímulos relacionados semánticamente no coincidieran en el mismo cuestionario (o bien que quedaran adecuadamente separados entre sí) y que los que tuvieran estructura de *script* se presentaran conjuntamente (Tomé Cornejo 2015: 363-370) (*vid. supra* § 3.4.2.1.). Retomaremos esta cuestión en relación con las encuestas hechas a estudiantes extranjeros de español (*vid. infra* § 6.3.1.2.).

En lo que se refiere a las contribuciones a este Proyecto⁵⁰, se puede decir que son ya numerosas. En Hispanoamérica, donde se iniciaron, han sido muchos investigadores los que se han detenido en trabajar la disponibilidad léxica. A modo de resumen, recogemos la nómina de referencias por países en los que se ha estudiado la disponibilidad léxica. En Puerto Rico se sitúa el promotor y director del Proyecto, López Morales (1973, 1999). En México investigan Justo Hernández (1987), Cañizal Arévalo (1987), Ruiz Basto (1987), Mesa Canales (1989) y López Chávez (López Chávez 1992, 1993, 1994, 1995; López Chávez *et al.* 1993; López Chávez & Strassburger Frías 1987, 1991a, 1991b; López Chávez & Pérez Durán 2014). Sobre la disponibilidad léxica de estudiantes del Dorado

⁵⁰ Para describir las aportaciones hechas al Proyecto Panhispánico de Disponibilidad Léxica y la nómina de investigadores y zonas, retomamos y actualizamos las referencias de Bartol Hernández (2006) y Martínez Olmos (2007).

existe también una tesis inédita de Román (1985). En Chile trabajan Mena Osorio (1986), Vargas Sandoval (1991), Valencia (1994a, 1994b, 1997, 1998a, 1998b, 2000, 2005), Echeverría (1991, 2001, 2002; Valencia & Echeverría 1998, 1999a, 1999b; Echeverría, Herrera, Moreno & Pradenas 1987, Echeverría, Urzúa & Figueroa 2005; Echeverría, Urzúa & Sáez 2006; Echeverría, Vargas, Urzúa & Ferreira 2008). Echeverría incluye el “índice de cohesión” en el análisis cuantitativo. Recientemente se ha trabajado sobre la disponibilidad léxica matemática en la línea de la lexicometría (Salcedo, Ferreira & del Valle 2014). En la República Dominicana esta línea de trabajo se ha visto representada por Alba Ovalle (1995a, 1995b, 1996, 1998, 1999). Costa Rica cuenta con la labor de Murillo Rojas (1993, 1994) y sobre el sociolecto infantil de Sánchez Corrales y Murillo Rojas (2006). La disponibilidad léxica de la ciudad de Córdoba, en Argentina ha sido analizada por Dalurzo y González (2010). En Colombia, el equipo coordinado por Henríquez Guarín y William Santiago junto con los investigadores Ruiz y Mateus Ferro (Mateus Ferro & William Santiago 2006; Henríquez Guarín, Mahecha & Mateus Ferro 2016). En el Caribe el grupo *GIESCA*, que dirige Rodríguez Cadena. En Panamá, dirigido por Vásquez, se ha publicado el léxico disponible por Villarreal Castillo (2013). En Honduras tenemos una publicación a cargo de Martínez Segura (2012).

En la Península Ibérica, después de la iniciativa pionera de Azurmendi en el 83, estos estudios los retoma Benítez Pérez (1991), cuando realiza una reseña al trabajo de la anterior investigadora, y posteriormente contribuye a la Disponibilidad léxica centrándose en la zona metropolitana de Madrid (Benítez Pérez 1992). A este le siguen otros artículos: sobre las listas que se piden a los encuestados (abiertas o cerradas) y varias críticas sobre el tipo de vocabulario que se incluye en los manuales de enseñanza del español, para lo que apuesta por métodos lexicométricos, como la frecuencia y la disponibilidad para su selección.

Para tener una visión general de las diferentes zonas de España que cuentan con trabajos o equipos que investiguen en disponibilidad léxica, haremos una recopilación muy esquemática. En el archipiélago canario se encuentra el equipo dirigido por Samper Padilla y Hernández Cabrera (2009).

Estos y otros investigadores han producido diversos artículos relacionados con esta cuestión (Samper Padilla 1998b, 1999, 2000; Samper Padilla & Hernández Cabrera 1997, 2003, 2009; Samper Padilla & Samper Hernández 2006; Hernández Cabrera & Samper Padilla 2007). También se ha publicado una aportación sobre la “Planificación y primeros resultados de la disponibilidad léxica en Tenerife” por parte de Torres González y García Falero (2009). También en la Universidad de Las Palmas, Regla Cruz Ventura (2016) compara la disponibilidad léxica de hablantes canarios adultos con corpus juveniles en su tesis doctoral.

Andalucía cuenta con numerosos estudios, que en conjunto cubren casi la totalidad de la región. López Morales y García Marcos (1995: 65-76) publicaron un artículo donde describen algunas líneas metodológicas de las investigaciones sobre disponibilidad léxica en la comunidad andaluza. Así, Mateo García trabajó sobre Motril y Almería (García Marcos y Mateo García 1997, Mateo García 1998). También sobre Almería ha publicado Muñoz Hernández (2005). Contamos también una tesis inédita de García Megía (2003). González Martínez (2002) sobre la norma léxica disponible de Cádiz, Bellón Fernández (2011) sobre la comunidad de habla cordobesa, Prado Aragonés y Galloso Camacho sobre Huelva (Prado Aragonés & Galloso Camacho 2005, 2006a, 2006b, 2006c), Trigo Ibáñez sobre Sevilla (2011), mientras que Pastor Millán y Sánchez García trabajaron sobre Granada (2008), Ahumada Lara sobre Jaén (2006) y Ávila Muñoz sobre Málaga (2006). Prado Aragonés y Galloso Camacho han publicado el léxico disponible de Extremadura comparándolo a su vez con el andaluz, sobre el que también investigan (Prado Aragonés & Galloso Camacho 2014). Casas, Díaz, Muñoz, Escoriza, Varó (dependiente de la Univ. de Cádiz) trabajan sobre Gibraltar (Escoriza Morera 2007). Ayora Esteban lo hizo en su tesis sobre Ceuta centrándose en los aspectos sociolingüísticos (Ayora Esteban 2006). Fernández Smith, Rico, Molina y Jiménez sobre Melilla (2006).

En Castilla y León existe el grupo de investigación de la Universidad de Salamanca, formado por Bartol Hernández, Borrego Nieto, Fernández Juncal, Galloso Camacho y Santiago Guervós. Este equipo se hace cargo del léxico disponible de Castilla y León y, como ya hemos comentado, mantiene la página

oficial del Proyecto Panhispánico de Disponibilidad Léxica. Galloso Camacho (1998, 2002b) realizó el léxico disponible de la provincia de Zamora. Al mismo tiempo publicó el del distrito de Salamanca (2002a), que luego amplió a *El léxico disponible de Ávila, Salamanca y Zamora* (2003), con resultados derivados de la tesis defendida en 2001. Bartol Hernández (2001, 2002, 2003, 2006, 2007, 2010, 2011, 2014) se encargó del diccionario de Soria (2004). Más recientemente ha sido publicado el léxico disponible de toda la Comunidad Autónoma (Cruz Alonso 2016).

En Asturias, Carcedo González ha recopilado el léxico disponible de los asturianos (2001). Fernández Juncal se centró en Cantabria (2013). En el País Vasco hay otro estudio que pertenece al Proyecto, que es el llevado a cabo por Etxebarría (1996) para el área metropolitana de Bilbao. De La Rioja se ocupa Balsameda Maestu, y, de Navarra, Saralegui y Tabernerero (2008) y Areta Lara (2009).

En Galicia existe el proyecto *DISPOGAL*, gracias al que se han editado el léxico disponible del español de Galicia (López Meirama 2008, 2011) y el léxico disponible del gallego (López Meirama & Álvarez de la Granja 2014). Ambas investigadoras han tratado las interferencias y trasvases entre una y otra lengua, al tratarse de una zona de contacto y convivencia de dos lenguas próximas (Álvarez de la Granja 2011; Álvarez de la Granja & López Meirama 2013).

En la Comunidad Valenciana cada provincia ha sido analizada por un grupo de investigadores diferente: Gómez Devís (2004, 2005, 2010) ha hecho el léxico disponible de la comunidad de habla de Valencia, Blas y Casanova el de Castellón (Blas Arroyo & Casanova Ávalos 2003). Gimeno, Carcedo González y Martínez Olmos el de Alicante y Llopis Rodrigo (2006) recoge la disponibilidad léxica en valenciano estándar de la comunidad de habla valenciana, lo que supone una novedad. Martínez Olmos (2007, 2015) recoge el léxico disponible de las comunidades de habla alicantinas y trata aspectos sociolingüísticos. Casanova Avalos (2006) también ha estudiado la disponibilidad léxica en la Comunidad Valenciana. En este proyecto también hay dos comunidades de habla catalanas: Calero y Blasco, trabajan en la disponibilidad léxica de Barcelona,

Forment y Pérez sobre Tarragona y Girona. Serrano Zapata (2002, 2003, 2006) ha publicado sobre el léxico disponible español de Lleida y el contacto con el catalán. Adscrita también a la Universidad de Barcelona, Pérez Edo se centra en las Baleares. Arnal (Arnal *et al.* 2004 y Arnal 2008) y Castañer (2008) trabajan sobre las comunidades de habla de Aragón.

De Castilla-La Mancha se encarga el equipo de investigación salmantino con Hernández Muñoz, quien ya ha finalizado toda la comunidad (2006), además de hacer otras aportaciones al tema (Hernández Muñoz 2004, 2005, 2007, 2010, 2014, 2015; Hernández Muñoz & Borrego Nieto 2004; Hernández Muñoz & Izura 2010; Hernández Muñoz, Izura & Ellis 2006). De las encuestas de Murcia se encarga Ortolano, investigadora dependiente de la Universidad de Huelva.

6.3. Aplicaciones de la disponibilidad léxica

Los diccionarios de léxico disponible resultantes, elaborados a partir de esta línea de investigación de carácter cuantitativo, ofrecen a la comunidad científica materiales que, a lo largo de las tres décadas pasadas y en la actualidad, están permitiendo el desarrollo de aplicaciones de índole multidisciplinar en los ámbitos de la dialectología, la sociolingüística, la psicolingüística y la enseñanza de lenguas.

6.3.1. Disponibilidad léxica, sociolingüística y contacto de lenguas

Una de las ventajas que nos ofrecen estos estudios es la posibilidad de estudiar la variación sociolingüística de una comunidad dada a través de su vocabulario. Al realizar las encuestas de disponibilidad se pide a los informantes una serie de datos relacionados con las variables sociales consideradas, que resultan ser de gran utilidad para este objetivo (Samper Padilla 2006). Samper Padilla y Hernández Cabrera (1997) supieron sacar partido de esta información para analizar la variación sociolingüística del léxico de Gran Canaria, por ejemplo.

También se han cruzado las perspectivas de la disponibilidad y el contacto de lenguas, incidiendo sobre conceptos como los de *transferencia*, *interferencia*, *cambio de código*, etc. Se amplía así el interés por el léxico

disponible desde un punto de vista monolingüe a otro bilingüe, con lo que se adopta una perspectiva translingüística que pone el énfasis en el reflejo del hibridismo en el ámbito léxico (Mas Álvarez & Santos Palmou 2014b).

Ya en la década de los 70, tanto Mackey (1971) como Bailey Victory (1971) y Njock (1979) habían llevado a cabo estudios con el propósito de aplicar la disponibilidad léxica a comunidades de habla bilingües o bien a zonas de contacto de lenguas. En realidad, Mackey hace su trabajo con la intención de comparar la disponibilidad del francés de Canadá con el de Francia. En el caso de Bailey se centra en la disponibilidad léxica de hablantes monolingües y bilingües de español e inglés en el estado de Texas, con propósitos pedagógicos. Por su parte, Njock establece un cotejo entre dos comunidades idiomáticas en Camerún, una francoparlante y otra de lengua basaa. En España la labor pionera en este sentido se debe a Azurmendi (1983), cuya tesis doctoral describe las dinámicas sociolingüísticas entre el euskera y el español en contacto en la comunidad bilingüe de San Sebastián a partir de los cuestionarios de disponibilidad léxica. Para ello tomó como referente el estudio de Mackey (1971), aunque introdujo modificaciones tanto en los centros de interés usados como en las variables sociales de los informantes, más propias de la sociolingüística pero poco habituales en la disponibilidad léxica, como “edad”, “tipo de centro” (público/privado), “lengua materna”, “lengua utilizada en casa”, “lengua empleada con los amigos” y “lengua en la que se realiza la encuesta”. Las variables de léxico las analizó atendiendo a diversos factores, como “cantidad de producción”, “categoría léxica”, “distancia interlingüística”, “corrección lingüística” y “divisiones semánticas”. Trabajó con listas cerradas (20 palabras por centro) y concedió 10 minutos para cada centro de interés, es decir, 5 minutos menos que el profesor canadiense. Estas alteraciones sobre la metodología tradicional de la disponibilidad la llevaron a recibir críticas por parte de autores reconocidos en esta línea de trabajo como Benítez Pérez (1991: 355-7).

Con respecto a los diferentes trabajos que tienen por objeto el estudio de la disponibilidad léxica del español en situación de contacto con otras lenguas, se

pueden clasificar en tres tipos⁵¹:

a) Por un lado están los que se centran en territorios bilingües, que en el caso de España se identifican con las comunidades autónomas con lengua propia. Así, encontramos los estudios situados en el País Vasco (Azurmendi 1983, Etxebarria 1996) y Navarra (Taberner 2008), en la Comunidad Valenciana (Blas & Casanova 2003, Gómez Molina & Gómez Devís 2004, Martínez Olmos 2007, Llopis Rodrigo 2008, Gómez Devís 2004 y 2010, Llopis Rodrigo 2008, Gómez-Devís & Llopis Rodrigo 2016), en Lleida (Serrano 2006), y en Galicia (López Meirama 2008), donde también se ha publicado el léxico disponible del gallego.

b) Una segunda clase son las propuestas que en mayor o menor medida analizan los resultados del contacto de lenguas que se aprecian en la disponibilidad léxica de hablantes bilingües, sin que lo sea todo el territorio, como son los casos de Gibraltar (Escoriza 2003, 2007), contacto con lenguas indígenas en Hispanoamérica, con el inglés en EE. UU. (Sancho Sánchez 2006, Moreno Fernández 2007, Verdeses-Mirabal 2012) o derivados de otros procesos migratorios (Fernández-Merino 2013).

c) El tercer tipo lo forman los que parten del contacto fronterizo de dos lenguas para recoger los listados sobre la disponibilidad léxica, como son los casos de la frontera entre España y Portugal (Prado Aragonés, Galoso Camacho & Conceição 2010) y de Ceuta con Marruecos (Ayora Esteban 2006).

En el caso del vasco, Etxebarria (1996) y Taberner (2008) estudiaron la relación entre estas dos lenguas y el grado de disponibilidad, teniendo en cuenta para ello la variación sociolingüística y los modelos de enseñanza de los hablantes del País Vasco y de Navarra. El objetivo de la segunda era obtener la disponibilidad del español y las transferencias del euskera al castellano en función de variables como la lengua materna, la lengua de uso habitual, zona dialectal o el modelo de enseñanza cursado. Los resultados muestran simplemente pequeñas diferencias entre unas y otras variables, por lo que no se puede deducir que la presencia de una segunda lengua suponga una influencia

⁵¹ Seguimos de la aproximación hecha sobre estas cuestiones por Mas Álvarez & Santos Palmou (2014b).

negativa en la disponibilidad léxica del español, ni en lo cuantitativo ni en lo cualitativo.

Para la Comunidad Valenciana, después de haber elaborado el léxico disponible de los estudiantes preuniversitarios valencianos bajo una perspectiva sociolingüística, Gómez Devís valora en sus encuestas las dos lenguas (estudio bilingüe con cuestionarios del catalán y del español) y las consecuencias que se derivan de su convivencia, puesto que en este caso el interés fue analizar las interferencias y sus diferentes tipos, y cuáles de los términos eran los más disponibles (fueran de una lengua u otra, o bien mezcla de las dos). Tiene como meta analizar la comunidad lingüística y su compleja realidad, no solo una de las lenguas existentes en ella (Gómez Devís 2004, 2010; Gómez Devís & Gómez Molina 2004). La otra cara de la moneda la constituyen los estudios de Llopis Rodrigo (2008) y Gómez Devís y Llopis Rodrigo (2016), en los que se centran también en analizar la disponibilidad léxica y en clasificar los fenómenos resultantes del contacto lingüístico, solo que en su caso lo hacen del valenciano.

La aportación de Serrano (2006) parte de los resultados de los listados del léxico disponible del español y el catalán en Lérida para centrarse directamente en clasificar las interferencias del catalán sobre 6 centros de interés del castellano. Su trabajo ofrece, así, una clasificación de los términos que suponen algún tipo de interferencia, además de una comparación entre las diferentes palabras que se obtienen en cada uno de los centros de interés para las dos lenguas. El trabajo de Serrano establece una tipología de estos fenómenos (préstamos no asimilados, préstamos asimilados, pseudopréstamos, derivados híbridos, combinaciones híbridas, raíz híbrida y calcos semánticos).

Los estudios sobre el léxico disponible del español en Galicia añaden también nuevos análisis a los estudios ya clásicos de este tipo. Así, López Meirama (2008) tiene en cuenta en los resultados obtenidos las interferencias del gallego con el castellano y ofrece comentarios de interés. A sus líneas se suman las de otros trabajos que se centran en el gallego y que tienen la disponibilidad léxica como objetivo principal. No obstante, tanto unos como otros tienen en cuenta esta situación de bilingüismo y ofrecen algún tipo de clasificación y explicación de los fenómenos derivados de ella.

Por su parte, las investigaciones llevadas a cabo por Ayora Esteban (2006)

y Escoriza (2007) parten de la existencia de dos comunidades lingüísticas diferentes (española y árabe en Ceuta, española e inglesa en Gibraltar). Escoriza lleva a cabo un estudio de la disponibilidad léxica de las dos lenguas, lo que posteriormente le permite centrarse de forma sucinta en las interferencias que se observan entre una y otra, en concreto su análisis se basa en observar extranjerismos e interferencias ortográficas o fónicas. Del mismo modo, los estudios que se centran en analizar la situación de Ceuta parten de una realidad en la que conviven dos comunidades lingüísticas y culturales bien diferenciadas, solo que en este caso además de la convivencia entre ellas se da una situación de frontera.

Sobre la frontera entre España y Portugal, los trabajos son análisis del léxico disponible de las regiones a ambos lados de la frontera (que se corresponden con las comunidades de Andalucía y Extremadura para el español y Algarve y Alentejo para el portugués). Se hacen encuestas de cada idioma y se suman los resultados, pero no se hacen calas comparativas. Llama la atención que no se ofrezca ningún tipo de cotejo entre los resultados a ambos lados de la raya, ni exista ningún apartado que sepamos, de momento, para analizar las posibles interferencias que deben producirse entre ambos idiomas.

6.3.1.1. Adaptaciones metodológicas

Aunque el conjunto de estudios que contribuyen al PPHDL se caracteriza por la homogeneidad metodológica, justificada por la finalidad última de componer el léxico disponible del español, es natural que esta metodología, prevista desde una perspectiva monolingüe, sufra adaptaciones cuando se aplica a situaciones en las que aparecen implicados más de un idioma. Las variables sociológicas comunes a los trabajos sobre léxico disponible son “sexo”, “tipo de enseñanza”, “ubicación del centro escolar” y “nivel sociocultural”. La incidencia del factor de contacto de lenguas afecta, por tanto, a diversos aspectos metodológicos, como la inclusión de una nueva variable sociolingüística, la lengua materna —junto con la selección de informantes condicionada por ella— y los centros de interés elegidos. Los entornos bilingües, sea en contextos naturales o en contextos migratorios, ofrecen por tanto una situación

sociolingüística bien diferente de la de las comunidades monolingües y ante ella la metodología de la disponibilidad léxica adopta nuevas estrategias.

Desde el inicio del PPHDL han surgido ya bastantes investigaciones referidas a comunidades multilingües. Aunque como hemos dicho la mayor parte de ellos analiza las comunidades de habla en las que se da una situación de contacto de lenguas, limitándose al mismo tratamiento que en los estudios de disponibilidad léxica que se realizan en una comunidad monolingüe, a menudo incluyen otro tipo de variables sociales de carácter sociolingüístico como “lengua materna” (Etxebarria 1996, Blas & Casanovas 2003; Ayora Esteban 2000, Serrano 2006, Escoriza 2007), “lengua de uso predominante fuera del recinto escolar” (Etxebarria 1996, Ayora Esteban 2000), “lengua habitual” (Gómez Molina & Gómez Devís 2004, Serrano 2006), “modelo de enseñanza bilingüe cursado” (Etxebarria 1996), “lengua utilizada en casa” (Ayora Esteban 2000) y “lenguas de uso del informante” (Escoriza 2007). De este modo, a pesar de centrarse en estudiar la disponibilidad léxica de una sola lengua, hacen referencia explícita al bilingüismo existente en dichas comunidades y en varios casos estudian las interferencias, las influencias etc. que la otra lengua plasma en el español. Sin embargo otros estudios han cambiado esta perspectiva y han analizado las comunidades multilingües (como por ejemplo Lleida, Valencia y Gibraltar) haciendo las encuestas en castellano y catalán (Serrano 2003), en castellano y valenciano (Llopis Rodrigo 2006; del castellano se encargaron Gómez Molina y Gómez Devís 2004; Martínez Olmos 2007) o en castellano e inglés (Escoriza 2007). Recordemos que el trabajo de Azurmendi (1983) sobre el español y vasco fue el pionero en este aspecto.

Así por ejemplo, el estudio llevado a cabo en Bilbao por Etxebarria en los 90, que tenía como finalidad evaluar la posible influencia de los modelos de enseñanza bilingüe en el País Vasco sobre el caudal léxico español de los escolares, registra tres variables relacionadas con la lengua: “lengua materna” (euskera/castellano), “lengua de uso predominante fuera del recinto escolar” (euskera/castellano/ambas) y “modelo de enseñanza bilingüe cursado” (Modelo A, enseñanza en castellano, el euskera se imparte como asignatura/Modelo D, enseñanza en euskera, el castellano se imparte como asignatura obligatoria) (Etxebarria 1996: 311).

En la Comunidad valenciana se recogió el léxico disponible del alumnado preuniversitario de Valencia a finales de los 90 adoptando la 'lengua habitual' como la variable social portadora de mejor información sobre las diferencias en la actualización léxica, lo que permitió obtener listados de léxico disponible contrastando dos grupos de informantes: los "monolingües en castellano, aquellos que tienen por lengua de uso en la familia, con los amigos y en el centro educativo, el castellano" y los "bilingües, aquellos sujetos que tienen como lengua de uso familiar, con los amigos y en el centro escolar (patio, pasillos), mayoritariamente el valenciano o ambas" (Gómez Molina & Gómez Devís 2004: 108-14). En las pruebas de Castellón, provincia en la que el bilingüismo social e individual está más extendido, según los investigadores del proyecto, se repartió a los informantes en tres grupos: los de lengua materna castellana, los que tenían el valenciano como lengua materna y "los de familias exolingües, cuya lengua dominante en los primeros años de vida fue indistintamente tanto el castellano como el valenciano" (Blas Arroyo & Casanova Ávalos 2003: 52).

El estudio llevado a cabo en la provincia de Lleida considera dos variables, la lengua materna y la lengua de uso habitual de los informantes (Serrano 2006: 815), mientras que una única variable "lengua materna" (español/árabe) fue la elegida en Ceuta, ciudad en la que la única lengua oficial es el español, mientras que la casi totalidad de escolares musulmanes tienen como lengua materna el *dariya* ("lengua de la calle"), un dialecto oral del árabe casi exclusivo de la zona fronteriza (Ayora Esteban 2006, § 1.1.5.3.). En la comunidad de habla de Gibraltar, caracterizada por la diversidad étnica y cultural de su población, el hecho considerado en el diseño del proyecto como más significativo desde el punto de vista lingüístico fue el contacto entre el inglés, lengua oficial, y el español, por lo que en las pruebas de disponibilidad léxica realizadas para ambas lenguas se registraron las variables "lengua materna" y "lenguas de uso de los informantes" (Escoriza 2007: 2480). Un planteamiento más ajustado a la situación sociolingüística de la comunidad de habla es el del léxico disponible del español de Galicia, ya que en él se distinguen, a partir de la variable "lengua habitual" cinco grupos de informantes, en proporciones acordes con la lengua habitual de la población juvenil gallega (López Meirama 2008: 21-30).

El enfoque en la investigación sobre la provincia de Lleida supuso la

novedad de que se pretendía describir y comparar el léxico disponible en las dos lenguas que conviven en la zona (español y catalán), algo que también se propone el proyecto *DISPOGAL*. Nótese que los diccionarios de estas zonas (Lleida, Galicia) se obtienen a partir de las encuestas pasadas a los mismos informantes, mientras que los de otras comunidades seleccionan para las pruebas de cada lengua distintos grupos de informantes, en virtud de su lengua materna. Este es el caso del estudio de Gibraltar para el español y el inglés (Escoriza 2007: 2479). Así pues, en algunos de estos estudios la selección de informantes se lleva a cabo en virtud de su lengua materna o en función de determinadas características asociadas a su historia (lingüística) personal. Téngase en cuenta a este respecto el caso de la selección de informantes hispanos bilingües en comunidades no nativas, a los que se aplica las variables de lengua habitual (uno de los dos idiomas o los dos), y nivel de lengua (Verdeses-Mirabal 2012)⁵².

También es relevante el hecho de que sean factores asociados a la lengua materna de ciertos informantes lo que los convierte en inapropiados (en el estudio de Galicia se excluyen los informantes procedentes de familias no autóctonas y en el de Navarra por su procedencia exógena)⁵³. Observamos aquí la paradoja de que lo que lleva a excluir a estos informantes como no representativos para un léxico disponible general del español convierte los diccionarios en no representativos del momento real actual de la situación lingüística: de esta manera se pierde una buena oportunidad de comprobar si hay léxico característico de la zona que forme ya parte del léxico disponible de estos informantes y, si efectivamente lo hay, de qué palabras y expresiones se trata; en caso de que tal léxico no aparezca en las listas de estos informantes,

⁵² Moreno Fernández (2012: 56) ha observado que la información sobre estos factores no está basada en datos comprobados como parte de la investigación, sino que se deben al juicio o la impresión de los informantes.

⁵³ Tal y como explican las responsables del proyecto en Navarra, “se han desechado, por no corresponder al objetivo de nuestra investigación, los datos de personas que no fueran naturales de Navarra, como es el caso de alumnos inmigrantes o de otras regiones españolas, con el fin de evitar la introducción de léxico ajeno a nuestra zona lingüística” (Saralegui & Tabernero 2008: 749). López Meirama, por su parte, expone que los listados de los escasos informantes que afirman tener como lengua materna una diferente del castellano o el gallego fueron excluidos de la muestra (López Meirama 2008: 29).

quedaría más patente su menor relevancia.

La incidencia de esta variable no se ha mostrado como significativa en los algunos de los trabajos mencionados, al menos en términos cuantitativos y globales. Los análisis más finos están pendientes de llevarse a cabo o de publicación, por lo que no es posible comentar datos concluyentes. Sí podemos apuntar una reflexión a propósito del doblete lengua materna/lengua habitual como variable, ya que puede constituir una muestra interesante de la imbricación de lenguas en contextos no monolingües, de ahí la importancia de tener en cuenta, por ejemplo, lengua habitual en los jóvenes de familias inmigrantes de procedencia hispana en una comunidad en la que “se da el bilingüismo activo” (Verdeses-Mirabal 2012).

6.3.1.2. Cambios y ampliación de los centros de interés

La tendencia general en los trabajos que contemplamos aquí ha sido la de ampliar el número de centros de interés, pero sin prescindir de ninguno de los 16 mencionados, en aras de respetar la metodología común al PPHDL, que permite comparar los resultados de los estudios entre sí y, sobre todo, permitirá obtener en un futuro, quizá no ya tan lejano, el léxico disponible del español. La ampliación, sin embargo, no parece tanto provocada por el contacto de lenguas como por motivos socioculturales, por ello mismo en ocasiones sí ofrece datos relevantes en el aspecto sociolingüístico.

Con la incorporación de nuevos centros de interés se pretende obtener palabras de categorías gramaticales diferentes al sustantivo —para designar defectos y cualidades físicos y morales, para referir acciones (Ayora Esteban 2006, Escoriza 2007)—; otras veces el objetivo ha sido completar la nómina habitual con otras áreas temáticas igualmente básicas —el de los colores es un centro de interés que incluyen muchos estudios, también los temas de vida y familia, sobre salud y enfermedades o el léxico de árboles y plantas (Moreno Fernández 2007), la religión (Escoriza 2007)—; por último, la adición de algunos nuevos centros de interés se justifica por tratarse de ámbitos propios del paisaje, costumbres, historia, economía y, en conjunto, de la cultura de determinadas comunidades de habla. Como se observa en Paredes García (2012), cualquier

investigación puede incluir campos nuevos en función de las características del territorio explorado o de intereses específicos. Este es el caso del centro de interés “el mar” (o “la mar”), incluido en los léxicos disponibles de Ceuta, de Gibraltar y de Galicia. Destacamos, en el caso de Galicia, la relevante presencia de palabras gallegas en los listados de léxico sobre el mar, con cerca de cien vocablos, entre nombres de embarcaciones, artes de pesca y utensilios para pescar o marisquear, y, principalmente, nombres de especies marinas (López Meirama 2008: 89). Parece indudable que en centros de interés tan esenciales como es el mar para determinadas comunidades de hablantes, el tema condiciona la presencia de palabras de la lengua propia, en este caso el gallego, en los repertorios de léxico disponible en español: con frecuencia conviven la palabra castellana y la gallega, pero no es extraño que los informantes actualicen términos gallegos exclusivamente. En cuanto al proyecto de investigación transfronterizo España-Portugal, se decidió incluir el centro de interés “Espanha”, para las pruebas de los estudiantes portugueses y el de “Portugal” para las de los españoles, con idea de detectar las connotaciones léxicas que la evocación de cada país vecino suscitara en los informantes (Prado Aragonés, Galloso Camacho & Conceição 2010)⁵⁴.

Hemos tratado en este apartado únicamente las modificaciones que se encuentran en los estudios centrados en zonas bilingües o de contacto de lenguas. Retomaremos las propuestas de modificación de los centros de interés para observar los que se han dado con estudiantes de español extranjeros (*vid. infra* § 6.3.3.1.). También hemos presentado anteriormente la alternativa ofrecida por Tomé Cornejo (2015) (*vid. supra* § 3.4.2.1.).

6.3.1.3. Trasvases léxicos entre lenguas

Uno de los hechos esperables tras pasar las pruebas de léxico disponible a hablantes en contextos de lenguas en contacto es que, como consecuencia de tal vecindad, se registren listados híbridos en los que haya vocablos pertenecientes

⁵⁴ Sin embargo los listados correspondientes no se han incluido en los diccionarios publicados en 2010, por lo que no es posible valorar en qué medida se nutren de estereotipos de cada país vecino.

a las distintas lenguas implicadas. De hecho, este es uno de los objetivos de los estudios llevados a cabo con hablantes hispanos en los Estados Unidos, el de detectar en qué medida penetran los anglicismos en el español de los nativos (López Morales 1999a, Moreno Fernández 2007), más patente cuando entre las instrucciones para la prueba se les indica a los informantes que escriban todas las palabras que les vengan a la mente sobre el tema propuesto y que, si no las recuerdan en español sino en inglés, las escriban en inglés (Verdeses-Mirabal 2012).

Los estudios estadísticos de tipo cuantitativo culminan con la reunión del léxico disponible en diccionarios que presentan las palabras de cada ámbito temático ordenadas según su índice de disponibilidad. El índice de disponibilidad (ID), recordemos, se obtiene teniendo en cuenta tanto la frecuencia como el orden de aparición de la palabra en los listados de respuesta. Lo que nos preguntamos de manera inmediata ante los listados de zonas o hablantes bilingües es en qué medida están presentes en ellos las otras lenguas diferentes del castellano y en qué medida podemos considerar que tal presencia es significativa. En cuanto a los resultados de las zonas bilingües de España, solo conocemos un acercamiento de conjunto, el que se hace en Bartol Hernández (2011). En este estudio se ofrece una valoración de la presencia de vocablos pertenecientes a las diversas lenguas vernáculas —catalán, con su variante valenciana, vasco y gallego— a partir de los diccionarios elaborados en cada zona. Lo que interesa es registrar aquellas palabras cuya presencia en los listados se explique por el contacto actual de lenguas, en lo que se ha venido llamando *interferencias*, *transferencias* o *cambios de código*, y valorar en qué medida estamos ante un conjunto relevante desde el punto de vista cuantitativo. Tras fijar el límite de lo que considera una disponibilidad representativa en un ID alto, de al menos 0,01, el autor tiene en cuenta también el número de informantes que lo han incluido en sus listas (como referencia, el 2%=16) para determinar si un vocablo está integrado (Bartol Hernández 2011: 175). Estas premisas arrojan los siguientes datos, por ejemplo, para Galicia, Navarra o Valencia: en Galicia sólo 19 palabras gallegas se consideran integradas en el léxico español (3,37%), de ellas las más utilizables son *pota*, *faiado* y *sachar*; en Navarra, solo 4 de los 52 términos vascos pueden considerarse integrados

(*patxaran, borragoma, txoko y gaztetxe*); en Valencia, de los 169 catalanismos solo seis son representativos, los más destacados *truc* y *sambori* (del centro de interés “Juegos y diversiones”).

Las conclusiones de Bartol Hernández apuntan en la línea de algunos de los trabajos de los que parte: la incidencia de los vocablos de las lenguas propias en los listados de léxico disponible de zonas de contacto es escasa, prácticamente anecdótica en términos cuantitativos. Es evidente que el énfasis de trabajos surgidos dentro del proyecto panhispánico, tanto en sus presupuestos, como en la edición y análisis de los datos va enfocado a subrayar la homogeneidad del léxico obtenido en las diferentes zonas, una homogeneidad reforzada, como proponen algunos, tanto como consecuencia de la labor uniformadora de la escuela como por la semejanza en las referencias culturales y vitales de las chicas y chicos en edad preuniversitaria. Sin embargo, otros acercamientos proponen una consideración más abierta de las hibridaciones lingüísticas motivadas por el contacto (Álvarez de la Granja 2011) o estudian en detalle la incidencia de cada una de las variables sociológicas contempladas en la actualización de léxico en una u otra lengua (Álvarez de la Granja & López Meirama 2013), en una línea que podría inscribirse en un nuevo paradigma de los estudios de contacto de lenguas (Gugenberger 2013).

De modo general, se observa que desde el pionero trabajo de Azurmendi el interés por analizar la influencia que el bilingüismo y el contacto lingüístico tienen sobre la disponibilidad del léxico ha ido en aumento, y la mayor parte de la producción científica en esta línea es del siglo XXI, por lo que no parece que este interés vaya a decaer. Además, se observan trabajos más recientes que se centran en la interferencia del contacto de lenguas en concreto, y no solo de modo tangencial. Se tendrá que ver en un futuro si esta motivación deja alguna impronta relevante a nivel teórico en el desarrollo de la línea de investigación de la disponibilidad léxica, dando cabida de algún modo a la variación lingüística y las interferencias que de ella deriven. El enorme interés de este terreno y la profusión de trabajos auguran buenas perspectivas, serán más prometedoras cuando se hayan recogido datos de las dos lenguas cooficiales en cada zona.

Observamos una tensión latente entre los análisis y las valoraciones cuantitativos y los acercamientos de carácter cualitativo. Los fenómenos de

transvase e hibridación que en los primeros suelen minimizarse y relegarse a segundo plano como irrelevantes, cobran protagonismo e interés cuando el enfoque del análisis va al detalle y bucea en las diferencias que ofrecen las variables o en los listados de determinados centros de interés. En nuestra opinión, esta segunda manera de enfocar el acercamiento a los datos puede aportar una perspectiva enriquecedora que no se limite a explicar las supuestas anomalías de los listados de léxico disponible como ejemplos de situaciones deficitarias o atípicas.

El punto de vista monolingüe se está viendo desbordado por la realidad de la lengua viva en situaciones de contacto, por la hibridación que ofrece el léxico disponible de jóvenes para quienes el castellano es una de las lenguas de referencia, solo una, junto a otra u otras. Además de los interesantes ejemplos de dialectalismos y extranjerismos —sobre todo de anglicismos— que todos los diccionarios de léxico disponible muestran, es relevante en estos estudios de disponibilidad léxica en español la presencia de formas de otras lenguas próximas, sea por su cooficialidad sea por el bilingüismo de los informantes. Los corpus de léxico disponible están proporcionando interesantes datos de hibridación de lenguas en contacto; qué sentido y relevancia les otorgue la comunidad científica a estos datos constituye un atractivo asunto para ulteriores indagaciones.

6.3.2. Disponibilidad léxica y dialectología: cotejo de resultados entre estudios y sintopías

Una de las zonas de habla hispana que ha sido analizada desde este punto de vista ha sido la isla de Gran Canaria. Los autores, Samper Padilla y Hernández Cabrera (2003), aprovechaban los resultados de un centro de interés concreto para explorar esta vía. Más tarde, los mismos investigadores han hecho más aportaciones a la fusión entre disponibilidad y dialectología (Hernández Cabrera y Samper Padilla 2007).

Dentro de la bibliografía sobre disponibilidad léxica han surgido otros estudios que muestran resultados interesantes. Nos referimos a los que cotejan los listados de vocabulario disponible recogidos en diferentes lugares haciendo

hincapié por ello en las diferencias dialectales del vocabulario de las comunidades de habla o bien comparando la producción léxica de estudiantes extranjeros.

López Chávez (1992a y 1992b) desarrolló una serie de fórmulas que permitían cotejar los resultados de los léxicos disponibles de dos o más lugares desde una perspectiva diatópica. Gracias a ellas podemos observar las unidades comunes y divergentes, tanto en total como por centros de interés, con el fin de alcanzar el grado de aproximación entre dialectos. En concreto, estadísticamente conseguimos obtener los siguientes datos (López Morales 1995-1996: 149-150):

- 1) la cardinalidad de los conjuntos (la comparación se hace de dos en dos),
- 2) la intersección,
- 3) la unión,
- 4) los complementos de ambos conjuntos,
- 5) la suma disyuntiva de los dos conjuntos,
- 6) la compatibilidad de los conjuntos,
- 7) la contribución de ambos conjuntos en la intersección,
- 8) el contraste de cada conjunto con respecto al otro y
- 9) los códigos exclusivos de cada conjunto.

Con estas fórmulas, López Chávez trabajó con los datos de las localidades de Madrid, Las Palmas, República Dominicana y Puerto Rico y llegó a la conclusión de que el mayor parecido se produce entre Madrid y Las Palmas, a continuación, Las Palmas y República Dominicana. El menor grado de coincidencia tenía lugar entre Las Palmas y Puerto Rico, a continuación Madrid y República Dominicana y el menor de todos era entre Madrid y Puerto Rico. Con todo, estas comparaciones de López Chávez han sido criticadas por su metodología, ya que daba lugar a coincidencias demasiado escasas. Hay, por ejemplo, otro artículo de Samper Padilla (1999) comparando la variación dialectal con datos de Puerto Rico y Gran Canaria en el que se proponen aproximaciones más idóneas.

El afán de comparación entre sintopías también se ha centrado en comparar la producción de estudiantes extranjeros y hablantes nativos de español. Un ejemplo es el de Galloso Camacho y Prado Aragonés (2005), que ponen en común los datos obtenidos por Sánchez Gómez (2003) con alumnos de

español como lengua extranjera y los que se obtuvieron con preuniversitarios nativos de Huelva, Zamora y Cádiz. En cuanto a los resultados, como en otros estudios, concluyen que el número de respuestas es menor en los no nativos, pero aumenta según el nivel de estudios, y que los centros de interés más y menos productivos en cuanto a número de respuestas son aproximadamente los mismos en los dos tipos de informantes. A continuación ofrece los 20 primeros vocablos más disponibles en los diferentes estudios, con lo cual se muestra que hay una gran coincidencia entre ambos, incluso en algunos casos aparecen las palabras en la misma posición. Se observa también una mayor semejanza en los centros más concretos, frente a los abstractos —lo mismo había observado Carcedo González (2003) cuando comparó las sintopías de Asturias, Puerto Rico y Cádiz—, por lo que parece que la cohesión tiene más que ver con el tipo de campo léxico (o centro de interés) que con el lugar de procedencia del informante o su carácter de extranjero o nativo.

Un trabajo muy similar es el estudio de Šifrar (2009), en el que se compara cuantitativa y cualitativamente la disponibilidad léxica de un grupo de alumnos eslovenos de español como lengua extranjera con la obtenida en otras dos investigaciones sobre la disponibilidad léxica en ELE, una llevada a cabo en Finlandia (Carcedo González 2000) y otra en Salamanca (Samper Hernández 2002). Algunas conclusiones de este trabajo ponen de relieve la similar competencia léxica básica que manifiestan los estudiantes ELE, a pesar de tener lenguas maternas tan diversas y vivir en culturas muy diferentes. Además, los tres estudios muestran mayor productividad de palabras en el tema de la comida y bebida, seguido por el centro de interés de la ciudad. El centro menos productivo en Eslovenia ha sido "la ropa"; en Salamanca y en Finlandia coincide el de "trabajos del campo y del jardín". En cuanto a los vocablos o palabras diferentes, en Eslovenia el centro de interés más variado es "la ciudad", mientras que en Salamanca y Finlandia es el de "alimentos y bebidas", mostrándose de nuevo como el campo más productivo, esta vez en número de vocablos. Sorprende la gran cantidad de coincidencias que se producen entre las listas de las tres investigaciones, llevadas a cabo en lugares tan diferentes de Europa. Las palabras con mayor disponibilidad que aparecen en las tres listas suman 12: *perro, coche, gato, cabeza, bicicleta, profesor, autobús, avión, tren, ojo, fútbol, calle.*

Entre Eslovenia y Salamanca hay 15 palabras en común, al igual que entre Eslovenia y Finlandia, mientras que entre Salamanca y Finlandia hay 14. Lo más sorprendente son, sin duda, las primeras posiciones, en las que hay gran similitud (*perro, coche, gato, profesor*). En cuanto a la comparación cuantitativa, la producción general de palabras y vocablos en Eslovenia es superior a la de Salamanca y Finlandia. Por otro lado, la coincidencia de los centros más productivos es sorprendente: como hemos dicho, los informantes de los tres estudios han mostrado mayor productividad de palabras en el campo sobre "la comida y bebida", seguido por el tema de "la ciudad".

6.3.3. Disponibilidad léxica y enseñanza-aprendizaje de español como lengua extranjera (ELE)

Como acabamos de ver, los estudios de disponibilidad léxica han alcanzado a los estudiantes de español como lengua extranjera y han ofrecido interesantes posibilidades de aplicación, como muestran los estudios del reciente volumen coordinado por Jiménez Catalán (2014).

Esta vía nos puede servir para:

examinar las distintas fases de aprendizaje del léxico español, descubrir tanto los errores más usuales como las tendencias de los hablantes no nativos a lo largo de este aprendizaje, planificar la enseñanza del léxico o realizar provechosas comparaciones con los listados de las distintas zonas hispanohablantes. [Samper Hernández 2002: 14]

Sin embargo, y al contrario de lo que sus orígenes pueden hacer pensar, la vía que más se ha explotado de la unión entre disponibilidad léxica y ELE ha sido la que estudia la disponibilidad del vocabulario de los discentes de español, y no la que se centra en planificar la enseñanza del léxico. Esta última ha sido siempre relegada a un segundo plano, mientras que la primera fue explotada desde el inicio de la década del 2000 y cuenta con numerosas aportaciones hasta la actualidad.

Por una parte, tenemos los trabajos que han usado la disponibilidad léxica como referente, bien comparando los resultados de las encuestas hechas a nativos con el vocabulario que se introducía en los manuales de ELE, bien

cotejando los datos de diferentes localidades, tanto con hispanoparlantes como con estudiantes de ELE. En esta línea los primeros esfuerzos se deben a un grupo de investigaciones que reivindican sumar a la enseñanza del vocabulario los resultados a los que se había llegado con las pruebas de disponibilidad en nativos (Benítez Pérez 1994a, 1994b, 1995 y 2009; y, más tarde, García Marcos & Mateo García 2003; Salazar García 2004; Guerra Salas & Gómez Sánchez 2004, 2005; Izquierdo Gil 2005, 2016; Núñez Romero 2008a, 2008b; Rodríguez Muñoz & Muñoz Hernández 2009, 2010; Bartol Hernández 2010; Palapanidi 2012). Inciden en la importancia que tiene el léxico en el aprendizaje de una lengua, aun a pesar de su complejidad, pues hablamos de paradigmas abiertos y muy variables. Así, todos coinciden en que la selección del vocabulario requiere criterios objetivos y científicos. Y es en este punto en el que se revela útil la disponibilidad, ya que ofrece amplios repertorios léxicos obtenidos mediante una metodología consolidada y permite, además, la elaboración de las unidades teniendo en cuenta variables e intereses concretos (como el nivel de lengua que se estudia, los lugares de donde se han obtenido las palabras y la comparación con otros puntos, la edad de los informantes, el estudio de campos específicos de vocabulario, etc.). La mayor parte de los autores son conscientes de estas ventajas, por lo que reivindican acudir a la disponibilidad como una de las fuentes objetivas.

Benítez Pérez recurre a la comparación del léxico de los diccionarios de disponibilidad léxica extraídos en territorio madrileño con el de manuales muy empleados en ese momento en la didáctica de ELE (Benítez Pérez 1994a). García Marcos y Mateo García (2003) tratan el centro de interés de “la alimentación”. No obstante, los resultados y conclusiones son compartidos, puesto que en ambos casos se echan en falta en los manuales de ELE palabras muy disponibles y, por el contrario, se encuentran otras que no gozan de un gran índice de disponibilidad. En concreto, es interesante observar cómo la selección aleatoria resulta incoherente a nivel cualitativo, pues hay palabras que se introducen en el aprendizaje de niveles que se no corresponden con su grado de dificultad o viceversa, además de que no siempre se opta por las variables (geográficas o diafásicas) más usuales o adecuadas en cada caso. En sus trabajos, Izquierdo Gil

(2005, 2016) se centra en un ámbito concreto, el francés, y aunque también compara los materiales y realiza encuestas a profesorado y alumnado, no tiene en cuenta listas de disponibilidad léxica elaboradas en España. Con todo, en su tesis de doctorado hace un amplio análisis de estas cuestiones y aporta numerosos puntos a tener en cuenta para mejorar la enseñanza del vocabulario a discentes franceses de español. Núñez Romero (2008b) enfrenta el vocabulario de dos manuales de nivel inicial con los resultados de los diccionarios de disponibilidad de Cádiz, Zamora, Valencia y Málaga para el centro de interés “Medios de transporte”.

En un segundo grupo están los trabajos que estudian la disponibilidad léxica en estudiantes de ELE. El iniciador en esta temática fue Carcedo González con toda una serie de trabajos con estudiantes de origen finlandés (Carcedo González 1998a, 1999a, 2000), que culminan con una amplia monografía (Carcedo González 2000). En ella se analizan las producciones de 350 informantes finlandeses de nivel preuniversitario y universitario (con finés o sueco como lengua materna) ante los dieciséis centros de interés tradicionales. Ofrece un análisis cuantitativo y cualitativo de los datos y examina la incidencia de las variables “sexo”, “nivel de estudios”, “lengua materna” y “conocimiento de otras lenguas románicas”, y compara sus resultados con los obtenidos en la República Dominicana, Puerto Rico, Cádiz y Zamora con hablantes nativos.

Un año más tarde, Samper Hernández (2001) publica el primer trabajo de disponibilidad léxica realizado con estudiantes de ELE de distintas lenguas maternas y que se ubica en España, es decir, en un contexto de inmersión lingüística. En este caso la muestra es de 45 alumnos de Cursos Internacionales de la Universidad de Salamanca. Samper Hernández se sirve, al igual que Carcedo González, de los dieciséis centros de interés tradicionales, y contempla las siguientes variables extralingüísticas: “sexo”, “conocimiento de otras lenguas”, “nivel de dominio de la lengua extranjera” y “lengua materna de los estudiantes”. Además del análisis cuantitativo y cualitativo de los resultados, la investigación se completa con el cotejo de los datos del alumnado finlandés extraídos de la monografía de Carcedo González (2000).

A partir de estos dos trabajos, se ha ido sucediendo toda una serie de investigaciones que los toman como referencia, muchos de ellos trabajos académicos como tesis o memorias de máster⁵⁵. Sirvan como muestra los puntos geográficos que han sido analizados hasta ahora en el extranjero: Finlandia (Carcedo González 1999a, 1999c, 2000), Islandia (Magnúsdóttir 2012), Polonia (López González 2009), Turquía, China (Jing 2006, 2012), Marruecos (Serfati 2010; Herreros Marcilla 2016 con marroquíes inmigrantes en España), Eslovenia (Šifrar Kalan 2009, 2011, 2014), Rumanía (Sandu 2009, 2010, 2012, 2013; Jimeno Panés 2006), Alemania, Portugal (Fernández dos Santos 2013, 2014; Gamazo Carretero 2014; Rubio Sánchez 2016; Santos Palmou & Mas Álvarez 2017), Italia (Roberto Rubio 2015), Brasil (Pedroni Torres 2015) y los lugares de la geografía española en los que se han hecho encuestas a estudiantes extranjeros en un contexto de inmersión lingüística, como Valladolid y más tarde toda Castilla y León (trabajando con inmigrantes Fernández-Merino Gutiérrez 2013, 2015), Salamanca (Cursos Internacionales de la USAL), Alcalá de Henares (centro de idiomas de la universidad), Madrid, Huelva, Sevilla, Cádiz, Ibiza, Navarra. Sánchez-Saus (2016) ha recogido el léxico disponible de estudiantes de español como lengua extranjera en las universidades de toda la comunidad andaluza.

Tomé Cornejo (2015: 338-9) nos ofrece una útil clasificación de los diferentes tipos de aproximaciones que se han hecho a la disponibilidad léxica en el ámbito del español como lengua extranjera⁵⁶. Partimos de ella para elaborar este esquema:

a) Trabajos que cotejan los resultados con el léxico disponible de hablantes nativos (Carcedo González 2000 y 2003; Galloso Camacho & Prado Aragonés 2005; Núñez Romero 2008a; Sandu 2010; Fernández-Merino 2013; Herreros Marcilla 2014, López González 2016), o bien con el de otros estudiantes de

⁵⁵ Para una revisión bibliográfica de este tipo de estudios hasta el año 2014 y de sus características básicas, se puede consultar el artículo de López González (2014).

⁵⁶ Para obtener un resumen de las características de cada una de estas publicaciones, consúltense las páginas 328-338 de la tesis de esta investigadora (Tomé Cornejo 2015), donde se presenta una nómina de autores y se resumen las características básicas de los trabajos.

español (Samper Hernández 2001; Šifrar Kalan 2009, 2012, 2014; Fernández-Merino 2011 y 2013; Gallego Gallego 2014). Otros establecen la comparación con el vocabulario actualizado en textos producidos por los propios participantes (Frey Pereyra 2007) o con manuales o repertorios léxicos de ELE (Fernández-Merino 2011; Jiménez Berrio 2012).

b) Algunos autores relacionan la disponibilidad con el nivel de español. Algunos cuentan con los resultados de informantes de distintos niveles (Carcedo González 1997-1998 y 2000; Samper Hernández 2001; López González 2010; Sánchez-Saus 2011; Jing 2006, 2012; Šifrar Kalan 2014), otros analizan resultados de un mismo grupo, pero en dos etapas diferentes del aprendizaje (López Casado 2013).

c) Otros clasifican las respuestas registradas en función de su naturaleza gramatical (Sánchez Gómez 2006; Sánchez-Saus 2011; Jiménez Berrio 2013), o del tipo de error que representan (Carcedo González 1999c; Frey Pereira 2007), también según el registro al que pertenecen, su composición, su condición de extranjerismos o de eufemismos, metáforas, etc. (Sánchez-Saus 2011).

d) Un análisis interesante es el que profundiza en las relaciones semánticas que se establecen entre las palabras más frecuentes mediante la generación de grafos (Sánchez-Saus 2011).

e) También los hay que se centran en estudiar centros de interés concretos Pérez Serrano (2009), e incluso nuevos centros de interés, como “Los negocios” (Santos Palmou & Más Álvarez 2017).

Frente a los trabajos con informantes nativos, que mayoritariamente han optado por establecer el corte de edad de los informantes en 17 y 18 años, los estudios de disponibilidad léxica de ELE manifiestan heterogeneidad en este aspecto, al tomar como objeto de estudio los diferentes niveles educativos en los que se desarrolla la enseñanza del español como lengua extranjera en diferentes contextos educativos: educación primaria, secundaria, enseñanza superior, escuelas de idiomas, etc. No obstante, al tratar con informantes extranjeros aprendices de español la edad queda en un segundo plano y el factor más

relevante pasa a ser el nivel educativo, pues este puede variar independientemente de la edad de los discentes. En este sentido, aunque algunos trabajos han separado las muestras en los niveles Elemental – Intermedio – Avanzado, o bien en los seis grados que contempla el *MCERL* (A1, A2, B1, B2, C1 y C2), los que se han detenido en estudiar en concreto esta división han concluido que “el crecimiento en el conocimiento del léxico se da entre los niveles inicial y medio, entre A y B, mientras que entre B y C este se estanca” (Sánchez-Saus 2011: 215).

Una reciente aportación a los estudios de disponibilidad y enseñanza de lenguas extranjeras es la obra colectiva *Lexical Availability in English and Spanish as a Second Language*, con investigaciones centradas en la recopilación y descripción de los léxicos disponibles de los aprendices, cuya edición corrió a cargo de Jiménez Catalán (2014). Con él se pretende tender un puente entre dos tradiciones de investigación en el ámbito de la disponibilidad: la inglesa y la española. Contiene una serie de estudios originales escritos por investigadores e investigadoras de la disponibilidad léxica dentro del ámbito hispanohablante, por un lado, y otros trabajos elaborados desde el terreno de la investigación del vocabulario del inglés; dos comunidades de trabajo con intereses compartidos, pero que con escasa frecuencia coinciden en los mismos foros de investigación, mucho menos en el espacio de un libro. Como se puede observar en este volumen, la investigación sobre vocabulario ha seguido diferentes vías en la lingüística aplicada del inglés y la del español. En la primera, los intereses principales se han centrado a lo largo de los últimos veinte años en desarrollar listas de frecuencia a partir de corpus, en elaborar diccionarios, o en diseñar pruebas de vocabulario para aprendices. En contraste, dentro de la tradición de lingüística aplicada del español la atención principal se ha destinado a la creación de un diccionario panhispánico nutrido a partir de los léxicos disponibles de los hablantes de diferentes zonas y países hispanohablantes (Jiménez Catalán 2014: 5-9).

En este libro, además de un capítulo introductorio de López Morales en el que traza el panorama del desarrollo de los estudios de disponibilidad léxica (y que presenta la historia, la terminología, los conceptos fundamentales y las

fórmulas que aparecerán después en el resto de capítulos), podemos conocer diferentes estudios de disponibilidad léxica aplicados a aprendices de inglés y español como lengua extranjera. La primera parte del tomo se centra en los del inglés, mientras que la última analiza los de español. Los capítulos incluidos en las dos partes contienen estudios empíricos. El interés común es la disponibilidad léxica en lenguas segundas o lenguas extranjeras, pero el centro de atención se fija más en los aprendices que en la enseñanza o en los materiales de trabajo. Cada parte comprende investigaciones abordadas desde diferentes puntos de vista, como la sociolingüística, la psicología cognitiva, los estudios de corpus lingüísticos o los de frecuencia léxica. Ambas partes incluyen estudios centrados en aprendices de educación primaria, secundaria y enseñanza superior, principalmente en España, pero también en Chile, Polonia y Eslovenia. Aunque se diferencian porque las lenguas maternas y las lenguas de destino observadas difieren: inglés como L1 y L2 en la Parte I (capítulos 2 a 6) y español como L1 y L2 en la Parte II (capítulos 7-10). La información que contiene este volumen puede resultar útil para el profesorado de lenguas, además de para la planificación y el diseño de actividades de vocabulario. La información sobre las palabras que emplean los aprendices de diferentes edades, sexos, niveles, lenguas meta, como inglés o español, puede proporcionar a los investigadores datos preciosos para investigar la naturaleza y organización del léxico de los aprendices de lenguas.

Existen dos trabajos recientes, López González (2014) y Paredes García (2015) que se detienen en estudiar la relación entre la disponibilidad léxica y la enseñanza de ELE, aprovechando para realizar un estado de la cuestión, por lo que sirven como modo de aproximación al tema. Así, a través de la lectura de sus páginas, se puede tener conocimiento de las cuestiones básicas que relacionan a ambas líneas de trabajo.

6.3.3.1. Los centros de interés: ampliación y nuevas propuestas

Fruto de la difusión y coordinación entre las diferentes investigaciones llevadas a cabo desde los años 90, sobre todo bajo el paraguas del PPHDL, en los estudios de disponibilidad léxica desde hace años se han establecido unas pautas

metodológicas muy precisas y consensuadas para la recogida de datos y su posterior tratamiento. Como hemos comentado, esta coherencia en el proceder permite sumar y comparar los resultados de unos y otros trabajos. Estos motivos llevaron a las primeras propuestas que se realizaron con informantes no nativos a seguir muy de cerca estos criterios previamente fijados, tanto en territorio español como extranjero. Se trata de los trabajos de Carcedo González (1998a, 2000) y de Samper Hernández (2002). No obstante, la comunidad científica en torno al léxico disponible es consciente de que los centros de interés tradicionales “presentan serios problemas teóricos que tienen que ver con la forma en que se enuncian, con su número, con su amplitud semántica y con los resultados cuantitativos que proporcionan” (Tomé Cornejo 2010: 186, Paredes García 2014), pero también es cierto que la homogeneidad del método compensa sobradamente las dificultades derivadas de su justificación (Moreno Fernández 2012: 53). En este apartado veremos esas modificaciones que fueron surgiendo en los trabajos de disponibilidad léxica que se elaboraron con estudiantes de español para extranjeros.

En un inicio, las modificaciones metodológicas llevadas a cabo pasaron por la mera adecuación al nuevo tipo de informantes en cuanto a las variables sociales que se tomaron en cuenta. Así, a las ya clásicas que se solían considerar con nativos, como el sexo, el tipo de centro escolar (público o privado), la localización del centro (rural o urbano) y el estrato sociocultural (y en ocasiones, cuando la prueba se realiza sobre comunidades bilingües, se recogen también los datos relativos a la lengua materna y al nivel de uso), se unen las de lengua materna, nivel de español (A, B, C), conocimiento de otra(s) lengua(s) románica(s) u otras lenguas extranjeras, frecuencia con que los encuestados utilizan el español y, en el caso de personas inmigrantes, cuestiones como las laborales o el tipo de centro donde han adquirido su formación de español. Como hemos ido explicando antes, algunas de estas variables se revelan importantes para explicar la variación en los resultados (Mas Álvarez & Santos Palmou 2014a).

A la hora de tratar los datos obtenidos de las encuestas, los criterios de edición seguidos también fueron los establecidos durante años por la comunidad científica especialista en disponibilidad léxica, pues se parte de las normas

fijadas por Samper Padilla (1998a). Sin embargo, se tiene en cuenta el tipo de informantes y se aplican con una mayor flexibilidad en la aceptación en las respuestas. Así lo expone Samper Hernández (2002), quien se detiene en mostrar los errores que son propios de los estudiantes de español como lengua extranjera y comenta que se aceptan si se puede rescatar la palabra que se pretende usar realmente; con los extranjerismos se rechazan los que no modifican la lengua de origen, pero se cuentan los aceptados por la RAE y los más usuales. Por último, este artículo subraya la amplitud de las relaciones asociativas en estos casos, una característica que debe tomarse en especial consideración en proyectos con estudiantes de ELE, pues cobra una dimensión especial al tratarse de informantes de otras culturas y con un menor conocimiento del idioma.

Después de obtener los primeros análisis con hablantes y estudiantes de español y ver qué variables influían más en los resultados, no se tardó en observar que las adaptaciones debían ir más allá, sobre todo las que afectaban a los centros de interés, que en un inicio fueron los 16 fijados ya por los investigadores franceses. El hecho de que se continúe trabajando con los mismos temas después de tantos años es una de las asignaturas pendientes de este campo de investigación, pues a la vista está que hay centros de interés que funcionan mejor y otros peor, que no se cubren todas las facetas de la vida ni se plasma de forma completa la competencia léxica de los hablantes, en parte debido al hecho de que se obtienen muchos sustantivos como respuesta y pocas palabras de otras clases. Es este un punto de inflexión en el que se encuentran los estudios sobre disponibilidad léxica en general, pues se trabaja para ofrecer mejoras que pasan por perfeccionar los centros de interés. En los diferentes trabajos se han ido ofreciendo soluciones parciales al problema, que la mayoría de las veces han supuesto añadir centros de interés de temas básicos y productivos para extraer vocabulario o categorías gramaticales diferentes, como “Los colores”, “Acciones cotidianas”, “España”, “Adjetivos”, “Familia”, “Universidad”, “La salud”, “Sensaciones y sentimientos”, etc. En otras ocasiones, ya lo hemos comentado, se han adaptado al entorno donde se desarrollaba la investigación, añadiendo centros de interés concretos de esos lugares, como algunos relacionados con el mar o con tareas del campo.

No obstante, en los estudios aplicados al entorno del ELE estas adaptaciones han sido más radicales y, después de observar los centros de interés más y menos productivos en los primeros trabajos, la mayor parte de los siguientes han eliminado los que obtenían menos respuestas (como “Trabajos del campo y de jardín” o “Iluminación, calefacción y ventilación”) para seleccionar solo algunos de los que obtenían más (como “El cuerpo humano”, “La ropa”, “Alimentos y bebidas”, “La cocina y sus utensilios”, “La ciudad, o “Juegos y distracciones”). En otros casos se han añadido, adaptado o fusionado algunos o se ha trabajado con centros de interés de acuerdo con las nociones específicas del *PCIC*. Así por ejemplo, Sánchez Gómez (2003) usa 12 centros de interés, escogiendo aquellos que se suponían más productivos según los estudios previos, de manera que eliminaron “Los muebles de la casa”, “Objetos colocados en la mesa para la comida”, “La cocina y sus utensilios”, “Iluminación, calefacción y medios de airear un recinto”, “Trabajos del campo y del jardín”; en el caso de “La escuela: muebles y materiales” pasa a denominarlo “La educación”, e incluye el de “La salud”. Otra propuesta que va en esta línea es la tesis de Sánchez-Saus (2011), en la que manifiesta una intención de alcanzar la mayor parte del léxico que los alumnos conocen o deberían conocer, según el *MCERL*, manteniendo al mismo tiempo los suficientes puntos en común con otras investigaciones. Según explica esta autora, no todos los centros de interés clásicos resultan de mucho interés cuando se trabaja con informantes extranjeros que no dominan el español; en algunos casos, debido a la escasa productividad del tema, en otros, a que son excesivamente específicos, mientras que se dejan de lado temas básicos que aparecen en cualquier listado de contenidos léxicos para los primeros niveles de aprendizaje de ELE. Por ello, se basó en la lista de temas de vocabulario que ha de conocer el alumnado que alcance el nivel A1 del *MCERL*. Algunos centros de interés que añade son “Ocio y tiempo libre”, “Tiempo meteorológico y clima”, “Acciones y actividades habituales”, “Aspecto físico y carácter”, “La familia”, “Viajes y vacaciones” o “El dinero”.

Las adaptaciones de los centros de interés al recoger y analizar el léxico de los nativos y los no nativos constituyen un tema central que ha dado pie a diversas investigaciones. Algunos trabajos, como el de Medina Arejita (2009), parten del *PCIC* para sus encuestas. El debate permanece abierto, como muestra

la polémica surgida en el número 16 (2014) de la *Revista Nebrija de Lingüística Aplicada a la Enseñanza de las Lenguas* entre uno de sus artículos de referencia (González Fernández 2014) y varios en la sección de discusión y comentarios (Paredes García 2014; Bartol Hernández 2014). Las cuestiones básicas que tratan giran en torno a la idoneidad de los centros de interés en ELE y las posibles mejoras, como eliminar o modificar algunos y tomar como referencia las nociones que fijan el *PCIC* y el *MCERL*.

Más en concreto, el primero de ellos (González Fernández 2014) tiene como objetivo analizar si los centros de interés clásicos son los más idóneos para estudiantes de español como lengua extranjera. Para ello, realiza un estudio de disponibilidad léxica con estudiantes turcos y, a partir de los resultados obtenidos, examina la productividad en cada uno de los centros de interés, al mismo tiempo que revisa la bibliografía sobre la disponibilidad léxica aplicada a la enseñanza de español como lengua extranjera. Por su parte, Paredes García (2014, 2015) reflexiona sobre la misma cuestión haciendo una revisión bibliográfica, y propone que en las nuevas investigaciones se adopten criterios que permitan, por una parte, mantener la continuidad con los estudios anteriores y, por otra, incluir nuevos temas o estímulos que permitan el conocimiento de otras parcelas léxicas. Algunas propuestas en esta línea, como la de Bombarelli (2005), Bartol Hernández (2010) o Sánchez-Sauz (2011), abogan por una mayor consonancia con las indicaciones del *MCERL* para la enseñanza del léxico. Así pues, podrían usarse como referencia las “nociones específicas” que se establecen en el *PCIC*, como apuntan los diferentes obras comentadas.

Pero, sin duda, la investigación que más ha sentado las bases para avanzar en este tema ha sido una tesis sobre cuestiones metodológicas relacionadas con el procesamiento y la aplicación del léxico disponible a la enseñanza de ELE, a la que ya nos hemos referido (Tomé Cornejo 2015) (*vid. supra* § 3.4.2.1.). En ella, y tras la realización de pruebas con estudiantes extranjeros, se establecen algunas pautas a seguir en este tipo de investigaciones, que hasta ahora no estaban totalmente asumidas, como usar los diccionarios de léxico disponible de nativos para seleccionar el léxico que se enseña en ELE o emplear los que se obtienen con estudiantes extranjeros para evaluación o diagnóstico de su conocimiento de la lengua. Además, la autora aborda cuestiones psicolingüísticas y factores

cognitivos que a la hora de estudiar el funcionamiento de los centros de interés resultan de ayuda. Así, podemos conocer mejor cómo obtener diferentes unidades léxicas o cómo evitar los efectos del *priming* semántico, por ejemplo. No obstante, la gran aportación de esta tesis de doctorado es su propuesta de mejora y ampliación de la nómina de centros de interés. Para ello, recurre a nombrar algunos como esquemas de conocimiento o *scripts* (por ejemplo el CI 39 “En un juicio”). Su propuesta, como ya hemos visto más arriba, cuenta con 47 centros de interés que reparte en cuatro cuestionarios, cada uno de ellos con dos modalidades. Recordemos que para la distribución de los centros de interés en las distintas pruebas se ha procurado que los estímulos relacionados semánticamente no coincidan en el mismo cuestionario (o bien que queden adecuadamente separados entre sí) y que los que tienen estructura de *script* se presenten conjuntamente. Estos 47 centros de interés están diseñados para cubrir las 20 nociones específicas del *PCIC* y sus subapartados. Como sabemos, se propone que se distribuyan en dos cuestionarios, aunque cada cuestionario presenta dos modalidades para realizar dos ordenaciones distintas de centros de interés (evitando que el cansancio de los informantes afecte siempre a los mismos campos nocionales) y para poder invertir el orden de los estímulos formados por más de una categoría semántica y contrarrestar los efectos del *priming* semántico. Con estos 47 centros se pretenden alcanzar una considerable amplitud semántica, que favorezca la evocación de múltiples subcategorías a la vez que evita la excesiva multiplicación de los estímulos temáticos. Para ello, a veces se recurre a unir bajo una misma etiqueta categorías semánticas próximas y a modificar el tipo de encuesta. Así, mientras en las pruebas de disponibilidad tradicionales los encuestados eran los mismos, para poder llevar a cabo estas nuevas pruebas la autora propone realizar cada uno de los dos cuestionarios con diferentes informantes. Por otro lado, también cambió los tiempos y la edad de los participantes, apostando por 3 minutos, en lugar de los 2 habituales, y por estudiantes universitarios de primer curso en lugar de alumnado del último año de bachillerato.

6.3.3.2 Propuestas de selección del vocabulario

Como vemos por la amplia bibliografía que lleva asociada, la disponibilidad léxica ha sido campo de trabajo fructífero y recientemente se ha detenido en probar su utilidad en relación al alumnado extranjero de español. No obstante, no ha habido un trabajo que se haya detenido en profundidad en seleccionar el vocabulario disponible ni en unirlo con el básico desde que Rojo Sastre presentó en 1979 su tesis doctoral *Bases lexicológicas en una metodología audiovisual de la lengua española para extranjeros*. En ella, da cuenta por extenso de las bases teóricas, métodos de selección léxica y resultados pedagógicos que le habían servido para los anteriores tomos. Según su estudio, el “español fundamental hablado”, como él lo denominó, estaba formado por 1.560 palabras, teniendo en cuenta las frecuentes y las disponibles. Este extenso trabajo sobre léxico fundamental fue la base para 4 manuales denominados *Vida y diálogos de España. 1er grado. Método audiovisual* (Rojo Sastre, Rivenc & Ferrer 1968-1969), publicados en París por Didier y en Nueva York por Chilton Books. Pero a la vista salta, por las fechas de ambos trabajos, que las metodología de la disponibilidad léxica y de los corpus de frecuencia requieren hoy en día una nueva aproximación. La nómina de voces que así lo señalan es numerosa, pero, como decimos, falta un trabajo que se detenga en ello. Quizá el más completo de entre todas estas visiones sea el de Izquierdo Gil (2005, 2016), que trata la selección del léxico en la enseñanza del español como lengua extranjera en Francia. A él nos hemos referido profusamente en capítulos anteriores de esta tesis, pues explora las posibilidades de los diferentes métodos de selección del vocabulario, si bien la suya es una visión concentrada en el contexto francés, con lo que se dirige a estudiantes de español con francés como habla materna y en el sistema educativo de ese país. Destaca, por ello, como modelo a seguir un análisis contrastivo, estudiando las diferencias, semejanzas, correspondencias, interferencias, cognados, parógrafos y homógrafos. Se suma, además, una nómina de unidades léxicas y métodos de enseñanza para el contexto en el que ubica su docencia. Con todo, su propuesta no solo es práctica, sino que ofrece también una gran base teórica e indica un camino a seguir en cuanto al uso del léxico disponible y el frecuente.

La tesis de Izquierdo Gil es una de las pocas investigaciones amplias y completas sobre este tema, como decíamos. Por el contrario, en relación a la disponibilidad léxica y la enseñanza del español a extranjeros, las aproximaciones que más abundan son las que tratan la cuestión de modo general y sin profundizar, reivindicando el uso de la disponibilidad para llevar a cabo la selección léxica, pero sin ir mucho más allá⁵⁷. De entre estas aproximaciones, algunas son más teóricas y breves, como la de Rodríguez Muñoz y Muñoz Hernández (2009), que puede servir de introducción al tema, o bien son un resumen general con algún tipo de cotejo entre el vocabulario que se ofrece en el *MCERL*, en el *PCIC* y el que se obtiene de los centros de interés, como las de Bartol Hernández (2010) y Paredes García (2015). Hay ejemplos de propuestas prácticas, como la de Gómez Devís (2005), en la que explora el aprovechamiento del léxico disponible para la enseñanza-aprendizaje del vocabulario, pero no especifica cuáles son los destinatarios en concreto. Lo que muestra en este breve trabajo es una propuesta didáctica para el aula de lengua española con las unidades léxicas del cuerpo humano, tanto nombres, como adjetivos, frases, etc. Para llevar a cabo la selección, parte del estudio sobre léxico disponible del que es coautora (Gómez Molina & Gómez Devís 2004: 297) y escoge las 15 primeras palabras de la lista del centro de interés “Partes del cuerpo”.

Otra cuestión puntual que se ha abordado más recientemente gira en torno a los criterios de selección, es decir, cómo escoger las palabras de entre las que se encuentran en los diccionarios de léxico disponible dentro de los centros de interés, y cómo graduarlas tomando como referencia los niveles de lengua. Esta cuestión, que a priori puede parecer simple, una vez que tenemos las unidades léxicas ordenadas por los índices de disponibilidad y por su aparición, se complica cuando se lleva a la práctica. Así por ejemplo, Fernández Leyva (2015) en un trabajo en el que compara el léxico de un manual de español con el disponible, apuesta por el índice de disponibilidad como parámetro y limita el vocabulario de un nivel inicial de español a las unidades que obtengan un valor

⁵⁷ Se puede acceder a un estado de la cuestión, así como a una revisión bibliográfica y metodológica sobre la selección léxica en Santos Palmou (2016).

igual o superior a 0,1. La elección no deja de ser subjetiva, pues no tenemos un motivo que indique que ese número de palabras son las que se resultan más útiles para un nivel inicial. Además, estas pueden variar, aunque ligeramente, si se trabaja con el diccionario de léxico disponible de un lugar o de otro, ya que en uno pueden quedar por encima de ese límite y en otros por debajo, o bien haber unidades diferentes, etc. El motivo de usar este punto de corte quizá radique en que ya se había usado para otros trabajos de cotejo entre sintopías. En el mencionado artículo de Paredes García (2015:17-19) se puede ver un resumen sobre los diferentes criterios y valores que se contemplan hasta ahora para esta selección:

- a) criterio de disponibilidad, a través del índice de disponibilidad;
- b) distribución social, mediante el porcentaje de aparición;
- c) peso relativo, con la frecuencia acumulada y
- d) prototipicidad.

Recientemente se ha hecho una propuesta de selección basada en la prototipicidad (Ávila Muñoz 2014, 2016). Este autor usa la teoría de los prototipos para explicar la estructura que forman las palabras en la mente de los hablantes, ya que se puede entender que las palabras más prototípicas son aquellas que debemos enseñar en un primer momento. Estas guardan relación directa con las más disponibles puesto que, partiendo de su naturaleza cognitiva, si accedemos a los mecanismos que utilizan los hablantes nativos para organizar el vocabulario, obtendremos el léxico necesario para la comunicación, según indica Ávila Muñoz. Este autor encuentra en un modelo, el de *Fuzzy Expected Value*, el medio adecuado para hacer los cálculos matemáticos necesarios que reflejen este grado de prototipicidad para la selección del léxico disponible. Así, partiendo de esta teoría sobre “conjuntos difusos”, establece diferentes grupos de palabras que organiza en 6 niveles, según sean más o menos compatibles con la idea que expresa el nombre del centro de interés. Según su hipótesis, cuanto más compatibles, más prototípicas. Por otra parte, se evita establecer cortes bruscos en la lista de palabras disponibles, ya que esta teoría permite graduar la transición entre los límites de los conjuntos. Sin embargo, ambos artículos resultan poco claros, pues no explicitan los cálculos, operaciones o *software* que

se usaron para establecer los niveles de compatibilidad. Y como los modelos matemáticos y los cálculos que se pueden aplicar siguiendo la teoría de los *fuzzy sets* son varios, la explicación resulta imprecisa y genérica como para poder seguir los pasos del trabajo que se han realizado, aun conociendo de antemano esta área de la lógica difusa.

El profesor Ávila Muñoz ya se había detenido en estudiar los criterios de selección del vocabulario disponible con anterioridad en colaboración con otros investigadores. Así, había recurrido a los principios de la teoría de los prototipos y a los “conjuntos difusos” para explicar la relación establecida entre las unidades léxicas en función de su grado de compatibilidad con respecto al conjunto con se mide (Ávila Muñoz & Sánchez Sáez 2010: 63-65). Por ejemplo, trabajando con el CI “Calefacción, iluminación y medios de airear un recinto” encuentran que de los 184 vocablos que componen el centro de interés, solo 47 pueden considerarse “muy compatibles” con el conjunto al que representan. En otro trabajo se sirve del grado de compatibilidad semántica para establecer 4 tipos de asociaciones en las listas de disponibilidad: 1) elementos nucleares, 2) asociaciones denotacionales, 3) derivaciones y 4) asociaciones individuales (Ávila Muñoz & Villena Ponsoda 2010: 183-84).

6.3.3.3 Propuestas para la enseñanza-aprendizaje del español para fines específicos (EpFE) desde la disponibilidad léxica

Una primera experiencia en rescatar el lenguaje técnico o específico fue la que llevaron a cabo Echeverría, Urzúa y Sáez (2006: 59-76). Aunque sin afán de aplicarla al ámbito de EpFE, usaron la metodología de la disponibilidad léxica aplicada a una lengua especializada, con unos centros de interés específicos creados *ad hoc* e informantes nativos conocedores de una jerga concreta. La terminología especializada de los diferentes ámbitos del saber se vincula a esos vocablos que el hablante no actualiza “a menos que se necesiten para comunicar una información muy específica” (López Morales 1996: 245).

Para la publicación de su artículo titulado “Disponibilidad léxica matemática. Análisis cuantitativo y cualitativo” los autores crearon 6 centros de interés específicos (“Cálculo”, “Algebra”, “Estadística”, “Física”, “Ecuación y

Geometría”) y los pasaron a 85 sujetos, de los cuales 71 formaban parte del alumnado y 14 del profesorado de la carrera Ingeniería Civil Matemática. El objetivo fue conocer cuánto y cuál es el léxico disponible que poseían docentes y discentes en el área de las matemáticas. Igualmente les interesaba la existencia de algún tipo de crecimiento en el léxico conocido por los alumnos a medida que aumentan sus años de estudio. Aunque su investigación siguió la metodología de la disponibilidad léxica, se alejó de sus parámetros tradicionales en lo que respecta a los centros de interés, pues crearon centros llamados “específicos” para extraer el léxico disponible del área en cuestión: la matemática. Así inauguran lo que tanto este artículo como más tarde Navarro Marrero (2009 y 2011) denominaron “disponibilidad léxica específica”. Según explican sus autores en el artículo, los resultados mostraron que tanto los análisis cuantitativos tradicionales como los biplots descriptivos apoyaron las hipótesis de trabajo⁵⁸.

De este modo, una línea interesante, aunque todavía poco explotada en la aplicación de la metodología de la disponibilidad a la docencia de ELE, es la que se lleva a cabo en el ámbito de la enseñanza del español con fines específicos (EpFE), teniendo en cuenta la relevancia del vocabulario técnico y específico. En el caso de las lenguas de especialidad, la mayor parte de los manuales de español para fines específicos de las diferentes especialidades con los que ya contamos, nos ofrecen el vocabulario que se suele denominar *general* y *semitécnico*, en detrimento del *técnico*. Las encuestas de léxico disponible pueden servir como un complemento para conseguir aquel tipo de términos que consideremos oportuno llevar al aula en cada caso, como podemos comprobar en las propuestas realizadas por Guerra Salas y Gómez Sánchez (2005), Navarro Marrero (2009, 2011) y Mesa Betancor (2012). En los tres casos se hicieron encuestas de disponibilidad léxica a estudiantes universitarios de una especialidad en concreto: Periodismo, Comunicación Audiovisual y Derecho en el primer caso, Diplomatura en Fisioterapia en el segundo y Español de los negocios (ENE), en una Facultad de Empresariales, en el tercero.

⁵⁸ EL biplot es un tipo de gráfico que permite posicionar grupos y vocablos a la vez, gracias al paquete estadístico InfoStat (2004).

El trabajo de Guerra Salas y Gómez Sánchez se puede considerar un ejemplo de la elaboración de un diccionario de léxico disponible con el vocabulario de “la prensa”, “la radio” y “la televisión” mediante encuestas a informantes nativos que se pueden considerar “modelo” (Guerra Salas & Gómez Sánchez 2004, 2005). En su caso hicieron 124 encuestas a universitarios de primer y cuarto curso de las carreras de Periodismo, Comunicación Audiovisual y Derecho (esta última facultad sirvió para contrastar los resultados con las anteriores), con tres centros de interés: “prensa”, “radio” y “televisión”. Los resultados que se obtuvieron muestran que el vocabulario resultante es de interés, pero que coincide con el general o semitécnico, muy probablemente por tratarse de estudiantes universitarios que aún no están en contacto con las situaciones prácticas en las que un determinado tecnolecto se usa. Se puede ver esto porque no se observaron grandes diferencias entre las respuestas del grupo de Derecho con respecto a los de Comunicación Audiovisual y Periodismo. Si lo que queremos es obtener las unidades léxicas y expresiones más específicas y técnicas, lo mejor es buscar grupos de profesionales que desarrollen un trabajo en un área concreta para hacer las encuestas de disponibilidad léxica.

Navarro Marrero (2009, 2011) se sirvió de la metodología de la “disponibilidad léxica específica” para llegar a la terminología especializada en el área de la Fisioterapia. Según esta investigadora explica, el léxico disponible específico es restrictivo cuantitativamente, al limitarse a un grupo restringido dentro de una comunidad de habla, y cualitativamente, en tanto que se vincula a un sector definido por la profesión o por la especialización adquirida a través del aprendizaje. Sus informantes eran más de 100 estudiantes universitarios de primero, segundo y tercer curso de la Diplomatura en Fisioterapia (Escuela Universitaria de Enfermería y Fisioterapia de la Universidad de Salamanca). Usó centros de interés específicos para el área de fisioterapia, aunque la mayoría de ellos son válidos también para otras áreas de las Ciencias de la Salud y solo el primero coincide con uno de los 16 tradicionales del PPHDL. Los centros usados fueron: 01. Cuerpo humano 02. Signos y síntomas 03. Pruebas de valoración 04. Patologías 05. Diagnóstico 06. Agentes y medios terapéuticos 07. Técnicas y métodos 08. Tratamiento 09. Recursos materiales 10. Prevención 11. Rehabilitación 12. Campos de actuación terapéutica 13. Salud 14. Paciente 15.

Profesiones sanitarias.

Otra fórmula fue la seguida por Mesa Betancor (2012), cuyo trabajo contrasta el léxico disponible de un grupo de 37 alumnos de español de los negocios (ENE) en el módulo de la banca, con la terminología real de textos auténticos del mundo de la banca española actual y con el vocabulario específico que recogen los principales manuales de enseñanza de ENE del mercado. Esta investigadora trabajó con un grupo de 37 alumnos de 9 nacionalidades diferentes durante un cuatrimestre en la asignatura de español de los negocios de la Universidad de Alicante, a los que se les realizaron dos pruebas de léxico disponible sobre el tema “la banca”. A continuación contrasta los resultados de estas pruebas con dos corpus: uno de terminología real de la banca española y otro con el vocabulario que extrajo de los manuales de ENE más usuales y reconocidos. Los resultados demostraron que existe discrepancia entre los corpus descritos y el léxico disponible de los alumnos, y por otro lado entre los dos corpus elaborados. Detecta por ello una necesidad de mejora en los inventarios léxicos recogidos en los manuales actuales y en la adecuación de los niveles de adquisición de la lengua. Al mismo tiempo apuesta por otra forma de obtener las nociones específicas y expresiones más técnicas, a través de los corpus especializados. Además de los glosarios y vocabularios específicos que puedan existir de un área concreta que queramos impartir, acercarnos a ese contexto y extraer de los documentos más usuales en las situaciones comunicativas que se desenvuelven en él puede ser un ejercicio mucho más productivo. Esta autora, escoge un área muy concreta, determina qué situaciones comunicativas y eventos se producen en ella, selecciona los documentos más usuales y hace una elección de las unidades léxicas y expresiones. Si nuestro curso de español para fines específicos (EpFE) es de profesionales de un ámbito concreto, desde luego es un método altamente útil y puede ser muy productivo, ya que tendríamos textos reales y un inventario léxico con los que trabajar. Así pues, tener un banco de datos con textos reales clasificados por eventos dentro de cada especialidad, con un análisis del vocabulario empleado que esté etiquetado de alguna manera, nos permitiría también saber qué cantidad de unidades léxicas de especialidad se encuentran y si estas son de un tipo u otro (técnicas, semitécnicas, generales).

La disponibilidad léxica se ha usado también para otros propósitos de carácter cultural: estudiar los elementos léxicos de ciertos campos semánticos que se pueden considerar aspectos culturales propios de un lugar. Así, Wingeyer (2014) estudia los términos disponibles de una región de Argentina que guardan relación con los mitos populares, la ciudad, el campo, insultos y diversiones nocturnas. Su objetivo fue aportar información relevante para programaciones interculturales y su correspondiente elaboración de materiales en el campo de la enseñanza-aprendizaje de ELE. Los materiales los obtuvo en las regiones de Resistencia, Corrientes y Posadas. Para ello cuestiona aspectos tales como “¿Cómo insultamos?” y “¿Qué hacemos para divertirnos los fines de semana a la noche?”, etc. De esta manera se afrontó el estudio de la cultura de un lugar utilizando pruebas asociativas, aplicadas a un total de 1.500 estudiantes secundarios de los últimos cursos de estas ciudades argentinas en 2008 y 2012. La idea que subyace a este estudio es la enseñanza de la cultura en el enfoque comunicativo y se sirve de la disponibilidad para conocer los elementos y referencias culturales que caracterizan a una comunidad.

Para terminar con este panorama mencionaremos los estudios que se llevan a cabo con inmigrantes en un contexto de inmersión, ya que las características del proceso de enseñanza-aprendizaje y las variables a tener en cuenta suelen manifestar diferencias notables con las de los estudiantes de un ámbito universitario o por un período breve de tiempo. En este tipo de investigaciones destaca la elaborada por Fernández-Merino (2013), en la medida en que tiene en cuenta las especificidades de este alumnado, ya que ofrece una amplia información sobre la inmigración en la Comunidad de Castilla y León, sobre los tipos de centros en los que los extranjeros siguen la docencia, así como una caracterización de variables que se revelan importantes, por ejemplo la profesión de los informantes. La intención de este autor es evaluar la eficacia de su estudio con las modificaciones que ha hecho en las variables y sobre todo en los centros de interés. Su referente son los dos primeros estudios que fijaron las pautas (Carcedo González 1998a, 2000 y de Samper Hernández 2002), pero modifica los centros de interés en función de los que se revelaron más o menos productivos en los anteriores en número de vocablos obtenidos. Esto le permite realizar una comparación cualitativa con estudios de disponibilidad léxica

hechos en Salamanca y Finlandia con informantes extranjeros. Un primer apunte, luego de analizar los resultados, está en el promedio de respuestas, ya que en nativos suele rondar las 20 palabras por informante y centro de interés mientras que para extranjeros no suele llegar a 12, número que todavía se reduce en el caso de informantes inmigrantes.

6.3.3.4. La disponibilidad léxica como medio de evaluación y diagnóstico de la competencia léxica

Otra de las utilidades que se han encontrado a la disponibilidad léxica es como instrumento de evaluación de la competencia léxica de los hablantes, sobre todo de ELE. Una de las primeras propuestas en esta línea fue el programa informático *Vocabulario disponible*, creado por el chileno Echeverría y su equipo. Le permite al alumno someterse a un test para conocer su dominio léxico en diferentes centros de interés y, además, ofrece actividades para subsanar posibles carencias en este sentido⁵⁹. De alguna forma estos materiales sirven como herramienta para el (auto)aprendizaje del léxico.

López González también se detiene en estudiar la importancia de la disponibilidad para la evaluación del desarrollo de la competencia léxica, aplicada tanto a la L1 como a la L2. En un artículo analiza la disponibilidad léxica en español como lengua extranjera de una muestra de 241 alumnos del programa de Secciones Bilingües de español en Institutos de Enseñanza Secundaria y Bachillerato de seis ciudades polacas, a partir de 16 centros de interés (López González 2010). Para ello, se centra en la evolución de su competencia léxica, tomando como puntos de referencia lo que en ese país se llama 3º de Gimnasio (B1) y 3º de Liceo (B2+). Sigue las pautas metodológicas del Proyecto Panhispánico. Gracias a este estudio se especifican, cuantitativa y cualitativamente, los incrementos y retrocesos de los estudiantes en el aprendizaje léxico.

⁵⁹ Puede verse el funcionamiento del programa en:

<http://www.c5.cl/ieinvestiga/actas/tise99/html/papers/metacognicion>

Más recientemente, Hernández Muñoz (2015) observa a través de la disponibilidad léxica si existen diferencias entre los estudiantes preuniversitarios de acuerdo con la especialización académica que han recibido a lo largo de su educación en los centros de enseñanza media. Mediante las pruebas de léxico disponible de seis categorías semánticas de las tradicionales evalúa la competencia léxica de 97 alumnos preuniversitarios de la provincia de Cuenca. Como hemos dicho los escoge en función de la especialización académica que han tenido en educación secundaria y bachillerato (ciencias o letras). Los resultados obtenidos son interpretados a la luz de las teorías de desarrollo de la competencia léxica adulta y los factores sociales, estratégicos y metalingüísticos que en ella influyen. La hipótesis de partida de la autora es que la formación académica previa al acceso al mundo profesional es uno de los factores ligados a la experiencia que modifican el conocimiento léxico de los hablantes de una lengua. Parte, además, de la idea generalizada en la mayoría de las investigaciones de que el conocimiento léxico semántico acumulado durante la vida adulta permanece estable hasta edad avanzada, aunque con cambios de la velocidad de transmisión neuronal, puesto que las redes de palabras que conforman la memoria semántica de los hablantes están en constante evolución a lo largo de la vida. Estas redes beben de la percepción que tenemos del mundo que nos rodea, que no deja de transformarse con el fluir del tiempo. El proceso consiste en que la experiencia hace que se produzca un reajuste cognitivo, y esto afecta a nuestras redes de conceptos y asociaciones. Incluso se dan cambios en lo que tiene que ver con las categorías semánticas que ordenan el conocimiento y los términos que funcionan como prototipos de ellas (Hernández Muñoz 2015: 80-82). Por estos motivos, mejorar la competencia léxica supone el aumento de unidades léxicas, y además también el incremento en sus posibilidades de actualización gracias al fortalecimiento de las redes léxicas (Ferreira & Echeverría 2010). Con estas ideas de fondo, el artículo intenta ver en qué medida dos factores determinantes *a priori*, como la edad y la educación, influyen en la competencia léxica adulta. No obstante, la conclusión global es que el léxico disponible del estudiante preuniversitario no depende en gran medida de la variable especialización académica, pues no presenta diferencias significativas

en cuanto al número total de palabras producidas (Hernández Muñoz 2015: 93-4).

Con una aproximación panorámica y una revisión bibliográfica, Tomé Cornejo nos indica las posibilidades que puede ofrecer la disponibilidad léxica para la evaluación y diagnóstico de la competencia léxica (Tomé Cornejo 2015: 341-6). Así, explica que al aplicar estas pruebas a extranjeros tenemos tres posibilidades diferentes de uso, en función de cuál sea nuestra finalidad:

1. *pruebas de diagnóstico*: para ver en qué áreas temáticas presentan más carencias;
2. *pruebas de dominio*: si queremos evaluar el nivel de vocabulario;
3. *pruebas de aprovechamiento*: si pretendemos saber si el alumnado ha asimilado los contenidos explicados y los objetivos previstos.

Otra distinción que se lleva a cabo en el uso de la disponibilidad léxica como instrumento de evaluación de la competencia léxica es la que busca su utilidad para medir el tamaño del vocabulario. Esto consiste en medir el número de palabras que un hablante conoce y su profundidad, es decir, su habilidad para usarlas. En estos casos las pruebas de disponibilidad léxica son de naturaleza independiente, descontextualizada y mayoritariamente objetiva en su corrección. No obstante, su utilidad no es la misma en todos los casos, pues resulta complejo saber si los hablantes son capaces de usar un determinado término y todo lo que ello implica. Usar las palabras de manera adecuada y efectiva supone el conocimiento de su forma, su significado y su uso. Además, la competencia en el uso adecuado se adquiere gradualmente y a diferentes ritmos. Con todo, a nivel cualitativo, las respuestas de los informantes pueden servir para estudiar las relaciones entre los vocablos actualizados, como son los *clusters*, las sinonimia, antonimia, hiperónimos, etc. Así lo han sabido aprovechar algunos trabajos con hablantes nativos (Gallos Camacho 2001) y con extranjeros (Sánchez-Saus 2011).

Estas posibilidades de análisis que ofrece la disponibilidad léxica se han explotado también por otra vía, que es la que se enfoca en los aspectos cognitivos del léxico mental y de nuestra memoria semántica, su estructura y asociaciones, los prototipos léxicos, las familias de palabras, etc. Una de las primeras

aproximaciones de este tipo en la disponibilidad léxica corrió a cargo de Hernández Muñoz, quien llevó a cabo diversos estudios interdisciplinares en los que suma las visiones de la psicología y los aspectos cognitivos que ofrece el estudio del léxico disponible. Su primer trabajo al respecto estudia el léxico desde la lingüística y desde la psicología (2005) y posteriormente profundiza en los aspectos cognitivos del léxico disponible (Hernández Muñoz 2006; Hernández Muñoz, Izura & Ellis 2006). Una de sus conclusiones a las que llega es que la disponibilidad de las palabras está influida por su tipicidad, su familiaridad y la edad a la que se adquiere, de tal modo que las unidades más disponibles se corresponderían con las más típicas de cada categoría, las más familiares para un determinado hablante y las que más tempranamente haya adquirido. Por el contrario, la frecuencia, la longitud de las palabras y la imaginabilidad, son factores que no resultaron importantes para la activación del léxico disponible. Posteriormente, prosiguió con las categorías semánticas (Hernández Muñoz 2007, 2010; Hernández Muñoz & Izura 2010), así como los efectos del bilingüismo en el *priming* semántico (2014), y, más recientemente, el comentado estudio sobre la evaluación de la competencia léxica bebe igualmente de este afán psicolingüístico (2015).

Ferreira y Echeverría reivindican también el tratamiento de los aspectos cognitivos en este ámbito (2010). En su artículo abordan las relaciones semánticas de las palabras en las pruebas de disponibilidad léxica en estudiantes de inglés como lengua extranjera (EFL) en nivel avanzado en comparación con las de los hablantes nativos de inglés. Para exponer las relaciones semánticas resultantes recurrieron a grafos o redes semánticas, demostrando con ellos que el alumnado de inglés como lengua extranjera solamente distingue grandes categorías semánticas, mientras que los hablantes nativos son capaces de organizar las palabras en subcategorías con un alto grado de especificidad.

Un año más tarde Sánchez-Saus (2011) trabaja con 322 estudiantes de ELE en universidades de Andalucía y con 18 centros de interés. Tiene en cuenta las variables extralingüísticas: “sexo”, “nivel de español (A, B y C)”, “lengua materna”, “frecuencia de uso del español” y, para informantes de lengua materna no románica, “conocimiento de otras lenguas extranjeras románicas”. Además de un estudio cuantitativo, lleva a cabo una cala cualitativa con el programa

DispoGrafo para examinar las conexiones entre las palabras, los núcleos temáticos en cada centro de interés, las categorías gramaticales que abundan, las variaciones diatópica, diafásica y especializada, los cambios semánticos, transferencias léxicas, creaciones, etc.

Desde la lingüística cognitiva se ha estudiado a fondo y con ayuda de análisis matemáticos el funcionamiento de los centros de interés por parte de Ávila Muñoz y Sánchez Sáez. Sus conclusiones van en la línea de la teoría de los prototipos para explicar la organización de la memoria semántica. La generación de las listas de léxico disponible en cada centro de interés supone un objeto de estudio para observar la organización cognitiva a través de la manera en que se relacionan unos elementos léxicos con otros en cada conjunto. Ya hemos comentado que, para obtener la representación de los prototipos léxicos, los autores se sirven del marco matemático de los conjuntos difusos. Este mismo modelo matemático de los conjuntos difusos y el marco de la teoría de los prototipos proporciona a Ávila Muñoz (2016) unas buenas herramientas para llevar a cabo una propuesta de selección léxica del vocabulario disponible de cada centro de interés para aplicarlo a la enseñanza del español como lengua extranjera.

Capítulo 7. El vocabulario fundamental

7.1. Origen del vocabulario fundamental y su metodología: el Francés Fundamental

Si bien no se ha escrito demasiado sobre el concepto y la metodología del vocabulario fundamental, su nacimiento fue parejo al del léxico disponible. Cuando los autores franceses Gougenheim, Michéa, Rivenc, y Sauvageot hicieron su propuesta de “français élémentaire” en 1954, recurriendo a una metodología diferente para establecer el repertorio léxico que enseñar a los aprendices extranjeros de francés, estaban poniendo la primera piedra de lo que dos años más tarde definieron ya como “français fondamental”. En realidad, su propuesta para suplir las deficiencias que los listados de frecuencia ofrecían no pasaba por usar la disponibilidad léxica como método alternativo, sino que en las dos obras que vieron la luz se recurría a la unión de ambos métodos como fórmula de selección del vocabulario (Gougenheim, Michéa, Rivenc & Sauvageot 1954 y 1958). Así, aunque por una parte estaban desarrollando toda una línea de investigación de gran profusión en el ámbito hispánico gracias al Proyecto Panhispánico (PPHDL), en conjunto su apuesta era sumar a las ventajas del léxico frecuente las virtudes que se descubrieron en las encuestas de disponibilidad, pues ambos procedimientos reducían sus limitaciones si se complementaban. De tal modo que el tipo de unidades léxicas que se conseguía seleccionar mediante esta nueva vía eran justo las que se echaban en falta en los estudios de frecuencias, y viceversa. Por lo tanto, estos investigadores franceses vieron muy pronto el potencial de aquella unión que empezaron a denominar Francés Fundamental, y que más tarde se llevaría a cabo en otras lenguas, como el irlandés, alemán, portugués y español.

En cambio en la bibliografía sobre disponibilidad léxica no se ha profundizado demasiado en explicar o seguir los objetivos de estas obras pioneras, sino que se han priorizado las encuestas de disponibilidad y el análisis de sus listados bien por separado, bien con fines dialectales, sociolingüísticos, etc. En toda la bibliografía sobre disponibilidad léxica en español, que a día de hoy es numerosa, son pocas las ocasiones en las que se cita esta posibilidad de aplicar el léxico disponible junto con el léxico frecuente a la enseñanza, y no pasan de unas breves líneas en las que se comenta esta posibilidad y se define el concepto, recordando las palabras de López Morales al respecto (1995-6). Según este autor el léxico fundamental de una comunidad de habla está constituido por el léxico básico y el vocabulario disponible. Cuando se citan las obras que iniciaron los estudios de disponibilidad en Francia se suele comentar simplemente su objetivo (seleccionar el léxico para la didáctica del francés a los extranjeros) y su nueva forma de operar (las encuestas de disponibilidad léxica).

7.1.1. El vocabulario fundamental y la didáctica de lenguas extranjeras

El contexto de las primeras publicaciones sobre el francés fundamental nos ayuda también a entender que esta nueva línea de selección léxica nació dentro de un proyecto con miras más amplias y bien orquestado, cuyo objetivo no se limitaba a la búsqueda de una mayor eficacia a la hora de decidir qué vocabulario incluir en los manuales de lenguas extranjeras. Muy al contrario, debemos pensar que, tras la Segunda Guerra Mundial, el inglés fue ganando terreno al francés como lengua internacional. Así, la intención que se esconde bajo estas publicaciones es el ánimo de contrarrestar esa tendencia y mantener el prestigio del francés. Es por este motivo por lo que Francia emprende en la década de los años 50 del pasado siglo una reforma concienzuda de la enseñanza y la difusión de su lengua, que afecta tanto a la metodología como a los medios materiales. De hecho estas obras están auspiciadas por el Ministerio de Educación de Francia y dirigidas por el *Centre de Recherche et d'Étude pour la Diffusion du Français (C.R.E.D.I.F.)*. Por ejemplo, de este centro dependía la Comisión para el Francés Fundamental, que coordinaba las encuestas y tomaba las decisiones para el proceso de selección del vocabulario. Como primer paso,

un equipo de expertos adscritos al C.R.E.D.I.F. —entre los que figuran el lingüista Gougenheim y el pedagogo Rivenc— elabora una lista con las palabras más frecuentes y las más útiles (los nombres obtenidos con el nuevo método del léxico disponible) de la lengua francesa, así como una gradación gramatical. La referencia para ello era el *Basic English* (de 850 palabras), confeccionado por Ogden, que junto con otras obras en lengua inglesa y francesa de léxico frecuente es citado por estos autores. El equipo persigue no solo elaborar un vocabulario y gramáticas básicas, sino que se centra también en el francés oral, ya que el método educativo que iba a llevar a la práctica sus contenidos era el audiovisual.

Recordemos que el método audiovisual es un modelo didáctico concebido para la enseñanza de la lengua extranjera a principiantes que en ese momento empieza a estar en boga. Este método es una de las concreciones didácticas de la lingüística estructuralista, las teorías conductistas y el análisis contrastivo. Su éxito fue rotundo —incluso a nivel comercial—, y enseguida comienza a explotarse en televisión. La idea no tarda en rebasar las fronteras, en especial hacia países anglófonos, donde se elaboran programas semejantes para la enseñanza del inglés y de otras lenguas.

Sus orígenes están en el denominado método audio-oral, presentado por el profesor Robert Lado en 1964 en la obra *Language teaching. A scientific approach*. En Inglaterra deriva en el método situacional, que consideraba que se debe tener en cuenta el entorno situacional, y en territorio francés añade la importancia de la comprensión global del texto y potencia un mayor uso de apoyos visuales, de ahí que se denomine a la corriente francesa a este método estructuro-global-audio-visual. La razón para el nacimiento de todos ellos es la creencia, motivada por el estructuralismo, de que la lingüística debía limitarse al estudio del comportamiento verbal. Stern (1983: 158) resumió las características de estos métodos del siguiente modo: la lengua es lengua hablada, una lengua es lo que sus hablantes nativos dicen, no lo que alguien cree que deberían decir, las lenguas son diferentes, una lengua es un conjunto de hábitos, el objetivo es enseñar la lengua, no enseñar sobre la lengua.

Aunque sin desatender el lenguaje escrito, se da la prioridad a la lengua oral, por eso las primeras sesiones (unas 20 horas) del método se dedican al

lenguaje oral y a continuación se comienza el aprendizaje del escrito. El aprendizaje de una lengua extranjera se canaliza a través del oído (escuchando diálogos) y de la vista (observando la situación), por lo que se emplean combinadamente grabaciones de diálogos en soporte magnético (en la época inicial se usaba un magnetófono) e imágenes en *filminas* (las primeras diapositivas). Aunque desde la perspectiva actual estos medios audiovisuales nos resulten rudimentarios, los recursos materiales reciben en el momento una atención especial y se procura que la presentación resulte lo más atractiva posible. Howatt (1987: 16) definió el método como “el primer intento serio de construir una descripción pedagógica de una lengua extranjera, basándose en transcripciones de conversaciones habladas”. De las teorías conductistas el método hereda la repetición y la práctica mecánica de estructuras, con el fin de automatizarlas: las reglas gramaticales se aprenden de manera intuitiva. Pero, dada la naturaleza social del lenguaje, es imprescindible presentar las estructuras debidamente contextualizadas en los diálogos grabados y en el contexto social que se aprecia en las imágenes; todo ello constituye un paso decisivo hacia un uso comunicativo del lenguaje (Martín Peris *et al.* 2008).

Algunas de sus premisas todavía hoy son la base del método comunicativo, como la de no emplear la L1 en el aula, entre otras razones, por considerar que la contextualización mediante imágenes ya garantiza la comprensión. O, por ejemplo, organizar una unidad en torno a una serie de situaciones en las que se presentan los diálogos con la puesta en práctica de una serie de funciones lingüísticas (saludar, pedir información, etc.) y presentando el vocabulario necesario para desenvolverse en ellas.

Debido a que el aprendizaje de la lengua extranjera se inicia con el lenguaje oral, se dedica una atención especial a la pronunciación, incluidos el ritmo y la entonación, para en una fase ulterior trabajar el lenguaje escrito. Una lección típica de este método puede empezar con una *presentación* mediante la cinta magnética y la filmina. Luego el docente procede a la *explicación*, mediante demostraciones, preguntas y respuestas, por ejemplo. En la tercera fase, la de *fijación*, se va repitiendo el diálogo —en clase o en el laboratorio de idiomas— con el fin de memorizarlo, pero se procura no fragmentar los patrones rítmicos y entonativos. Vemos, pues, que se introduce el trabajo en el laboratorio de

idiomas. A continuación llega la fase de *explotación*, en la que la práctica es más libre, y se permite al alumnado la creación de sus propios diálogos —mediante preguntas y respuestas, juegos teatrales, etc.—, ya sea a partir de las mismas imágenes de las filminas, o bien en relación con su propia vida, su entorno familiar, social, etc.

Dos aportaciones del método audiovisual a la didáctica de la lengua extranjera son los medios audiovisuales (que favorecen el empleo de un lenguaje más natural y próximo a la realidad de los hablantes nativos) y la intención de suministrar a cada grupo meta un tipo de lengua acorde con sus características (un paso decisivo hacia lo que años más tarde se denominará análisis de necesidades).

Este es el contexto en el que nace en 1954 el *Français Élémentaire*, cuya versión revisada, *Français Fondamental*, se publica en 1958 y se toma como el inventario léxico-gramatical para la confección de los materiales didácticos audiovisuales con que culmina el proyecto. En *L'Élaboration du Français Fondamental (1^o Degré)* de 1964 se añade un Apéndice con muestras de toda una serie de manuales pedagógicos y publicaciones que derivaron de este trabajo: Davesne (1955), Mauger y Gougenheim (1955), Gougenheim (1958)⁶⁰, C.R.E.D.I.F (1958, 1962, 1963), Lagrave (1961), Beaumont (1963), Fourré (1963), Mauger (1963).

También se incluyen en este apéndice muestras de algunos capítulos de los diferentes libros y una breve descripción. Todos ellos se han elaborado teniendo como vocabulario de referencia el derivado de las dos primeras publicaciones del francés elemental y fundamental. Como vemos, estas no nacen solas, sino insertadas en todo un proyecto educativo para facilitar el aprendizaje del francés entre la población de las antiguas colonias francesas africanas. Hay por detrás, pues, un objetivo pedagógico y de expansión del idioma, que dio lugar a la creación y desarrollo de métodos de selección léxica y de enfoques y técnicas

⁶⁰ Un diccionario con 3.000 palabras, es decir, que a pesar de que parte del francés fundamental de 1956 se añaden casi 1.700 palabras, según los autores con el “mismo espíritu” que en el anterior, esto es, con base en los datos estadísticos de las palabras frecuentes y las palabras de las encuestas mediante centros de interés. La selección de los textos está hecha en este caso por Vander Beke. Para la definición de las palabras se usaron también las del francés fundamental.

de aprendizaje. Sin olvidar que esta intención de promoción cultural y de los valores de Francia es heredada de su pasado como gran potencia internacional e imperialista.

En este Apéndice del *L'Élaboration du Français Fondamental (1^{er} Degré)* se pueden observar fragmentos del libro del profesorado y algunos textos y actividades de varias lecciones del volumen para el alumnado. En el libro de los aprendices, además de textos, se recurre a tiras de imágenes con diálogos. Es digno de reseñar el hecho de que uno de los métodos sea un manual con solo imágenes y diálogos provenientes de Francia, que otro de los manuales sea un diccionario de imágenes con el nombre asociado y el hecho de que se cree además un método de aprendizaje audiovisual para niños. Esto se debe a que la creación de estos manuales va de la mano del método audiovisual, algo que también ocurrió, como veremos, en el caso del Español Fundamental.

7.1.2. El Francés Fundamental

La expresión “léxico fundamental” se acuñó en la obra de Gougenheim *et al.* (1958). En sus 300 páginas se incluyen 4 apartados, un primer capítulo dedicado a estudiar los diccionarios básicos anteriores y coetáneos, un segundo apartado que explica cómo se desarrolló el trabajo para llegar al léxico frecuente, un tercero dedicado a la disponibilidad y un último capítulo en el que se unen ambos resultados y se establecen el léxico y la gramática del Francés Fundamental. El punto de referencia para el primer capítulo son las obras del ámbito inglés, que por aquellos años intentaron seleccionar un vocabulario básico con fines didácticos, como el conocido Inglés Básico al que ya hemos aludido en varias ocasiones. En general, la tradición de las listas de frecuencia tuvo su germen en la lengua inglesa, que todavía hoy cuenta con diferentes corpus de frecuencia para elaborar diccionarios (tanto generales como diccionarios para aprendices de inglés), lecturas graduadas, listas de frecuencia para el estudio de la lengua en general o para determinar un registro específico, como el inglés académico, etc. Así, la mayor parte de las referencias de estos primeros apartados provienen del ámbito anglosajón, que suponen los inicios de la léxico-estadística, con autores como Odgen, West, Thorndike, etc. Se pasa a

continuación a describir las anteriores aportaciones de las listas de frecuencias y los léxicos fundamentales en el ámbito francófono, cuyas primeras obras fueron las de Hemmon (1924) y Vander Beke (1935). De ambas obras se hace un análisis profundo en cuanto a su composición, metodología y resultados, comparando los datos de uno y otro, si bien el estudio de la obra de Vander Beke es más completo y de mayor tamaño. Así mismo, se tratan otros recuentos inspirados en el trabajo de Vander Beke en el ámbito francés, como el de Tharp (1934), el de Verléé (1954) o el de Schlyter (1951). Se tratan también otros dos diccionarios básicos de francés, el que Aristizábal (1938) publicó en Bélgica con fines pedagógicos para la etapa escolar de primaria y el de Dottrens y Massarenti (1948), junto con los de otras lenguas, como el de García Hoz para el español, otro para el ruso, el esloveno, etc.

En el capítulo siguiente se explican detenidamente cuáles fueron los pasos seguidos para determinar el vocabulario frecuente. Hay una serie de características que se siguen en el trabajo del léxico fundamental del francés que se continuarán prácticamente igual en el resto de trabajos sobre otras lenguas, incluidos el español y el portugués, que más adelante analizaremos. Para empezar, se trabaja con un corpus extraído de conversaciones que han sido grabadas (en los primeros trabajos con magnetofón) y que luego se transcriben. Con esto se persigue analizar el lenguaje oral y más coloquial, por lo que se aprovecha para presentar algunas de las características y diferencias entre ambos registros. Para esta parte del trabajo se escogen como informantes (275 en total) no solo adultos, hombres y mujeres, sino también niños en edad escolar, y se intentan representar diferentes profesiones y procedencias geográficas. El resultado son 1.063 palabras de frecuencia igual o mayor a 20 apariciones, de las cuales no todas pasarán a la lista final del francés fundamental. Se generan así dos listas, una por orden de frecuencia y repartición (también llamada rango en otros estudios de frecuencias) en la que se da preferencia a la frecuencia para ordenar los elementos, y otra por orden alfabético. Las palabras más frecuentes son: *être, avoir, de, je, il, ce, la, pas, à, et, le, on, vous, un, ça*, etc. Como se puede observar en estos ejemplos, los resultados no difieren demasiado del resto de listados de frecuencias en cuanto al tipo de vocablos que se obtienen, pues

mayoritariamente son palabras gramaticales. Como la procedencia de los datos en este volumen son conversaciones orales transcritas, a diferencia de los recuentos del francés citados, se considera oportuno incluir una comparación entre los resultados del primero y los de los segundos, que se justifica como una comparación entre “lengua oral” y “lengua literaria”, por tratarse de corpus obtenidos de diferentes fuentes. Se aplican también técnicas que en ese momento eran de actualidad para medir el orden de frecuencia de las palabras en la lista, como era la ley de Zipf (formulada en la década de los 40 por este lingüista de la Universidad de Harvard), y posteriormente el uso de uno de los primeros ordenadores, el IBM modelo 7094, que fue lanzado al mercado en el año 1962 para aplicaciones científico-técnicas.

La tercera parte, dedicada a la disponibilidad, es quizás la que más ha trascendido a través de las referencias que se han hecho a esta obra por parte de los trabajos de disponibilidad en español. Las críticas que se exponen con respecto al tipo de palabras que se obtienen mediante los estudios de frecuencia y lo que se puede esperar de esta, así como el motivo de por qué crear encuestas de léxico disponible para complementarla, se repiten en la bibliografía de la disponibilidad léxica. Como ya hemos comentado en cierto detalle, en primer lugar se parte del estudio del tipo de palabras que ofrecen los listados de frecuencias, esto es, palabras gramaticales, verbos, adjetivos y nombres de carácter general; por otra parte se observa la ausencia de palabras muy concretas (*coche, metro, chaqueta*) que son poco frecuentes pero muy conocidas y básicas en el vocabulario de cualquier hablante nativo de francés. Si se considera solamente el índice de frecuencia para establecer un repertorio léxico, se observa que hay palabras muy conocidas y usadas por los hablantes que no aparecen o que sí lo hacen pero obtienen una frecuencia muy baja. Por ello, para estos investigadores se mostró evidente que las frecuencias no podían considerarse por sí solas los índices más adecuados para la elección de aquellos vocablos que debían constituir el léxico fundamental del francés. Por lo tanto, buscaron otro método también lo más objetivo y científico posible que complementara las listas de palabras frecuentes. Nace así el concepto de palabra disponible y su metodología, como una forma de corregir los vacíos de la

frecuencia. Se establece la conocida diferencia entre palabras frecuentes y palabras disponibles, unas y otras con sus propias características. Las primeras son atemáticas. Las segundas por el contrario forman parte de aquel léxico asociado a un tema concreto. Pero ambas resultan igualmente útiles. Los autores lo expresan así:

La suite de nos réflexions communes nous a amenés à opposer aux mots fréquents les mots disponibles. Tel est le nom que nous donnerons à ces mots d'une fréquence faible et peu stable, qui sont cependant des mots usuels et utiles. Nous les appelons ainsi parce que, quoiqu'ils ne soient pas souvent prononcés ou écrits effectivement (hormis le cas où l'écrit ou la conversation traitent d'un sujet déterminé), ils sont à notre disposition, quitte à nous échapper momentanément, ou même à disparaître sous l'effet de la maladie. [Gougenheim *et al.* 1964: 145]

Ya en la primera publicación de 1956, *L'élaboration du français élémentaire*, en la que crean un recuento de 1.475 palabras mediante listas de frecuencias con el objetivo de construir un vocabulario elemental para la enseñanza del francés a extranjeros, sus autores habían detectado la ausencia de términos cotidianos (*métro, lettre, autobus y timbre*), por lo que constataron la insuficiencia de un método basado únicamente en el índice de frecuencia para la selección léxica. Aunque la primera referencia a estas cuestiones se la debemos a Michéa, quien anteriormente había escrito sobre la noción de disponibilidad en un artículo de 1953, "Mots fréquents et mots disponibles", en el que expone los principios que regirán las investigaciones sobre disponibilidad. Este autor ya constataba que las palabras frecuentes, que denominó atemáticas, alcanzan una gran estabilidad estadística. Por el contrario, existe otro tipo de palabras muy inestables desde el punto de vista de la frecuencia, que llamó temáticas, que suelen ser sustantivos concretos. Este investigador notó que este tipo de palabras no se abarcaban con el aumento de las muestras introducidas, ya que estos términos escapaban a la estadística general basada en la frecuencia. Así, llegó a la conclusión de que las palabras estadísticamente más inestables estaban vinculadas a áreas temáticas o situaciones específicas y que encontraban su estabilidad si se utilizaba como punto de partida las situaciones mismas. Observó además la inmediatez de las unidades léxicas disponibles y destacó en su definición el carácter latente en la mente del hablante en cuanto se hace alusión a un tema. Por consiguiente, en la selección léxica debemos procurar ambos tipos

de unidades, ya que para hablar necesitamos expresar conceptos vinculados a temas y situaciones, a su vez, ligados a automatismos psicológicos. Ya sabemos que de ahí surgió la idea de trabajar con pruebas asociativas, que possibilitaban que los individuos actualizaran estas unidades léxicas a partir de determinados “centros de interés”, los cuales actuaban como estímulos verbales para los hablantes.

Esta distinción entre palabras frecuentes y palabras disponibles ha sido mantenida por la tradición hasta nuestros días. En realidad, buscar asociaciones mediante estos “centros de interés” suponía el uso de las técnicas asociativas ya utilizadas en la psicología cognitiva empírica. Estos principios fueron los que se siguieron para *L'élaboration du français fondamental* (Gougenheim *et al.* 1964: 152-153). Como se explica en su cuarto apartado, dedicado a la disponibilidad, pidieron a 904 adolescentes (de entre 9 a 12 años) de 4 provincias francesas que escribieran las primeras 20 palabras que se les ocurrieran de forma inmediata para cada uno de los 16 centros de interés propuestos —que, recordémoslo, se han mantenido casi intactos en la tradición española gracias al Proyecto Panhispánico—⁶¹. Gougenheim suponía que estos centros de interés representaban los sectores semánticos más universales, es decir, los más comúnmente compartidos por todos los hablantes, cualquiera que fuera su procedencia social, geográfica, etc. Evidentemente, al preguntar única y exclusivamente por 20 sustantivos estaban limitando la clase de palabras, el tipo de unidades léxicas y el número de respuestas que podían obtener. Con las encuestas se enviaban indicaciones que se debían seguir en los diferentes colegios, es decir, eran los propios centros escolares los que se encargaban de recoger los datos. Gracias a esta lista con 12 indicaciones sabemos cuáles eran en concreto algunos de los procedimientos seguidos, como que cada centro de interés tenía que ir en una hoja separada, que el tiempo de espera para cada uno de los centros de interés no debía ser superior al cuarto de hora, etc. Al igual que

⁶¹ Recordemos que estos 16 estímulos denominados CI eran: 01, Partes del cuerpo; 02, La Ropa; 03, Partes de la casa (sin muebles); 04, Los muebles de la casa; 05, Alimentos y bebidas; 06, Objetos colocados en la mesa para la comida; 07, La Cocina y sus utensilios; 08, La escuela: muebles y materiales; 09, Iluminación, calefacción y medios de airear un recinto; 10, La ciudad; 11, El Campo; 12, Medios de transporte; 13, Trabajos de campo y del jardín; 14, Los Animales; 15, Juegos y distracciones; 16, Profesiones y oficios.

se hace hoy en día, a cada encuestado se le pedían algunos datos además del nombre y el centro en el que estudiaba, como el sexo, la edad, el curso escolar, la profesión de los padres, o el tipo de estudiante que era cada encuestado, a ojos del docente. Con estos datos, se elaboran estudios sociológicos para analizar las posibles diferencias en las respuestas en función del sexo y el entorno rural o urbano, lo cual permitía posteriores estudios de geografía lingüística o sociolingüística, recurrentes en la época en las que se llevaron a cabo estos trabajos. En general las diferencias no son muy relevantes, aunque sí se muestran algunas preferencias de uso en función del medio. Las divergencias encontradas entre las respuestas de chicos y chicas son, según los autores, ligeramente más significativas en algunos centros de interés como La ropa o Los muebles de la casa.

El apartado que más interés despierta para nuestros objetivos es el cuarto, en el que se presenta cómo fue elaborado el Francés Fundamental, pensado para la enseñanza inicial de esta lengua a extranjeros. Una primera cuestión es que persigue no solo elaborar un recuento léxico, sino también una gramática básica. Para ambos objetivos había una comisión que analizaba los resultados y tomaba decisiones para añadir, ampliar o descartar elementos. Esta comisión dependía del *Centre de Recherche et d'Étude pour la Diffusion du Français*. Así, de los estudios de frecuencias realizados seleccionan un total de 805 palabras, estableciendo como corte una frecuencia igual o superior a 29 y que las palabras hubieran sido nombradas al menos en cinco textos (conversaciones con informantes). A esta lista de palabras frecuentes se le añaden los nombres concretos que han sido obtenidos mediante las encuestas de disponibilidad, alcanzando así un total de 1.475 palabras. No obstante, esta cantidad final no solamente es la unión de las palabras más frecuentes y las más disponibles, sino que la comisión encargada de llevar a cabo la selección final tomó la decisión de añadir otras de interés educativo, ya que este vocabulario iba a ser utilizado en el contexto del aula, y algunas otras de carácter general o universal, pues la intención era superar los intereses locales de las zonas francesas donde se habían hecho las encuestas. Recordemos que la intención de esta selección léxica era el aprendizaje del francés por parte de los inmigrantes

llegados a Francia y por las poblaciones de las antiguas colonias francesas. Otra característica que manifiestamente se persigue en la elaboración de esta lista de palabras es imprimirle un carácter conversacional. De ahí la metodología usada en la elaboración del corpus de frecuencia: conversaciones grabadas. El vocabulario fundamental pretende ser sobre todo concreto y práctico, no abstracto y literario, por eso se recurre a la lengua oral cotidiana, pues se pensaba que era esta la que poseía estas características. La presencia de interjecciones en el listado se ve como una muestra de ello. El resto de la estructura lexicológica del francés fundamental indica que la mayor parte de las palabras son sustantivos (46,9%), seguidos de los verbos (22,9%), más numerosos en comparación con el *Basic English*. Contiene también palabras gramaticales (17,10%), y llaman la atención los pocos adjetivos (6,6%) y adverbios (5,6%) que lo constituyen.

De las encuestas estadísticas se extrae también una serie de cuestiones gramaticales esenciales que se organizan en dos fichas, una propiamente gramatical (con siete secciones: expresiones de futuro y pasado, construcciones verbales para introducir el discurso directo e indirecto, construcciones condicionales, condicionales de presente y de pasado, subjuntivo, gerundios y participios de presente, pluscuamperfecto de indicativo) y otra dedicada a las interrogaciones, directas e indirectas.

Para medir la utilidad de la selección léxica y gramatical hecha, se intenta verificar su grado de comprensión con diferentes tipos de textos, es decir, algo como el porcentaje de cobertura textual que el francés fundamental permite obtener en cuatro situaciones comunicativas diferentes: una conversación, diferentes artículos de periódico, una novela y un texto científico. Por supuesto los porcentajes de comprensión varían mucho de un contexto a otro, pasando de porcentajes de entre 90% o 80% en el caso de las conversaciones, o alrededor de 60% y 70% en el de los artículos y las novelas, a unos porcentajes de entre 16,38% y 9,73% en el caso de un texto científico. No obstante, los objetivos que pretendían alcanzar con la creación de esta nómina léxica y gramatical eran las situaciones comunicativas cotidianas, por lo que las primeras cifras son las que deberían ser relevantes, y no tanto las de un texto de carácter científico.

7.2. El Alemán Fundamental

A raíz de la publicación del Francés Fundamental se llevaron a cabo otras tentativas con el mismo espíritu y métodos. En la tesis de Rojo Sastre (1979: 205-6) se da noticia de proyectos de este tipo anteriores al del español: el francés de Canadá, el irlandés, el alemán, algunas lenguas africanas y un proyecto posterior, el del portugués. En el tomo primero del Portugués Fundamental, en el que se publica el vocabulario resultante, se citan en su introducción proyectos en curso sobre el español, el alemán, el árabe y el wolof. Efectivamente, en un artículo de Rivenc (1967) se anunciaban y describían someramente las encuestas hechas en Canadá por el Comité Interuniversitario de las Universidades de Laval, Macgill y Montréal, las hechas por el Padre O'Huallachain para el irlandés y las que se llevaron a cabo por el *Centre de Linguistique Appliquée de l'Université de Dakar* sobre algunas lenguas habladas en África. El proyecto del equipo de investigadores portugueses, que tenía como mecenas a la Fundación Gulbenkian y como director a Lindley Cintra, empezó sus trabajos después de conocer los resultados franceses y españoles.

De todos ellos, solo hemos podido localizar los trabajos sobre el alemán y los de la Península Ibérica. Para el alemán, la obra resultante que conocemos fue el volumen de Pfeffer (1964). En sus 79 páginas cuenta con un vocabulario en orden alfabético, otro listado ordenado por clases de palabras y un tercero por frecuencia y origen. Los materiales usados fueron 400 entrevistas grabadas y 5.400 registros escritos de asociación espontánea, es decir, se recurrió a las mismas técnicas que se habían usado en Francia para lograr un léxico frecuente y un léxico disponible. En este caso se registraron 21 centros de interés, según indica Rojo Sastre. De los datos recogidos, cada conversación fue transformada mediante la técnica de cintas perforadas en un centro de informática, que también habían usado para el francés y posteriormente para el español. Se excluyeron los nombres propios, nombres de lugares y los gentilicios. Se usaron las técnicas que los avances informáticos del momento proporcionaban al análisis lingüístico. El resultado final consistió en un vocabulario de 1.269 palabras.

Del mismo modo que la lengua inglesa ha considerado siempre los corpus y las listas de frecuencias léxicas para la lexicografía y la enseñanza, la tradición didáctica alemana otorga también importancia a las listas de vocabulario, tanto para el aprendizaje de otras lenguas por parte de los alemanes como al enseñar su lengua a extranjeros. No obstante, no sabemos la repercusión que este *Basic (Spoken) German Word List* pudo tener en su época, pero sí sabemos que recientemente un manual de alemán para extranjeros recurría a él como fuente para ofrecer un vocabulario de varias páginas alemán-inglés e inglés-alemán como apéndice. Nos referimos a *Kaleidoskop* de Moeller, Adolph, Mabee y Berger, editado por Heinle, que en su octava edición de 2012 contenía un vocabulario bilingüe de aproximadamente 50 páginas con las palabras usadas en el cuerpo del volumen y marcaba con un asterisco las que estaban incluidas dentro de una lista básica de 1.200⁶².

7.3. El Español Fundamental

El primer Español Fundamental que conocemos se lo debemos a Rojo Sastre, que aun siendo de origen burgalés y licenciado por la Complutense de Madrid en Filosofía y Letras, tuvo acceso al proyecto del léxico fundamental del francés durante su formación en una diplomatura en el *C.A.V de L'École Normale Supérieure* de Saint Cloud de París. Así, en colaboración con uno de los autores del anterior estudio, Rivenc, y junto con Ferrer, saca a la luz una serie de obras para la enseñanza del español a extranjeros mediante el método audiovisual, siguiendo muy de cerca algunos de los manuales que hemos citado para el francés. Nos referimos a los cuatro volúmenes de *Vida y diálogos de España. 1er grado. Método audiovisual* (Rojo Sastre, Rivenc & Ferrer 1968-1969), publicados también en París por Didier, como muchos de los anteriores, y en Nueva York por Chilton Books. Los 4 volúmenes son: V.1 Libro del profesor, V. 2. Textos de lectura. Ortografía del español, preparación para el dictado, V. 3. Lista alfabética de formas empleadas, V. 4. Libro de imágenes. En la década de los 70 hay varias

⁶² Para determinarlas, recurría a varias listas de frecuencias: *A Frequency Dictionary of German* (Randall L. Jones & Erwin Tschirner), *Das Zertifikat Deutsch als Fremdsprache* (Deutscher Volkshochschul-Verband & Goethe-Institut), *Grundwortschatz Deutsch* (Heinz Oeler) y el *Grunddeutsch: Basic (Spoken) German Word List* de J. Alan Pfeffer.

ediciones, alguna ya a cargo del Ministerio Español de Educación y Ciencia, Comisaría de Extensión Cultural. Así, mientras que Rivenc puso de moda el método estructuro-global audiovisual en Francia, Rojo Sastre y Ferrer lo hicieron en España. Desde luego, en la década de los 60 y 70 el español no gozaba de la difusión y el interés como lengua objeto de aprendizaje por extranjeros que tiene hoy, pero, aun así, se trabajó bastante en nuestro país con estos manuales y todavía no hace mucho tiempo se podían encontrar en algunos laboratorios de lenguas de Asia. En una crítica a los libros de enseñanza del español más antiguos, Soler-Espiauba comenta que los textos están basados en la repetición de estructuras fuera de contexto, diálogos demasiado secos y solemnes, poco expresivos o espontáneos, a veces incluso estereotipados, sin apenas contenidos culturales que nos sitúen en territorio, ya no hispano, sino español, ya que las imágenes y los decorados son unidimensionales y modélicos. Todo ello lleva a esta investigadora a definirlo como un método “estoico” (Soler-Espiauba 1994: 483). Lo que la autora de estas críticas desconoce es el verdadero año de creación de estos manuales y el año de la primera edición, pues los sitúa en 1976, ocho años después de la fecha real. Y no debemos olvidar que son ocho años realmente significativos para las circunstancias que vivía España, pues la primera edición se publicó en Francia y Nueva York cuando Franco estaba aún con vida y la edición que cita Soler-Espiauba es de la España en proceso de transición democrática. Esto le impide tenerlo en consideración como aquello que realmente es: un manual pionero en la enseñanza de ELE que pone en práctica de modo notable la metodología didáctica que imperaba en los años 60. De todos los que esta autora cita, es el más antiguo, y como bien explica Rojo Sastre en su trabajo doctoral, pertenece a una época en la que todavía no había ningún departamento universitario ni profesorado especialista en la lingüística aplicada a la enseñanza de lenguas extranjeras. Es, por lo tanto, una publicación merecedora de respeto que abre un nuevo camino, el de la docencia de ELE, como así lo reconoce la difusión y distribución que tuvo en años posteriores. Y como también lo avala el hecho de que tuviese por detrás todo un proyecto de selección del vocabulario basado en métodos lexicométricos, lo cual ni tan siquiera ocurre en la gran parte de los manuales del siglo XXI.

Estas características son compartidas por los manuales de Rojo Sastre Rivenc y Ferrer, en la medida en que seguían los principios del método audiovisual y el estructuralismo, y son hijos de su época e intenciones. El propio Rojo Sastre comenta sobre la metodología en su tesis que se habían llevado a cabo pruebas de fundamentación entre los años 1971 y 1973 con estudiantes de 34 nacionalidades diferentes para medir la adecuación lectora entre los contenidos visuales de la imagen y su expresión lingüística. De ellas se dedujo que se debían buscar formas unívocas de expresión visual, perceptibles a nivel universal. Podemos, por lo tanto, criticar la rigidez y poca alegría del método, cuando, en realidad, sus autores eran más pragmáticos y primaban los resultados que obtenían con ellos.

Al igual que en Francia, la publicación de estas obras didácticas fue acompañada de un proyecto de selección léxica que pretendía objetividad y procedimientos empíricos. Así, diez años más tarde Rojo Sastre presenta los resultados en su tesis doctoral, defendida en la Facultad de Filosofía y Ciencias de la Educación de la Universidad Complutense de Madrid en 1979. Bajo el título *Bases lexicológicas en una metodología audiovisual de la lengua española para extranjeros*, se esconde todo un proyecto didáctico que incluye desde las bases teóricas, hasta los métodos de selección del léxico fundamental y los resultados pedagógicos. No parece casualidad que la dirección corriera a cargo de García Hoz, pedagogo y autor de uno de los listados de frecuencias con intenciones didácticas.

El trabajo de doctorado de Rojo Sastre se divide en tres partes. En la parte primera e introductoria, y como también se había hecho en la publicación francesa, el autor repasa las diferentes propuestas de recuentos léxicos y proyectos basados en la creación de una lengua básica para un aprendizaje rápido de extranjeros: Ogden, *Basic English*; West, *The new method English dictionary*; Thorndike, *Teacher's Word Book*; Hemmon, *A French Word Book*; Vander Beke, *French Word Book*; García Hoz, *Vocabulario usual, común y fundamental*; Rodríguez Bou, *Recuento de vocabulario español*; Juilland y Chang-Rodríguez, *Frequency Dictionary of Spanish Words*. Además, Rojo Sastre incluye un capítulo sobre los diferentes métodos didácticos de aprendizaje de lenguas

extranjeras, entre los que el autor considera el método gramatical, el método de lectura de textos, el de traducción, el método directo y el audio-oral.

La segunda parte se inicia con un capítulo para explicar las diferencias entre la lengua hablada y escrita, cuáles son las características de la primera y reivindicar la selección de esta frente a la segunda, pues es la lengua hablada la que se considera que nos garantiza muestras representativas de la “lengua real” y cotidiana. No olvidemos que los trabajos de vocabulario fundamental estudiados aquí se denominan vocabulario de la lengua hablada, y trabajan con corpus orales, con un método consistente en grabar conversaciones con informantes, transcribirlas y seleccionar de ellas las palabras más frecuentes. El propio autor comenta en la introducción de este trabajo que los resultados estadísticos lo llevaron a fijar un “español fundamental hablado” (Rojo Sastre 1979: 16) y se refiere a la publicación del *Français Fondamental* como “un hito en los recuentos de lengua hablada y las comunicaciones concomitantes presentadas al leer” (Rojo Sastre 1979: 18). La razón de que busquen la lengua oral es que en el método didáctico con el que trabajan se pretende enseñar a hablar a los aprendices, mientras que se trabaja la escritura en un segundo momento y supeditada a la lengua hablada. Esta metodología no es la usual en el resto de los listados de frecuencias, de los cuales solo algunos más recientes se fijan en el registro oral en su totalidad. La inclusión de muestras de lengua oral en los corpus de frecuencias es algo reciente, aunque los criterios y métodos que se siguen en la actualidad son diferentes, pues se usan muestras orales espontáneas y no conversaciones con un entrevistador para generarlas. Nos referimos, por ejemplo, a conversaciones de radio, telefónicas, televisión, etc. que no son creadas *ex profeso* para ningún estudio. En esta parte es en la que se explica todo el trabajo realizado para la creación del Español Fundamental, desde la definición de los conceptos, hasta la metodología de trabajo y análisis, las muestras y los resultados, el trabajo de la Comisión que se creó a nivel internacional para tomar decisiones, etc.

La tercera parte es en la que desarrollan los fundamentos de la metodología estructuro-global audiovisual y se explica el sentido de los manuales didácticos *Vida y diálogos de España*, creados para desarrollar este

método con los contenidos léxicos y gramaticales que se habían obtenido en el proyecto del Español Fundamental. Se comenta cómo se llevó a la práctica el trabajo con estos manuales, desarrollando toda una programación didáctica, con planificación de unidades didácticas, cronología de los contenidos, de las sesiones en el aula, las actividades que se llevan a cabo, el trabajo con las imágenes, audios, vídeos, un grupo de alumnos meta, etc.

En el proyecto del Español Fundamental de Rojo Sastre, con la ayuda del profesor Rivenc, se realizaron numerosas conversaciones y encuestas para obtener 1.200.000 palabras gráficas, 800.000 de las conversaciones grabadas. Estas dieron origen al léxico frecuente, siguiendo la misma técnica de cinta perforada para su transcripción que en el proyecto del francés. Las otras 400.000 son de léxico disponible, esta vez con 25 centros de interés, más de los que se usaron en el Francés Fundamental. Otra diferencia es que en este caso se pedía a los informantes que escribieran en columnas separadas los 10 sustantivos, 5 adjetivos y 5 verbos más disponibles. El resultado son 2 volúmenes de 285 y 287 páginas en los que se incluyen varias listas de palabras: en las frecuentes se indica el índice de frecuencia y de distribución, y se pueden ver en un listado por orden decreciente de frecuencia así como en otra lista alfabética lematizada. Las palabras que selecciona para el léxico fundamental suman un total de 1.560 palabras diferentes, de las cuales 1.142 son del léxico básico y 664 son disponibles. Para los criterios de selección usará un límite de frecuencia de 50 o más, con la que llega a un total de 975 palabras frecuentes y 652 disponibles, aunque después incluye palabras entre la frecuencia 50 y 20, en el caso del léxico básico, y de una frecuencia de entre 50 y 25, en el del disponible. Este aumento surge de los consejos de la comisión que revisa los trabajos, cuyo criterio sirve de asesoramiento técnico. De la suma de estas dos listas de forma independiente se esperaría una cantidad de 1.806 palabras, pero debemos tener en cuenta que hay muchas palabras, que el autor denomina “zona mixta”, que aparecen en ambos tipos de léxico, por lo que la lista final mixta, contando palabras diferentes, está formada por esas 1.560, que forman el “español fundamental” para Rojo Sastre. La lista unitaria muestra estas palabras en dos columnas contiguas con la indicación de su frecuencia, de tal modo que se puede observar si se han

obtenido mediante la grabaciones del léxico frecuente, mediante las encuestas del léxico disponible o bien si aparecen en ambos listados.

Ofrece también las lista de palabras que se agrupan en series: las nacionalidades, los colores, los días de la semana, los meses del año, las series de numerales (cardinales y ordinales), los múltiplos, los divisores partitivos, las profesiones, la familia, los puntos cardinales, las estaciones del año, los pesos y medidas, así como los saludos y excusas.

Tanto la investigación sobre el Español Fundamental como los volúmenes del *Vida y diálogos de España* se llevaron a cabo con la ayuda del Ministerio de Educación y Ciencia. En el caso del proyecto del Español Fundamental, además, la Secretaría General Técnica del Ministerio organizó varias reuniones de especialistas para emitir sus observaciones y críticas, con el fin de garantizar la calidad del trabajo realizado. Así, en la cátedra del profesor Rafael Lapesa se organizó un coloquio que duró una semana. Se creó una Comisión Internacional, al igual que se había creado una Comisión en Francia y que más tarde también crearían los portugueses. En el caso del español, además de pertenecer a ella Lapesa, también contaban con varios profesores de la Universidad Complutense de Madrid como Manuel Alvar, Manuel Criado de Val, Emilio Lorenzo, un profesor de la City University of New York (Jerald R. Green), otros dos de la Universidad de Toulouse (Paul Merimée y Paul Rivenc), otro de la Sorbona (Bernard Pottier), un miembro de la RAE (Alonso Zamora Vicente) y el profesor Rojo Sastre en calidad de Co-director del Programa y Director de la Fundación Internacional Lengua Española. Al igual que en Francia y Portugal, esta comisión establecía las líneas a seguir, observaba las listas con los resultados, tomaba decisiones en cuanto a los criterios para seleccionar las palabras que descartaba y las que eran seleccionadas, etc.

7.4. El corpus del Portugués Fundamental

También con la ayuda de Rivenc, bajo la dirección de Luís Filipe Lindley Cintra, se llevó a cabo un proyecto de las mismas características para establecer un vocabulario y gramática del portugués europeo, que se elaboraron en el seno

del *Centro de Linguística da Universidade de Lisboa*. Este Centro, todavía hoy en activo, ha desarrollado numerosos proyectos y muchos corpus de diferente tipo, gracias a que cuenta no solo con la colaboración de profesorado universitario, sino también con varios investigadores y becarios que dependen directamente del Estado portugués⁶³. El Portugués Fundamental tiene una envergadura considerable y, aunque no se llegó a publicar la gramática fundamental, vieron la luz tres tomos diferentes: un primer tomo con el vocabulario del Portugués Fundamental, y otros dos tomos que recogen los resultados, las transcripciones de las conversaciones, los listados resultantes y la metodología de análisis y selección. El listado final es de 2.217 palabras, que constituyen el portugués fundamental. El proyecto echó a andar en los años 70, pero la publicación de estas obras no tuvo lugar hasta los años 84 y 87⁶⁴. Recordemos que Portugal, al igual que España, sufrió un proceso de cambio de una dictadura a una democracia a mediados de los años 70, por lo que los investigadores decidieron retomar el proyecto y elaborar nuevas encuestas, más ajustadas a la nueva realidad que se vivía en el país. El objetivo era crear un vocabulario del portugués más frecuente usado en situaciones de la vida cotidiana, pero recurriendo a un corpus de frecuencia y un corpus de disponibilidad. Al igual que en el caso del francés y el español, el primero es un corpus de lengua hablada, recogido en el periodo prerrevolución de entre 1970 y 1974, tanto del continente como de las islas. En total son 1.800 grabaciones en situaciones comunicativas que intentan ser espontáneas, sobre diversos temas cotidianos (organizados en torno a la vida personal, familiar, profesional y social), con informantes de diferentes edades y características sociológicas diversas. De estas casi 500 horas de grabaciones se seleccionaron 1.400 fragmentos, con un total de 700.000 unidades, cuyo resultado fueron 25.107 formas léxicas diferentes. La lista fue lematizada posteriormente y se organizan los vocablos según su índice de frecuencia. Para seleccionar de ella las palabras más frecuentes se establece como umbral una frecuencia igual o superior a 40, aunque se consideran también las que están dentro de la horquilla de frecuencia de entre 25 y 40, por contar

⁶³ Pueden verse algunos de sus proyectos en su página web: <http://www.clul.ul.pt/en/resources>

⁶⁴ Malaca Casteiro (1984), Bacelar do Nascimento, Garcia Marques & Segura da Cruz (1987).

con alta repartición entre los diferentes textos (lo que Rojo Sastre denominaba distribución).

Por su parte, el corpus de léxico disponible tuvo encuestas complementarias realizadas a inicios de los 80 con 3 nuevos centros de interés, además de las elaboradas antes de la revolución del 25 de abril con 27 centros de interés. El motivo fue incluir temas que con los acontecimientos sociales y políticos pasaron a ser comunes. Los autores justifican su metodología como forma de complementar el listado de frecuencia con aquel vocabulario temático que tiene menos probabilidades de aparecer en un corpus oral espontáneo, pero imprescindible en determinados dominios de referencia, aunque muchas veces se usen deícticos para sustituirlos. Para ello se enviaron encuestas a cada capital de distrito, distribuyendo los 30 centros de interés en una ficha separada, en la que se preguntaban sustantivos, adjetivos y verbos para cada tema. El montante obtenido fueron 481.800 palabras disponibles. Al igual que en el francés y en el español, no se limitaba el tiempo de la encuesta, sino el número de respuestas. En este caso se preguntaba por 15 sustantivos y 10 adjetivos y verbos, excepto en el último centro de interés, que se pedían solo 20 verbos. En las encuestas hechas en los 80 se aumentaron estas cifras a 30 nombres y 20 verbos en dos de los temas, y 20 nombres y verbos en el otro.

Los resultados se muestran en varias listas de palabras: la primera es una lista organizada por orden alfabético. Otra por clases de palabras, entre las que se incluyen sustantivos, adjetivos, verbos, pronombres personales, artículos, demostrativos, posesivos, indefinidos, numerales, relativos, interrogativos, exclamativos, preposiciones y locuciones prepositivas, conjunciones y locuciones coordinativas y conjunciones y locuciones subordinativas.

En el caso del portugués también se creó una comisión para dirigir y coordinar el proyecto. La primera iniciativa para llevarlo a cabo ya se había tomado en 1964 por la Fundación Gulbenkian y posteriormente se encargó a Lindley Cintra la elaboración. Por diferentes dificultades el trabajo no se llegó a concretar en ese momento, y se tuvo que esperar a 1968 para que se retomara la idea. Entonces, la responsabilidad había pasado al Ministerio de Educación Nacional a través del Instituto de Alta Cultura, y la realización pasó a manos del

Centro de Estudos Filológicos de Lisboa y el *Centro de Linguística Geral e Aplicada de Coimbra*. Finalmente el trabajo se dejó en manos del primer centro y bajo la tutela del profesor Lindley Cintra, que recibió también la colaboración de Rivenc y de Roche de la Universidad de Toulouse, pues el primero era coautor del proyecto del Español Fundamental. Así pues, el Portugués Fundamental es fruto de un trabajo colaborativo de diferentes organismos e investigadores, tanto en su desarrollo de los 70 como en la fase posterior de complementación que se llevó a cabo en los 80.

Los años en que estas publicaciones salieron a la luz permiten otros desarrollos técnicos y metodológicos, cuyo resultado se puede observar a través de la página web del *Centro de Linguística da Universidade de Lisboa*, donde se puede acceder gratuitamente a una nueva versión del corpus, que incluye los ficheros de audio en formato WAV, las transcripciones ortográficas correspondientes, bien alineadas con el sonido o bien por separado, y ficheros en TXT con las anotaciones morfosintácticas.

7.5. Los vocabularios fundamentales del francés, del español y del portugués: semejanzas y diferencias

Aunque los tres trabajos corrieron suertes diferentes, ya que del primero no se continuó su estela, del segundo solo se publicaron los manuales didácticos *Vida y diálogos de España* y del tercero se publicaron tres volúmenes en los años 80, todos ellos son proyectos homólogos y con un origen similar. El Francés Fundamental fue el primero y el modelo a seguir, pues se inició en la década de los 50. No obstante, su difusión y uso en los diferentes manuales fue a finales de esa década y en la siguiente, cuando también se trasladó la misma idea a otras lenguas. No tenemos información de cómo se elaboró el proyecto alemán, aunque por los años en los que se desarrolló deducimos que no hubo demasiados cambios sobre el trabajo francés, fuera de la ampliación a 21 centros de interés que sí conocemos a través de la aportación de Rojo Sastre.

Metodológicamente, el estudio español es casi idéntico al del francés, aunque de mayor envergadura en los corpus y en los centros de interés, estos últimos tanto en número como en palabras obtenidas. La técnica de la cinta

perforada también se mantuvo en ambos casos. Así, si comparamos la extensión, en francés se trabajó con un inventario de palabras frecuentes de 312.133, en el español con 800.000 y en el portugués con 700.00, y estos números no solo constituyen un aumento cuantitativo sino también cualitativo, pues se elaboraron sobre bases más sólidas y rigurosas. Para el español y el portugués se definieron los datos estadísticos de cada entrevista, así como los índices de medición de frecuencia y repartición o distribución. En el caso del portugués, la automatización del tratamiento ganó con los medios informáticos y tecnológicos con los que se contaba en ese momento. En el inventario de palabras disponibles, el aumento de centros de interés y de categorías gramaticales hace que el portugués recoja 481.800, frente a las 23.680 del francés.

Una diferencia cualitativa muy significativa que se observa en la lista del Portugués Fundamental es la inclusión de locuciones gramaticales (prepositivas, conjuntivas, pronominales y adverbiales) y lo que se denomina “grupos de palavras”, que no son sino colocaciones y locuciones léxicas. Las primeras se buscaron en el corpus obtenido de palabras frecuentes, además de los análisis de homografías más convencionales. Para obtenerlas tuvieron en cuenta el carácter indisoluble de los constituyentes, el valor “monossémico” del grupo y la equivalencia morfológica y sintáctica con las clases gramaticales correspondientes. Una vez distinguidas estas locuciones, se incluyeron solo aquellas que alcanzaron el umbral de frecuencia establecido para el resto de palabras. Así, por ejemplo, algunas de las locuciones que incluye el portugués fundamental son: *além de, a não ser, a partir de* (como locuciones prepositivas); *de maneira que, assim como, mesmo que* (como locuciones conjuntivas); *a gente, cada um* (como locuciones pronominales); y *de facto, por acaso, em geral* (como locuciones adverbiales). Entre las que se incluyeron entre la lista de palabras disponibles encontramos por ejemplo *nível de vida, fazer as malas* (como colocaciones) y *no entanto* (como locuciones). En la lista en la que se organizan las palabras por clases gramaticales, se observan los apartados específicos de “conjunciones y locuciones coordinativas”, otro de “conjunciones y locuciones subordinativas”, otro de “preposiciones y locuciones prepositivas”, “adverbios y locuciones adverbiales”. Con ello se puede observar cómo la perspectiva sobre el tipo de unidades básicas a tener en cuenta y el concepto de palabra, es mucho

más actual que en muchos trabajos que se publicaron en décadas posteriores. En el listado de clases de palabras se incluye el artículo antes de cada sustantivo, lo cual resulta altamente útil si tenemos en cuenta las diferencias y dificultades sobre el género de las palabras que se pueden dar en el aula de lenguas extranjeras.

Además de la lista por orden alfabético y la lista por clases de palabras, se añade otro apartado que se denomina “Listas especiais”, que incluye 11 tipos de palabras y locuciones: cronología (división del tiempo, momentos del día, relaciones de tiempo, días de la semana, meses, estaciones, épocas, etc.), puntos cardinales, nombres de parentesco, colores, elementos de identificación, saludos, formas de cortesía, al teléfono, formas de tratamiento, interjecciones y palabras de valor interjetivo y lo que en portugués se llaman “bordões” y vienen siendo una especie de muletillas del tipo *bom, bem, pois, não é*, etc. Para buscar los elementos pluriverbales en la lista unitaria por orden alfabético, se puede buscar por la palabra que se considera principal, por ejemplo encontramos debajo de la entrada *vez(es)* una serie de expresiones en letra más pequeña: *em vez de, cada vez mais, de vez em quando, outra vez, uma vez, às vezes, muitas vezes, outras vezes, por vezes*; debajo tanto de *querer* y como de *dizer*, la expresión *quer dizer*. Es decir, las coocurrencias fueron indexadas junto a cada uno de los elementos que las constituyen, excepto en el caso de los artículos, de las preposiciones y de las conjunciones *que* y *e*. Aunque se encuentran como entradas propias las locuciones cuya frecuencia era igual a la del núcleo, bien porque este no aparece como palabra aislada (ejemplo: *de repente*), bien porque no existe fuera de la locución (ejemplo: *no entanto*). Todas estas distinciones resultan muy útiles para ser aplicadas a la docencia y discernir qué locuciones o colocaciones son más usuales, por lo tanto deben ser enseñadas desde un inicio.

Hemos indicado que el trabajo del Español Fundamental amplía sus centros de interés a 25, mientras que el portugués trabaja primero con 27 y luego con 3 centros de interés complementarios del periodo postrevolucionario. En la siguiente tabla se pueden ver estas diferencias:

	<i>Français Fundamental</i>	<i>Español Fundamental</i>	<i>Português Fundamental</i>
1	<i>Les parties du corps</i>	<i>El cuerpo humano</i>	<i>O corpo humano</i>
2	<i>Les vêtements</i>	<i>Los vestidos</i>	<i>O vestuário</i>
3	<i>La maison</i>	<i>La casa</i>	<i>A casa e os móveis da casa</i>
4	<i>Les meubles de la maison</i>	<i>Los muebles de la casa</i>	<i>Refeições, alimentos e bebidas</i>
5	<i>Les aliments et boissons des repas</i>	<i>Alimentos y bebidas</i>	<i>Cozinha e objetos que vão à mesa</i>
6	<i>Les objets placés sur la table et dont on se sert à tous les repas de la journée</i>	<i>Los objetos colocados en la mesa</i>	<i>Estabelecimentos de ensino (pessoas e coisas)</i>
7	<i>La cuisine, ses meubles et les ustensiles qui s'y trouvent</i>	<i>La cocina</i>	<i>A cidade</i>
8	<i>L'école, ses meubles et son matériel scolaire</i>	<i>Vida escolar</i>	<i>Aldeia e trabalhos do campo</i>
9	<i>Le chauffage et l'éclairage</i>	<i>La ciudad</i>	<i>Meios de transporte</i>
10	<i>La ville</i>	<i>El pueblo</i>	<i>Animais</i>
11	<i>Le village ou le bourg</i>	<i>Los medios de transporte</i>	<i>Divertimentos e passatempos</i>
12	<i>Les moyens de transport</i>	<i>Trabajo en el campo</i>	<i>Profissões e ofícios</i>
13	<i>Les travaux des champs et du jardinage</i>	<i>Animales</i>	<i>Saúde e doença</i>
14	<i>Les animaux</i>	<i>Juegos y distracciones</i>	<i>O café</i>
15	<i>Les jeux et distractions</i>	<i>Oficios y carreras</i>	<i>Meios de informação</i>
16	<i>Les métiers</i>	<i>Salud y enfermedad</i>	<i>A religião</i>
17		<i>La calle</i>	<i>Desportos</i>
18		<i>La información</i>	<i>Plantas, árvores e flores</i>
19		<i>Vida religiosa</i>	<i>Casas comerciais</i>
20		<i>Deporte e instalaciones deportivas</i>	<i>O tempo (condições atmosféricas)</i>

21		<i>Plantas, árboles y flores</i>	<i>Higiene pessoal</i>
22		<i>Compras</i>	<i>Viagens</i>
23		<i>Clima</i>	<i>A família e a vida familiar</i>
24		<i>Acciones físicas sin herramienta</i>	<i>A vida sentimental</i>
25		<i>Acciones con herramienta</i>	<i>O correio</i>
26			<i>A arte</i>
27			<i>Verbos referentes à vida mental</i>
28			<i>Vida política</i>
29			<i>Relações de trabalho</i>
30			<i>Problemas económicos de carácter colectivo</i>

TABLA 17. Centros de interés usados en los proyectos del francés, el español y el portugués fundamental

Como se puede observar, las modificaciones prácticamente consisten solo en ampliar los 16 centros de interés del francés, bien para intentar captar léxico de ámbitos que no se habían incluido en el primero y que son de carácter general (como Plantas, árboles y flores, Salud y enfermedad, Clima, etc.), bien para tratar temas que son propios de algunos de los países en los que se trabaja (como en el caso del portugués con los centros de interés “O café, Vida política, Problemas económicos de carácter colectivo”). En algunos casos se han unido dos centros de interés (como los iniciales “La maison” y “Les meubles de la maison” que pasan a “A casa e os móveis da casa”). Resulta también propio de los tiempos vividos en la España y el Portugal de la época la inclusión en ambos de Vida religiosa y “A religião”. En general, los tipos de centros de interés que se incluyen en los dos proyectos posteriores son del mismo tipo y siguen los mismos criterios que los franceses iniciales. No se lleva a cabo una transformación metodológica en este sentido, y poco se pueden mantener hoy en día esas ampliaciones, pues no ayudan demasiado a obtener resultados que se acerquen más a los intereses que perseguimos en la actualidad.

Por otra parte, la metodología seguida en los tres casos para las encuestas limitaba las clases de palabras por las que se preguntaba (sustantivos en Francia; sustantivos, adjetivos y verbos en el caso español y portugués) y se acotaba la encuesta no mediante el tiempo destinado a obtener las respuestas, sino mediante el número de respuestas que se pedían en cada una de ellas. Debemos pensar que esto tiene también que ver con lo que se perseguía con las encuestas de disponibilidad léxica, es decir, en un inicio estaban pensadas para corregir la lista de palabras más frecuentes en la que se detectaba la escasez de sustantivos concretos y específicos. Posteriormente la aportación de la disponibilidad léxica se ve como complementaria y se habla de fusionar ambos tipos de léxico, más que de corregir el primero, pero siempre se trata de buscar aquel tipo de vocabulario que no se conseguía obtener con los corpus de frecuencias léxicas, que en los casos de los proyectos de léxico fundamental siempre fueron corpus definidos como orales y extraídos de conversaciones transcritas. Posteriormente, los trabajos de disponibilidad léxica se llevaron a cabo de forma independiente y como método autónomo de obtención de vocabulario. De ahí que se intentara obtener el mayor tipo de categorías posibles y se profundizara en qué condiciones de las encuestas son las que ayudan en mayor medida a obtener el léxico que está disponible en nuestro lexicón cuando activamos un tema concreto. En este sentido, se cree que la condición no debe ser el número de palabras ni las clases de palabras sino que se debe dejar libertad en ello al informante y limitar la encuesta a un tiempo determinado. Es ese tiempo, que debe ser breve pues luego las palabras que se obtienen ya no serían las que se activan primero sino que aparecen también las más rebuscadas en nuestra memoria, el que resulta más adecuado. Hasta ahora se ha trabajado a raíz del Proyecto Panhispánico con 2 minutos, aunque hay propuestas recientes que apuntan a 3 minutos (Tomé Cornejo 2015), también hay estudios que no toman el tiempo como limitación.

Los tres trabajos tienen como objetivo ofrecer a la didáctica de lenguas extranjeras un método de selección léxica objetivo, basado en técnicas científicas de estadística léxica como son el índice de frecuencia, repartición o distribución, la frecuencia de las palabras disponibles y su rango. Nacen, por lo tanto, con una finalidad didáctica muy concreta y tienen en cuenta claramente los enfoques y

métodos a los que pretenden aportar su vocabulario. En el caso del francés y el español van de la mano del método audiovisual. En el del portugués, sus autores reconocen en la introducción que su propuesta es la misma que la de sus homólogos franceses y españoles: la selección gradual del vocabulario y el establecimiento de una gramática básica que destierren los criterios subjetivos e intuitivos en la enseñanza de las lenguas extranjeras. Así, sitúan sus desarrollos técnicos en el ámbito de la Lingüística Aplicada y reconocen que varias décadas después de la publicación del *Francés Fundamental*, las nuevas tendencias didácticas se relacionan con el concepto de competencia comunicativa, que deriva en un nuevo enfoque. No obstante, la necesidad de un inventario léxico que nutra a los materiales didácticos de las palabras y expresiones más usuales en las fases iniciales, así como las más propias de un registro oral o de la escritura, sigue vigente con cualquier método y enfoque, por lo que el léxico fundamental sigue siendo un recurso como forma de establecer jerarquías, graduaciones y distinciones en el vocabulario mediante criterios científicos.

Además, el hecho de que estos proyectos tengan un objetivo claro de partida, se puede relacionar con la metodología utilizada. Así, los posteriores proyectos de léxico disponible en el ámbito hispánico, que hasta hoy han descartado una finalidad única relacionada con la enseñanza de lenguas, han optado exclusivamente por desarrollar la disponibilidad como método de obtención de listas de vocabulario con fines científicos y apartados de la Lingüística Aplicada a la enseñanza de lenguas. Por el contrario, los estudios de léxico fundamental pretenden ofrecer un vocabulario lo más completo posible, y por ello parten de la frecuencia y la disponibilidad para conseguirlo, conscientes ya desde la publicación del *“Français Élémentaire”* de las limitaciones de los corpus de frecuencia y las posibilidades alternativas de las encuestas de disponibilidad. Por consiguiente, ya desde aquel momento se comenzó a explicar que con las dos vías se obtienen clases de palabras diferentes y de distintas categorías, por lo que una solución es unir ambas. En este sentido, ninguna de ellas ofrece una solución satisfactoria por sí sola a la cuestión de la selección léxica. La disponibilidad léxica, por mucho que mejore y amplíe sus centros de interés y su metodología de análisis, solo puede aspirar a ofrecer aquellos conceptos y clases de palabras que surgen en relación a temas determinados, y

por ello puede ser una vía más que útil en el caso de cubrir las nociones específicas del *PCIC*. No obstante, estas limitaciones se observan más recientemente, cuando se empieza a reivindicar la utilidad de la disponibilidad léxica para la selección léxica, pues en la enseñanza del español no se había recurrido demasiado a criterios claros, objetivos y provenientes de los proyectos de léxico-estadística. Es, pues, ahora cuando la disponibilidad léxica se empieza a reorientar, o más bien especializar, a la aplicación como método de selección léxica, y es por este motivo por lo que retomamos la idea del léxico fundamental como un método más completo para este fin.

7.6. Metodología y criterios de selección de las palabras en los proyectos de vocabulario fundamental

Desde un principio, para la definición del vocabulario fundamental se ha recurrido a especificar el tipo de unidades léxicas que lo componen, los métodos que se usan para obtenerlas y el objetivo al que va destinado. Así, si se observa la explicación de este concepto en la publicación del *Francés Fundamental* de 1964, se puede ver una defensa del concepto como forma de facilitar la adquisición de la lengua francesa por parte de los extranjeros. Pensemos en la difusión de esta lengua como medio de comunicación a nivel internacional desde el siglo XVIII y como lengua oficial de muchas de sus excolonias, sobre todo el norte de África. De modo similar al *Basic English*, pretende ayudar a una rápida adquisición de la lengua francesa. No obstante, no se considera un código cerrado como en el inglés, sino que se crea un vocabulario y una gramática de base, delimitando una serie de palabras para ser usadas en la enseñanza en el nivel inicial de aprendizaje. Así, los franceses definieron su vocabulario fundamental como un instrumento de difusión rápida de la lengua que aporta una lista definitiva de 1.475 palabras, compuesta por 1.222 palabras léxicas y 253 palabras gramaticales. Pero, repetimos, no se pensó como una propuesta cerrada y completa para adquirir la lengua, sino como una introducción a ella en un primer momento, pues se pensaba que su conocimiento permitía un dominio básico del idioma en un periodo corto de tiempo. Se diferencia así de la propuesta del inglés, en que este era un código completo, y no una primera fase de la lengua

inglesa. Los franceses concibieron el vocabulario fundamental como una técnica innovadora para la creación de una lengua básica, que combina los siguientes elementos:

1. la frecuencia de las palabras de la lengua hablada,
2. la búsqueda de las palabras disponibles que resulten más útiles y
3. el empirismo racional en la metodología.

Recurrir a la lengua hablada y a la noción de disponibilidad constituyen dos innovaciones a las anteriores propuestas de lenguas básicas. Con esta lista de palabras se alcanza el objetivo de establecer los contenidos léxicos de una primera fase de enseñanza del francés como lengua extranjera.

Lo que se observa sobre el concepto en proyectos posteriores de otras lenguas es casi idéntico en cuanto a objetivos, tipos de unidades y metodologías usadas. En el caso del español y del portugués siguen estos mismos parámetros. En el del español se pone énfasis en que es un método de selección léxica que sigue, por primera vez en esta lengua, técnicas objetivas con criterios precisos, claros y complementarios, basados en:

1. el índice de frecuencia (F) e índice de distribución (ID), sacados ambos exclusivamente de conversaciones;
2. el índice de disponibilidad de palabras concretas obtenidas según la técnica de centros de interés;
3. el reajuste correctivo de los resultados provisionales mediante técnicas empíricas.

Con todo, en los proyectos francés y español no se observa una definición clara de la noción de vocabulario fundamental. Los investigadores franceses se limitan a definir su objetivo —asegurar y agilizar el aprendizaje del francés por los inmigrantes y habitantes de las excolonias, así como por los extranjeros que usen esta lengua como instrumento de comunicación a nivel diplomático e internacional— y a describir el método para llegar a él —la corrección de un corpus oral de frecuencia mediante los sustantivos concretos que se obtienen de las encuestas temáticas—. El Español Fundamental desarrolla y perfecciona las

cuestiones metodológicas, pero no ahonda en el concepto. Es el Portugués Fundamental el que avanza una definición más clara y concisa. Así, el primer párrafo de su introducción está dedicado a establecer la naturaleza y objetivos de un Vocabulario Fundamental:

O Português Fundamental, estabelecido de acordo com critérios de frequência de emprego, segundo princípios de estatística lexical, a partir de amostragens do português europeu falado contemporâneo, pode definir-se como um vocabulário indispensável a uma efectiva capacidade de comunicação em situações da vida corrente. Este trabalho tem dois objetivos principais. Um é fornecer os dados que permitam estabelecer conteúdos lexicais adequados ao ensino do Português, língua estrangeira, desde o nível mais elementar da aprendizagem. O outro é apresentar os resultados de análises gramaticais sobre aspectos morfo-sintáticos do discurso oral, de interesse mais imediato para o ensino do Português. [Malaca Casteiro 1984: 15]

Como se puede observar, la finalidad de la elaboración de un Vocabulario Fundamental es constituir un conjunto de unidades léxicas indispensables para llevar a cabo una comunicación efectiva en situaciones cotidianas y corrientes. La metodología usada para llegar a establecer estas unidades es el uso de métodos de léxico-estadística para analizar muestras orales de una lengua. Y el objetivo de este vocabulario es aportar el contenido léxico y morfosintáctico necesario y adecuado para un nivel inicial de aprendizaje de esta lengua por parte de extranjeros.

La necesidad de desarrollar este método viene dada por el objetivo, es decir, el de facilitar el aprendizaje de estas lenguas (francés, portugués, español, ...) por parte de extranjeros. Se necesita delimitar unos contenidos lingüísticos que permitan el dominio del idioma en un periodo relativamente corto de tiempo, así como establecer una progresión en el aprendizaje. Es por ello por lo que se hace necesario reducir todo un sistema lingüístico a un vocabulario y un conjunto de reglas gramaticales que resulten las necesarias y más útiles para comunicarse en situaciones cotidianas. Por otra parte, dejar en mano de los autores de manuales y profesorado la selección de estos contenidos usando su intuición o experiencia personal derivaba en desajustes y diferencias notables, pues se asentaba en un método subjetivo. Los vocabularios fundamentales y gramáticas de base ofrecían un modo de operar de carácter estadístico, asentado

en cálculos e índices matemáticos que permitían trabajar sobre muestras reales para obtener ese conjunto léxico y gramatical que enseñar. Al mismo tiempo, se partía de que las unidades léxicas y reglas gramaticales necesarias para la comunicación más básica eran las más frecuentes, por lo que ya teníamos un objetivo, un tipo de unidades concretas y un medio estadístico para conseguirlas.

Sobre la necesidad de recurrir a un corpus de léxico disponible y no solo a un vocabulario frecuente, todos ellos se expresan en los mismos términos. En el caso de los recuentos de frecuencias se observa que solo se consigue obtener un tipo de unidades léxicas, por lo que no se puede confiar a este método la labor de selección del vocabulario por completo. Las listas de frecuencias nos facilitan:

- las palabras de uso gramatical (*mots gramaticaux*) con una significación limitada (por ejemplo en el caso del francés: *de, je, il(s), ce* (pronombre), *la* (artículo), *pas* (negativa), *a* (preposición), *et, etc.*);
- algunos verbos de frecuencia muy elevada (*être, avoir, faire, dire, aller, voir, savoir, etc.*);
- sustantivos de carácter general, pero con poca carga semántica (*chose, heure, temps, an, monsieur, fille, type...*);
- algunos adjetivos muy comunes (*tout, petit, mon...*);
- muy pocos sustantivos de significado concreto.

Aunque los ejemplos son del francés, todos los trabajos de frecuencias léxicas de cualquier otra lengua consiguen resultados de este tipo: vocablos gramaticales y de significado muy general, o como lo denominaron en francés, palabras *passe-partout*. Es por ello por lo que constituye un instrumento de selección insuficiente para enseñar una lengua, pues nos ofrece un tipo de unidades, pero no todas; faltan los sustantivos, adjetivos y verbos con una mayor carga semántica y más específicos. Por este motivo intentan desarrollar un nuevo método con el que completar y corregir la lista de palabras frecuentes obteniendo aquellos elementos que se echan en falta: los sustantivos concretos y con alta carga semántica en un primer momento, y los sustantivos, adjetivos y verbos con posterioridad. Como ya sabemos, fue pensando en cómo acceder a ellos cuando se detectó que solo aparecían cuando se nombraban determinados

temas, y así se llegó a la lista de centros de interés y la metodología de encuestas de disponibilidad léxica. Por lo tanto, habían definido una metodología que ayudaba a delimitar de modo más completo y eficaz el vocabulario de una lengua.

En estos primeros trabajos, aunque se hable por un lado de palabras frecuentes o atemáticas y de palabras disponibles o temáticas por otro, la base de ambos corpus se asienta en la frecuencia. Se obtiene el corpus de frecuencia mediante conversaciones grabadas tratando temas diversos y el corpus de léxico disponible mediante encuestas individuales, escritas y con temas concretos y bien delimitados, pero, a la hora de medir matemáticamente la frecuencia de las palabras en este corpus la fórmula es la misma: medir la frecuencia que esos vocablos obtienen en ese conjunto de palabras. Más tarde se tendrá en cuenta, recordemos, la dispersión, llamada *repartición* o *distribución* en el caso de la frecuencia, y el *rango* en el de la disponibilidad, si bien el concepto es muy similar en ambos corpus, pues de algún modo mide el número de informantes que nombran esa palabra y su distribución en el conjunto. En el caso del corpus de frecuencia se tiene en cuenta las veces que aparece el vocablo en los diferentes textos transcritos y en el caso de la disponibilidad se fija en la posición en que esa palabra es nombrada en las encuestas. Con todo, en ambos casos son medidas correctoras de la frecuencia. Pensemos que estamos hablando de trabajos que no tienen en cuenta las fórmulas más complejas para obtener el índice de disponibilidad que crearon años después Lorán (Lorán & López Morales 1983; Lorán 1987) y, perfeccionadas, López Chávez y Strassburger (1987). De ahí que en los proyectos del francés y el español se sumen los resultados en un listado unitario tomando como equivalentes sus unidades y las cifras de frecuencia que alcanzan unas y otras en las listas separadas. En el caso del corpus de palabras frecuentes, la frecuencia se corrige con la *repartición* (en el francés y portugués) o *distribución* (en el español), que de alguna forma se puede comparar con las fórmulas para medir la dispersión simple y compleja desarrolladas por Juilland y Chang-Rodríguez en los léxicos de frecuencias. Sin embargo, estos autores establecían una serie de conjuntos en los que se clasificaban todos los textos estudiados, mientras que, en el caso de los

vocabularios fundamentales, la comparación se hace entre todas las conversaciones.

El francés y el español fundamentales se limitan a establecer un umbral de frecuencia a partir del cual incluyen las palabras en las listas definitivas o no: en el francés, una frecuencia igual o superior a 29 ($F \geq 29$) y que las palabras hayan sido nombradas al menos en cinco textos; en el español, una frecuencia igual o superior a 50 ($F \geq 50$), con inclusión de algunos términos cuya frecuencia está entre 50 y 20 posteriormente. El español hace explícito que también seleccionó las palabras disponibles del mismo modo, estableciendo un límite de frecuencia de 50 en un primer momento y de entre 50 y 25 *a posteriori*. En el caso del francés deducimos que algún criterio de selección tuvieron que seguir, pero no se explica cuál. Quizás es debido a que la mayor parte de los vocablos (palabras diferentes) obtenidos en las encuestas pasaron a esa lista final. El portugués, por su parte, va más allá y no solo establece un umbral de selección de las palabras frecuentes, sino que también desarrolla una fórmula para escoger las palabras disponibles. En el caso de las palabras frecuentes, el límite es de frecuencia igual o superior a 40 ocurrencias ($F \geq 40$), teniendo en cuenta ya la repartición y luego de haber probado con otro límite de entre 50 y 30. Para fijar la barrera de frecuencia en 40 apariciones, se consideraron varios factores: la extensión del corpus, el porcentaje de formas homógrafas a analizar, el número aproximado de términos que el vocabulario debería contener y las experiencias de los vocabularios fundamentales de otras lenguas románicas.

Para crear la fórmula de selección del vocabulario disponible portugués, se tuvieron en cuenta diferentes aspectos de este tipo de encuestas. En un primer momento se planteó si el orden en el que las palabras eran nombradas influía para establecer el grado de disponibilidad de los vocablos. Como hemos dicho, a principios de los ochenta se estaban barajando estas posibilidades, pero el Índice de Disponibilidad como hoy lo entendemos, es decir, teniendo en cuenta en su fórmula el número de ocurrencias y el orden en que aparecen, todavía no se había logrado de forma eficaz. Se consideraron, no obstante, las propuestas y trabajos sobre esta cuestión, como los ofrecidos por Rivenc y por Müller. Estos investigadores se habían preguntado si el lugar de las palabras en cada lista

constituía o no un valor pertinente para la selección del vocabulario disponible en cada tema. Así, al estudiar las listas de vocabulario disponible del Español Fundamental, habían colocado sobre la mesa la posibilidad de que para una misma frecuencia o para frecuencias próximas, los vocablos que habían sido más citados en los primeros lugares de las listas pudieran ser afectados de un grado de disponibilidad más elevado que los que aparecían más veces en los lugares más alejados de la cabeza de la lista. Müller, además, había establecido una fórmula matemática, antecedente del posterior Índice de Disponibilidad, que permitía calcular aquello que designó como “índice de espontaneidad”. Sin embargo, los cálculos efectuados en el subcorpus del vocabulario disponible del español no hicieron patentes diferencias significativas entre las listas ordenadas por frecuencias decrecientes de los vocabularios, como se había hecho desde el inicio, y las listas ordenadas de acuerdo con los índices de espontaneidad, por lo menos en las posiciones significativas de las listas, que incluían según los casos los 20 o 30 primeros vocablos. Es por esta conclusión por lo que en el Portugués Fundamental se renunció a usar este nuevo índice y se optó por tener en cuenta exclusivamente la frecuencia que concreta el grado de disponibilidad de los vocablos en cada una de las listas de los centros de interés. Esto es algo que hoy en día nos puede resultar cuando menos curioso, pero hay que tener en cuenta que en ese momento la metodología sobre disponibilidad léxica no contaba con las fórmulas y el desarrollo de hoy, al igual que ocurre con la cuestión de solicitar a los informantes un número concreto de unidades en cada clase de palabra.

Aun así, y aunque sus opciones no resulten hoy por hoy estadísticamente rigurosas, se hace manifiesto que la forma de proceder del equipo de investigadores portugueses en cuanto a establecer los umbrales de selección se lleva a cabo con criterios empíricos metódicos, teniendo en cuenta la variación de ciertos valores significativos. De este modo, manifiestan que tienen en cuenta tres factores para la creación de los diccionarios de vocabulario disponible: la dispersión, la frecuencia del primer vocablo y el número de palabras que se pide a los informantes en cada centro de interés y categoría.

Con la *dispersión* nos referimos a la extensión del conjunto de vocabulario ofrecido por los encuestados para un determinado centro de interés en cada

categoría gramatical por la que se pregunta. Así, cuanto más elevada es esta extensión, menos estables son las frecuencias de los vocablos situados en las primeras posiciones de la lista. En cambio, cuanto más baja es su dispersión, más rápidamente crece su frecuencia. El rendimiento de un centro de interés es, por lo tanto, inversamente proporcional a la dispersión de su vocabulario.

En cuanto a la frecuencia del primer vocablo de la lista, también hay una correlación. Cuanto más elevadas y estables son las frecuencias de los vocablos que se sitúan en los primeros lugares de la lista más alta será la de la primera palabra. El rendimiento de un centro de interés es, por lo tanto, proporcionalmente dependiente de la frecuencia global del primer lema de la lista del diccionario de léxico disponible. Como se observó que los resultados no variaban sensiblemente aunque se tomase solo la frecuencia media de los 5 o 10 primeros vocablos de la lista o la frecuencia del vocablo que aparece en primer lugar, se optó por esta última opción por resultar menos laboriosa.

En tercer lugar, y aunque de modo menos regular, cuanto mayor es el número de palabras que se pide a los encuestados en cada centro de interés, mayor es también el rendimiento de los informantes, pues anotan un mayor número de formas.

Estas son las cuestiones que se tuvieron en cuenta para crear la fórmula con la que los investigadores portugueses seleccionaron las palabras de cada centro de interés para llegar al listado de palabras disponibles. El resultado fue el siguiente:

$$F_L = \frac{F_1}{D} \times N \times K$$

En esta fórmula, podemos observar que **F1** es la frecuencia del primer lema que está a la cabeza de la lista; **D** es la dispersión del subcorpus recogido para determinado centro de interés y para una categoría dada; **N** es el número de palabras de una misma categoría gramatical pedida a cada encuestado (15 sustantivos, 10 adjetivos y 10 verbos); **K** es un coeficiente arbitrario, establecido en función del número de vocablos que se pretende obtener para cada centro de

interés (fijado en 10 para el conjunto de los temas de la encuesta principal, excepto en cinco casos). Con estos cálculos seleccionaron el vocabulario disponible que mostraron a la Comisión que tomaba las decisiones finales sobre el proyecto. Uno de los problemas o deficiencias que aparentemente se manifestaron con este método fue las enormes diferencias entre la cantidad seleccionada para dos de los centros de interés frente al resto. El desequilibrio es notable, pues en la mayor parte de los centros de interés el número de vocablos seleccionados está entre los 17 y los 40 (18, 19, 32, 27, 25, 28, 30, 41...), mientras que en el CI *Aldeia e Trabalhos do Campo* resultaron escogidos 92 sustantivos, 72 adjetivos y 43 verbos. En cambio en el CI *Animais* se escogieron 10 sustantivos, en el CI *O Tempo* 10 adjetivos y en el CI *Casas Comerciais* solo 8 verbos. El motivo que expresan los autores es que se unieron lo que se había previsto como dos centros de interés (*A Aldeia y Trabalhos do Campo*) con lo que el centro de interés resultante mostraba una dispersión de 1826 vocablos para los sustantivos, 894 para los adjetivos y 821 para los verbos. Por el contrario, hay 5 centros de interés que muestran una dispersión muy inferior a la media.

Capítulo 8. Hacia un vocabulario fundamental del español

Como hemos comentado, si bien la tradición de la disponibilidad léxica en el ámbito hispánico se ha desarrollado al margen de los corpus y las listas de frecuencia, en sus inicios se creó para ser un método corrector de sus deficiencias, no para ser una alternativa a ellos, como parece ser la tendencia en la actualidad. Después de varias décadas de contribuciones en esta línea, muy poco se ha avanzado en intentar unir las palabras disponibles a las palabras frecuentes y en su aplicación didáctica. Apenas tenemos como referencia nada más que la definición de lo que sería el léxico fundamental para el responsable del Proyecto Panhispánico, López Morales, quien no percibía la disponibilidad léxica como una forma de completar el vocabulario básico, recordémoslo, sino como uno de los elementos que junto al léxico básico constituirían el léxico fundamental. Debido también a esta separación y autonomía de la disponibilidad léxica, se ha observado un cambio en lo que se espera de ella, de tal modo que, si en sus principios se preguntaba solo por los 20 sustantivos más disponibles de un número reducido de centros de interés, en las investigaciones actuales sobre el español no solo se pretende obtener nombres y, además, se han ido aumentando algunos centros de interés. Algunos proyectos incluso han solicitado sustantivos, adjetivos y verbos explícitamente, para evitar que las palabras que aportaran los encuestados fueran únicamente las primeras. No obstante, en los estudios más recientes no se limita la categoría léxica, con la esperanza de que puedan surgir otras clases de palabras y expresiones pluriverbales.

Con todo, y aunque la disponibilidad léxica se ha difundido considerablemente, no se han explotado, quizá, todas sus posibilidades para la

docencia. Podemos decir lo mismo de los corpus lingüísticos, cuyo uso goza de arraigada tradición en la enseñanza de otras lenguas, como el inglés o el alemán, en las que la metodología de selección léxica y la lexicografía continúan teniendo como base la lingüística del corpus y los estudios de frecuencias. Y esto se observa tanto para la didáctica a extranjeros —a través de diccionarios, manuales, lecturas graduadas, recursos electrónicos de diverso tipo, etc.—, como para la elaboración de grandes obras lexicográficas, pensadas ya para usuarios que dominan el idioma y para hablantes nativos.

En el español, la preocupación de la lingüística aplicada por la enseñanza de la lengua a extranjeros es mucho más reciente y la creación de grandes corpus para seleccionar el vocabulario es mucho más escasa. De hecho, en inglés se pueden obtener diccionarios que recurren al índice de frecuencia y escogen lo que se puede entender como el “inglés más frecuente” para estudiantes de nivel inicial e intermedio bajo. Un ejemplo sería el *COBUILD Essential English Dictionary* de Harper Collins, que en su segunda edición de 2014 cuenta con algo más de 8.000 entradas. Se anuncia como una herramienta atractiva para los aprendices desde A1 hasta B1 y es muy difícil encontrar un equivalente de estas características en español. Algo similar puede decirse de algunas lecturas graduadas en inglés, que no dudan en indicar el nivel para el que están pensadas y el número de palabras que introducen, en consonancia con lo anterior. Una apuesta igualmente interesante es la clasificación del vocabulario de la lengua por niveles a través de los corpus de aprendices. El inglés, por ejemplo, cuenta con una herramienta en línea de este tipo, *English Vocabulary Profile*⁶⁵, un proyecto en el que han colaborado varias universidades (University of Cambridge, University of Bedfordshire, English UK) y el *British Council*. Consiste en una base de datos en la que se clasifica el vocabulario en los 6 niveles del *MCERL*, distinguiendo entre el inglés británico y el americano. Pero a diferencia del *MCERL* o de los planes curriculares, lo que se ofrece no es lo que los usuarios deben saber hacer, sino lo que realmente saben hacer, ya que para crearlo se han

⁶⁵ Se puede consultar en:

<http://www.englishprofile.org/wordlists>

<http://vocabulary.englishprofile.org/staticfiles/about.html>

basado sobre todo en el *Cambridge Learner Corpus* (CLC), aunque también en el *Cambridge English Corpus*. Contiene unas 7.000 entradas e incorpora expresiones fraseológicas, como por ejemplo los llamados *phrasal verbs* y lo que llaman *phrases*, que son coocurrencias léxicas restringidas. Un ejemplo sería *round of applause*, para un nivel C2. La búsqueda simple tiene dos opciones: obtener una lista de unidades léxicas de un nivel o de varios o bien el nivel de una palabra en concreto. Si buscamos la palabra *apple*, la herramienta dice que pertenece al nivel A1, pero también podemos buscar la lista de palabras de nivel A1 sobre el tema de la alimentación, por ejemplo. Desde las opciones avanzadas se realizan búsquedas más complejas restringidas por filtros (categoría gramatical, tipo de entrada, tema, marca de uso, etc.), con las que podemos obtener, por ejemplo, el listado de los sustantivos del nivel B2 sobre un tema o lanzar otra búsqueda y filtrar los adjetivos informales (como marca de uso) para determinado nivel. La búsqueda avanzada también dispone de un filtro que selecciona solo *phrases*.

Tampoco contamos por ahora con estudios que nos guíen sobre lo que realmente conocen nuestros estudiantes de ELE, ni las palabras que deben conocer para poder llevar a cabo la lectura de textos de forma eficiente. Podemos pensar que hacia estos objetivos se encaminan las encuestas de disponibilidad léxica hechas a extranjeros, que han proliferado desde hace más de una década, pero en ellas no se ofrecen pruebas del uso real de las palabras en contexto, por lo que nada sabemos sobre el verdadero grado de conocimiento y eficacia de uso del léxico por parte de los aprendices. Por otra parte, los corpus de aprendices del español, con los cuales sí contamos, todavía están en una fase inicial, y no se han aprovechado como herramientas útiles para otros proyectos. Como hemos indicado en los primeros capítulos, en los que tratamos el número de palabras necesarias para llevar a cabo una lectura eficaz, en inglés hace décadas que han surgido estudios de este tipo. En ellos, tras trabajar con diferentes grupos de alumnos y llevar a cabo diversas pruebas, se ha llegado a la conclusión de cuántas palabras son necesarias para que se dé una buena u óptima comprensión. Se ha puesto en relación la comprensión lectora con el tipo de vocabulario y la materia de los textos, así como con los corpus de frecuencias y con la cobertura textual que estos ofrecen (Cartaya Febres 2011; Hu & Nation

2000; Laufer 1992, 1997a, 1997b; Laufer & Ravenhorst-Kalovsky 2010; Nation 2001, 2006; Pulido 2009) (*vid. supra* § 4.1.1, 4.2 y 4.3).

En el ámbito de ELE, sin embargo, es difícil encontrar un manual, lectura graduada o libro de vocabulario que explique el número de palabras que ha incluido y cómo se ha hecho esa selección, si es que se ha llevado a cabo con criterios objetivos y científicos de algún tipo. Todo lo contrario, echamos en falta diccionarios básicos actuales que se preocupen por ello y quizás por el mismo motivo los trabajos de disponibilidad léxica se alejaron de este objetivo pedagógico inicial a favor de miras más teóricas y filológicas.

En este sentido, nuestra propuesta de vocabulario fundamental pretende cubrir el vacío que se observa en este campo en la lengua española, pues intenta aprovechar los frutos de dos disciplinas con gran trayectoria (como son los corpus de frecuencias y los léxicos disponibles) para ofrecer contenidos léxicos seleccionados mediante criterios objetivos y que pueden ser graduados en los diferentes niveles del *MCERL*. Además, hoy en día la disponibilidad léxica toma estos niveles del *MCERL* como referente para la elaboración de sus núcleos temáticos, como también hacemos en la conformación y organización de nuestro trabajo. Esto supone considerar el *PCIC* como guía, desde la elaboración hasta la presentación de las palabras, por lo que debemos entender el vocabulario fundamental como una apuesta complementaria, más que como una alternativa al *Plan curricular*, si bien las diferencias y semejanzas que se observan entre ambos nos pueden servir para reconsiderar y analizar más en profundidad los contenidos. En el capítulo del *PCIC* en el que se presentan las nociones específicas de los 6 niveles, así se expone. Aunque la cita es larga, merece la pena su reproducción, pues pone de manifiesto que la naturaleza del documento no es la de ser un texto cerrado ni prescriptivo:

Se trata de una presentación que ha permitido incorporar los contenidos léxicos necesarios para cumplir los objetivos de cada nivel, pero que *debe consultarse, en todo caso, como una lista de carácter orientativo y no como un resultado final e inamovible*. A este respecto, *el inventario debe considerarse abierto* en los distintos sentidos que se exponen a continuación.

[...]

Teniendo en cuenta las consideraciones anteriores, hay que insistir en que el inventario de *Nociones específicas es solamente una base de orientación general* que podrá servir de referencia a quien lo utilice a la hora de seleccionar y

distribuir por niveles las unidades léxicas que precise para sus propios fines. Es, por tanto, *un inventario abierto* en los distintos sentidos que se acaban de exponer, por lo que *los exponentes o las series de exponentes que se incluyen en las entradas de la lista no deben entenderse como elementos cerrados o finales*. [PCIC, Centro Virtual Cervantes. La cursiva es nuestra]

Por otra parte, no podemos olvidar que el método de elaboración de este inventario ha sido la intuición y la experiencia docente de las personas que lo han elaborado, como así se explica:

Como criterios de selección, siempre partiendo de la apreciación intuitiva basada en la experiencia docente, han primado la frecuencia y la rentabilidad comunicativa, es decir, que se tratará de exponentes nocionales necesarios para realizar las funciones comunicativas y para desenvolverse con los textos orales y escritos correspondientes a cada nivel. [PCIC, Centro Virtual Cervantes. La cursiva es nuestra]

Y, como se puede observar en sus propias palabras, estos criterios se suman a la frecuencia y a la rentabilidad comunicativa, lo cual nos lleva a cuestionarnos cómo se pueden conjugar ambas cosas, es decir, cómo medir la frecuencia de las palabras de forma intuitiva y cómo saber cuáles son las más rentables a nivel comunicativo sin hacer estudios de campo como los que han llevado a cabo autores a los que nos hemos referido (Cartaya Febres 2011; Hu & Nation 2000; Laufer 1992, 1997a, 1997b; Laufer & Ravenhorst-Kalovsky 2010; Nation 2001, 2006; Pulido 2009, por citar solo algunos de ellos).

Por nuestra parte, para la elaboración de un vocabulario fundamental del español, hemos analizado la cuestión desde múltiples perspectivas y nos hemos hecho numerosas preguntas, como se puede observar a lo largo de los capítulos de este trabajo. Por una parte hemos estudiado las diferentes propuestas de vocabulario fundamental que se han hecho con anterioridad, pero también hemos analizado los recientes avances de la lingüística del corpus y de las disponibilidad léxica por separado. Nos hemos cuestionado sobre el concepto y la naturaleza del vocabulario fundamental para dar contenido a nuestra propuesta, al mismo tiempo que hemos estudiado en conjunto cuestiones como el tipo y el número de unidades léxicas que son necesarias para cada nivel educativo. Para ello, hemos revisado bibliografía de varios ámbitos, en español y otras lenguas. Hemos intentado establecer patrones de selección del vocabulario

frecuente y disponible lo más explícitos, coherentes, objetivos y empíricos posible atendiendo a criterios matemáticos y estadísticos, aunque tampoco hemos renunciado a estudiar los diferentes términos concretos que obtenemos en cada uno de los inventarios que hemos analizado: vocabularios básicos, diccionarios de léxico disponible y los contenidos en el *PCIC* para las nociones específicas. Todo ello no solo ha supuesto un laborioso trabajo de análisis y síntesis, sino también un constante cuestionamiento de los criterios y métodos que debíamos seguir para conseguir nuestra meta, en aras de hallar procedimientos objetivos para realizarla. Así, nos hemos preguntado hasta dónde debemos llegar usando métodos empíricos, matemáticos y estadísticos, y en qué momentos podemos recurrir a otros parámetros igualmente útiles con los que *de facto* se opera en la docencia de lenguas extranjeras. Sin olvidar que son el resultado de un consenso tácito entre los profesionales del sector, que ha sido alcanzado tras años de trabajo y reflexión.

8.1. Sobre el concepto de vocabulario fundamental: definición

A lo largo del capítulo anterior hemos comentado lo que entendían por vocabulario fundamental las diferentes propuestas que se han elaborado en francés, español y otros idiomas europeos. En ellas se partió del vocabulario frecuente oral para complementarlo con sustantivos de carácter temático y, en menor medida, adjetivos y verbos. Recordemos que la definición que se suele ofrecer de “léxico fundamental” en la bibliografía de disponibilidad léxica es “la unión orgánica del léxico básico y el léxico disponible”. Debemos tener en cuenta que cuando se lanzaron las propuestas del *Basic English* y del *Français Fondamental* todavía no existían los documentos de referencia con los que hoy trabajamos para la enseñanza y acreditación de lenguas extranjeras. En el caso del *Basic English* se ofrecía poder comunicarse de la misma forma que con el inglés normativo, pero usando un vocabulario más simple y elemental compuesto por menos de mil palabras, es decir, renunciando a los términos más complejos e inusuales, así como a las irregularidades gramaticales. En el caso del francés, el proyecto se pensó para la docencia, por lo que se pretendía escoger las palabras más necesarias y útiles para enseñar en los primeros niveles de

enseñanza a extranjeros. Se asociaba eminentemente con la lengua hablada, aunque no se orientaba a ser un nuevo código alternativo que sirviese para la comunicación en todos los contextos, sino a ofrecer un nivel inicial del francés, dividido en dos estadios. En el caso del inglés no había una asociación estrecha con un tipo de unidades léxicas concretas, mientras que en el francés sí que existía una relación directa.

Por nuestra parte, hemos preferido concebir el vocabulario fundamental desde dos perspectivas diferentes, aunque complementarias: una definición **analítica**, para lo que hemos contemplado los tipos de unidades léxicas que lo componen, palabras usuales y disponibles; y a una visión **funcional**, que estudie su composición y considere sus destinatarios. Sin olvidar que en este momento ya contamos con la descripción de los diferentes niveles de lengua y del dominio que es necesario en cada estadio, algo que no existía cuando se realizó el Español Fundamental de Rojo Sastre o el Francés Fundamental. En realidad, estamos ante el mismo asunto, pero encarado desde diferente óptica, en función de si ponemos el énfasis en el usuario como aprendiz de una lengua o como hablante de ella. Si nos situamos en la primera perspectiva, entendemos el vocabulario fundamental de una forma totalizadora y comunicativa. En este caso, un vocabulario fundamental debe ser aquel que capacite al individuo para desenvolverse con eficacia en situaciones comunicativas cotidianas tratando temas habituales. Esto nos sitúa en lo que llamamos un nivel umbral: el grado mínimo de dominio que se debe alcanzar para desenvolverse en situaciones del día a día y el vocabulario que se relaciona con ellas. Es decir, un nivel superior al de la simple supervivencia, equivalente al nivel B1 del *MCERL*, y que es el que les corresponde a los que denominamos “usuarios independientes” de un idioma. Por esta vía, observamos el vocabulario fundamental poniendo el énfasis en lo que nos permite hacer como hablantes, y no cabe duda de que hasta el nivel umbral no alcanzamos un dominio suficiente para ser capaces de desenvolvernó en la mayor parte de las situaciones, así como para afrontar la comunicación con interlocutores nativos de modo más general. Según el *MCERL*, quien obtenga este nivel de lengua será capaz de:

comprender los puntos principales de textos claros y en lengua estándar si tratan sobre cuestiones que le son conocidas, ya sea en situaciones de trabajo, de estudio o de ocio. Sabe desenvolverse en la mayor parte de las situaciones que pueden surgir durante un viaje por zonas donde se utiliza la lengua. Es capaz de producir textos sencillos y coherentes sobre temas que le son familiares o en los que tiene un interés personal. Puede describir experiencias, acontecimientos, deseos y aspiraciones, así como justificar brevemente sus opiniones o explicar sus planes. [MCERL: 26]

Se identifica el B1 como el reflejo de la especificación del nivel umbral para un viajero que va a un país extranjero y posee dos capacidades principales: consigue mantener una interacción y logra hacerse entender en varias situaciones (MCERL: 37). He aquí las competencias recogidas en el MCERL:

- generalmente comprende las ideas principales de los debates extensos que se dan a su alrededor siempre que el discurso se articule con claridad en nivel de lengua estándar;
- ofrece y pide opiniones personales en un debate informal con amigos;
- expresa de forma comprensible la idea principal que quiere dar a entender;
- utiliza con flexibilidad un lenguaje amplio y sencillo para expresar gran parte de lo que quiere;
- es capaz de mantener una conversación o un debate, pero a veces puede resultar difícil entenderle cuando intenta decir exactamente lo que quiere;
- se expresa comprensiblemente, aunque sean evidentes sus pausas para realizar cierta planificación gramatical y léxica y cierta corrección, sobre todo en largos períodos de producción libre.

Por otra parte, tiene la capacidad de saber cómo enfrentarse de forma flexible con problemas cotidianos, por ejemplo (MCERL: 37):

- se enfrenta a situaciones menos corrientes en el transporte público, a las típicas situaciones que suelen surgir cuando se realizan reservas para viajes a través de una agencia o cuando se está viajando;
- participa en conversaciones sobre asuntos habituales sin haberlo previsto;
- es capaz de plantear quejas;
- toma la iniciativa en una entrevista o consulta (por ejemplo, sabe cómo iniciar un nuevo tema), aunque aún se le nota que depende del entrevistador durante la interacción;
- sabe cómo pedir a alguien que aclare o desarrolle lo que acaba de decir.

Si nos centramos en el vocabulario que este hablante debe dominar, sabemos que en este nivel B1 su repertorio lingüístico es lo suficientemente amplio como para desenvolverse sin graves problemas en temas como su familia, sus aficiones e intereses, su trabajo, sus viajes y acontecimientos actuales. También sabemos que su vocabulario es adecuado para expresarse, aunque con

dudas y con circunloquios (*MCER*: 32). Al explicar el conocimiento del vocabulario que debe tener un hablante de este nivel, el *PCIC* expresa lo mismo: que es el propio de aquella persona que cuenta con una riqueza léxica suficiente para expresarse sobre la mayoría de los temas pertinentes para su vida diaria, como familia, aficiones e intereses, trabajo, viajes y hechos de actualidad, aunque lo haga con algún circunloquio. En cuanto al control que debe tener, se especifica que debe dominar un vocabulario elemental, aunque cometa errores cuando expresa pensamientos más complejos, o cuando aborda temas y situaciones poco frecuentes.

En este sentido, la selección del vocabulario que llevemos a cabo debe permitirnos alcanzar estos objetivos marcados previamente, por lo que debemos calcular cuántas unidades y cuáles son necesarias para que este dominio de la lengua se produzca. En este caso, partimos de un tipo de usuario de la lengua y nuestro fin es eminentemente comunicativo y práctico. Para nosotros, este vocabulario fundamental estaría compuesto de una selección de palabras básicas y disponibles, aunque no solo, ya que podríamos complementarlo con otro tipo de unidades, como algunas que contemple el *PCIC*, algunas de vocabulario instrumental, otras unidades pluriverbales, etc.

Partiendo de esta óptica, podemos representar gráficamente la definición del vocabulario fundamental del siguiente modo, como un primer paso que nos ayude a alcanzar la base de lo que sería nuestro repertorio léxico fundamental:

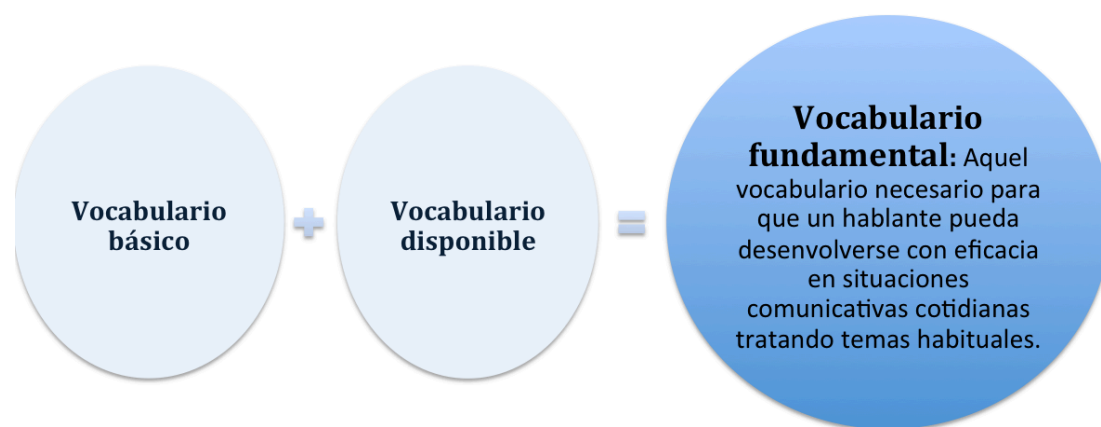


IMAGEN 5. Definición del vocabulario fundamental desde una óptica funcional

Y en un segundo momento lo podemos implementar, por ejemplo, para la enseñanza a estudiantes de una lengua materna próxima al español:

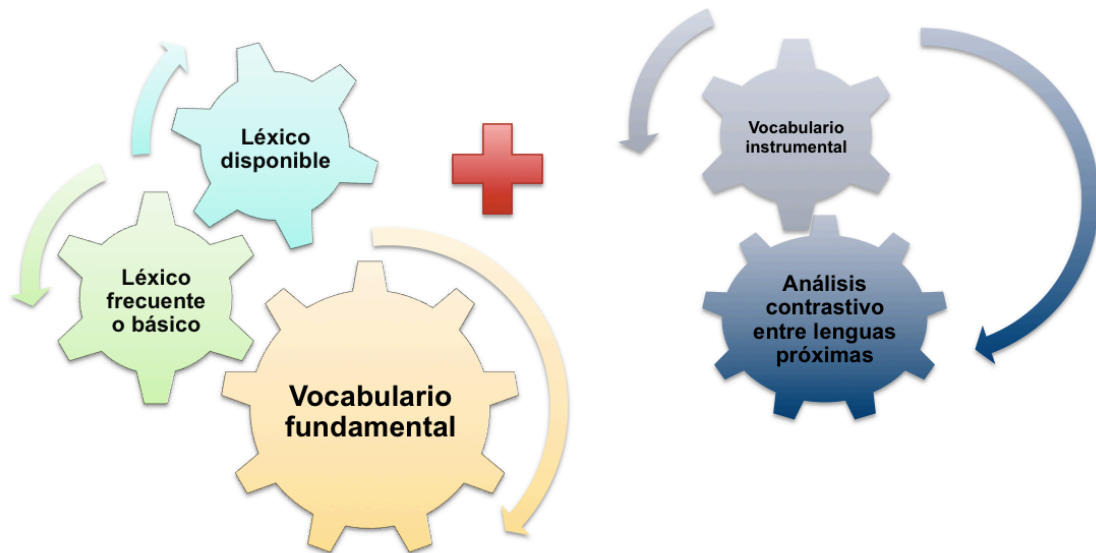


IMAGEN 6. Métodos objetivos de selección léxica

Por otro lado, si observamos la cuestión desde la perspectiva más **analítica**, aunque sin olvidar un fin didáctico, entonces debemos partir de las unidades que componen la lengua meta. En este caso, el vocabulario fundamental no es solo una cantidad concreta de palabras, sino una metodología para clasificar y graduar las unidades léxicas de un idioma siguiendo los niveles del *Marco*. Partimos, pues, de que el léxico es extenso, abierto y complejo de delimitar, por lo que el vocabulario fundamental sería un método de análisis para organizarlo y secuenciarlo en los escalones que se contemplan en el aprendizaje. De este modo, con este vocabulario fundamental intentamos alcanzar las diferentes unidades que componen la lengua y luego las organizamos en función de métodos matemáticos y estadísticos según sean más o menos frecuentes y disponibles. Con esta vía llegamos a los términos más básicos y elementales, los que conforman el núcleo del idioma, al mismo tiempo que separamos el resto de palabras en otros dos grupos, cuyas unidades tienen un nivel de frecuencia y disponibilidad medio y bajo. Con esto logramos poner orden en un campo muy

amplio, etiquetando en varios conjuntos sus componentes de modo objetivo y pormenorizado. Si bien no siempre es fácil llegar a un grado de especificidad tal que nos permita distinguir de forma plausible los 6 niveles del *MCERL*, sí conseguimos separar aquellas palabras más usuales y accesibles para los hablantes de las que no lo son. Desde esta óptica, el vocabulario fundamental resulta más una metodología que un producto, ya que nos ofrece un procedimiento empírico a la hora de establecer una selección y gradación del léxico de una lengua, pensando en la enseñanza a extranjeros sobre todo. Al igual que antes hemos comentado, en este caso también podríamos contemplar otro tipo de léxico y unidades, en función del nivel y el contexto educativo, y no solo las disponibles y frecuentes.

Si representamos gráficamente esta concepción del vocabulario fundamental, el resultado se puede plasmar de esta manera:

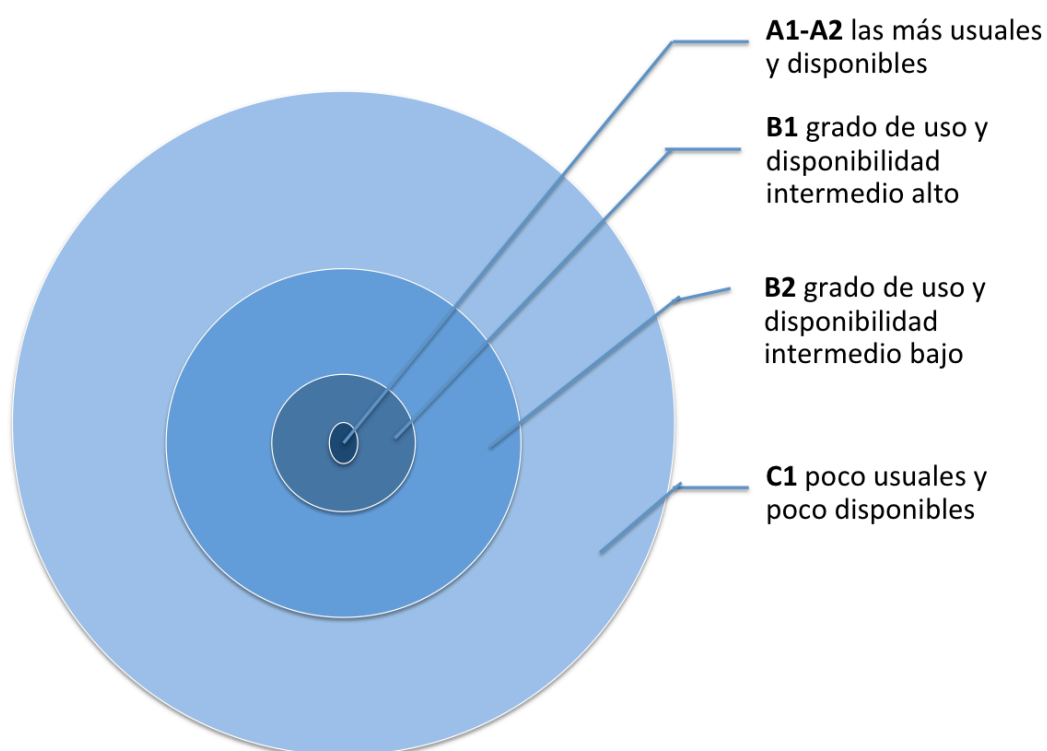


IMAGEN 7. Definición del vocabulario fundamental desde una óptica analítica

En esta imagen, el tamaño y la intensidad en el color de cada segmento muestran la amplitud y la importancia del vocabulario que representan, ya que

después de haber observado los diferentes tipos de unidades hemos concluido que las palabras que conforman el núcleo de una lengua son una cantidad mucho menor, mientras que las que están en el tercer y cuarto círculo son más numerosas y forman parte de lo que se podrían denominar la periferia. Las del núcleo son las más frecuentes y disponibles, por el contrario, en la periferia están las menos usuales y menos disponibles. Entre ambos conjuntos se observa una tendencia gradual de aumento en número de palabras y descenso en nivel de importancia. Esta evolución da un salto más brusco entre los niveles B2 y C1, como después mostraremos.

En el caso de los términos frecuentes, son aquellos que obtienen una mayor cobertura léxica textual, por lo tanto, solemos usarlos en una gran cantidad de textos. La relación entre ambos aspectos varía en función del corpus con el que trabajemos (si es oral o escrito, los registros que recoja, los conjuntos que se establezcan y la forma de hallar las palabras usuales), pero en general sabemos que si hablamos solo de frecuencia (es decir, un corpus de frecuencias léxicas) y en cálculos de Davies (2006), con las 1.000 palabras más frecuentes podemos alcanzar entre un 76% y un 80% de cobertura textual, si se trata de la escritura, y un 88% de cobertura textual, si se trata de la lengua hablada. Con las 3.000 más frecuentes nos situamos entre el 88% y el 90%. Como hemos visto en el capítulo dedicado a la frecuencia léxica, estos cálculos varían en función de los autores, tipos de textos, etc., pero en general entre todos ellos se observa una gran coherencia. Entre las primeras posiciones de las listas de frecuencia, sobre todo en las primeras 1.000, se concentran la mayoría de las palabras gramaticales.

Al hablar de las palabras disponibles, la tendencia en cuanto al índice de disponibilidad y a su grado de frecuencia es también la misma, es decir, las palabras más disponibles son muy escasas, mientras que el conjunto de las poco disponibles es muy numeroso.

En nuestro caso, hemos optado por recorrer ambos caminos, el *funcional* y el *analítico*, y hemos llegado a un punto de encuentro entre ellos en el nivel B1. Por consiguiente, hemos tomando como meta hallar el contenido léxico para este

nivel lingüístico. Así, si por una parte hemos establecido unos criterios de gradación y selección de los vocabularios básicos y disponibles —respondiendo a las preguntas de investigación ¿cuántas palabras seleccionamos? y ¿cómo las seleccionamos?—, por otra vía hemos analizado las unidades concretas que componen el vocabulario fundamental del español para un nivel umbral, cuya elaboración es uno de nuestros objetivos principales.

8.2. Análisis y comparación de cuatro repertorios: dos vocabularios disponibles, el vocabulario básico temático de Davies y las nociones específicas del Plan curricular del Instituto Cervantes (PCIC)⁶⁶

Después de hacer un estudio bibliográfico sobre los vocabularios básicos españoles y los diferentes corpus de frecuencia, así como de los diversos trabajos y avances del campo de la disponibilidad léxica, hemos analizado qué unidades léxicas los componen en concreto. Para ello las hemos comparado con las que el *PCIC* establece para los niveles iniciales e intermedio hasta el nivel umbral, es decir, los estadios A1, A2 y B1. Tomamos el *PCIC* como el texto de referencia para la selección de contenidos en la enseñanza-aprendizaje de ELE, por lo que nos sirve como punto de partida y como modelo para contrastarlo con el vocabulario que obtenemos con los diccionarios disponibles y básicos. El léxico frecuente temático de Davies (2006) sirve como una referencia más, aunque somos conscientes de que parte de sus contenidos quedan fuera de las 3.000 palabras frecuentes que contemplamos para nuestro trabajo (*vid. infra* § 8.3.2.1.). Las cuatro obras que hemos usado para esta comparación son: López Meirama (2008), Tomé Cornejo (2015), Davies (2006) e Instituto Cervantes (2006)⁶⁷.

⁶⁶ Para observar los datos y los comentarios de estos análisis con una mayor profundidad recomendamos la consulta del Anexo 4.

⁶⁷ López Meirama, B. (2008). *Léxico disponible en el español de Galicia*. Santiago de Compostela: Universidade de Santiago de Compostela, Colección Lalia (Series Maior, nº 22). (Hemos usado la versión en línea, disponible con contraseña y alta de usuario en la página web de la USC).

Tomé Cornejo, C. (2015). *Léxico disponible. Procesamiento y aplicación a la enseñanza de ELE*. Tesis Doctoral defendida en el curso 2014-2015 en la Universidad de Salamanca.

Davies, M. (2006). *A Frequency Dictionary of Spanish: Core Vocabulary for Learners*. New York: Routledge.

Instituto Cervantes (2006). *Plan Curricular del Instituto Cervantes. Niveles de referencia para el español*. Madrid: Instituto Cervantes / Edelsa, Biblioteca Nueva (3 vols.).

La comparación no siempre es fácil, ya que en muchas ocasiones no se producen equivalencias claras entre ellas. La obra de López Meirama y el trabajo de Davies, por ejemplo, cuentan con menos núcleos temáticos, mientras que la propuesta de Tomé Cornejo ha ampliado sus centros de interés a 47 y ha tenido siempre como punto de referencia el *PCIC*, por lo que es más fácil establecer líneas de unión por temas entre estos dos últimos repertorios. En nuestro caso, hemos partido de los 16 campos semánticos con los que se trabajaba en los estudios de léxico disponible tradicionales y hemos buscado sus equivalentes en el resto de obras siempre que ha sido posible. Hemos obviado el CI 09. *Iluminación, calefacción y medios de airear un recinto*, ya que no es un tema que se considere muy productivo ni que incluyan el resto de repertorios. En el léxico disponible consideramos los vocablos que obtienen un índice de disponibilidad igual o superior a 0,1. En el caso del *PCIC* hemos limitado la comparación al nivel B1, por lo que vaciamos los contenidos de las nociones específicas de los niveles A1, A2 y B1. En el trabajo de Davies, nos limitamos a las palabras que esta obra agrupa por temas, aunque solo seleccionamos las que están dentro de las 5.000 palabras más frecuentes. Los temas que hemos valorado y su correspondencia entre obras son los siguientes:

Léxico Disponible de López Meirama	Léxico Disponible de Tomé Cornejo	Léxico Frecuente de Davies	Nociones específicas PCIC (A1-A2 y B1)
1. El cuerpo humano	1. El cuerpo humano: partes, acciones y posturas	<i>Body</i>	1.1. Partes del cuerpo 1.3. Acciones y posiciones que se realizan con el cuerpo
2. La ropa	28. En la tienda de ropa	<i>Clothing</i>	12.1. Ropa, calzado, complementos
3. Partes de la casa (sin muebles) 4. Los muebles de la casa	21. La casa y las tareas domésticas 20. Buscar una vivienda	No se contempla	10. Vivienda 10.1 Acciones relacionadas con la vivienda 10.2. Características de la vivienda 10.3. Actividades domésticas 10.4. Objetos domésticos
5. Alimentos y bebidas 6. Objetos colocados en la mesa para la comida 7. La cocina y sus utensilios	8. Alimentación y nutrición 9. Preparar una comida 10. En el restaurante 11. En el bar	<i>Food</i>	5.1 Dieta y nutrición 5.2 Bebida 5.3. Alimentos 5.4 Recetas 5.5 Platos 5.6. Utensilios de cocina y mesa 5.7. Restaurante

8. La escuela: muebles y materiales	12. El sistema educativo: etapas, personal y centros de enseñanza 13. En clase	No se contempla	6. Educación 6.1. Centros e instituciones educativas 6.2. Profesorado y alumnado 6.3. Sistema educativo 6.4. Aprendizaje y enseñanza 6.5. Exámenes y calificaciones 6.6. Estudios y titulaciones 6.7. Lenguaje de aula 6.8. Material educativo y mobiliario de aula 11.5. Servicios educativos (Carmela en su tabla de equivalencias entre sus CI y el PCIC no incluye el 6.7. y el 6.8.)
10. La ciudad 11. El campo 13. Trabajos del campo y del jardín 14. Los animales	27. La ciudad y el campo 42. La geografía y los accidentes geográficos 43. El tiempo atmosférico 44. Animales y plantas 45. Medioambiente y desastres naturales	<i>Animals</i> <i>Natural</i> <i>Features and plants</i> <i>Weather</i>	11.8. Servicios de abastecimiento público 12.1. Lugares 20.2. Geografía 20.3. Espacios urbanos o rústicos 20.4. Clima y tiempo atmosférico 20.5. Fauna 20.6. Flora 20.7. Problemas medioambientales y desastres naturales
16. profesiones y oficios	46. Profesiones y oficios 14. El mundo laboral	<i>Professions</i>	7.1 Profesiones y cargos 7.2. Lugares, herramientas y ropa de trabajo 7.3 Actividad laboral 7.4 Desempleo y búsqueda de trabajo 7.5. Derechos y obligaciones laborales 7.6. Características de un trabajador 15.4. Entidades y empresas

TABLA 18. Equivalencias entre los temas considerados en López Meirama (2008), Tomé Cornejo (2015), Davies (2006) y el PCIC

Una vez seleccionados los temas y las unidades léxicas que se ajustan a estos parámetros, las ordenamos alfabéticamente y en columnas, para poder observar sus encuentros y divergencias. Así, cada listado se coloca en paralelo, pero usando diferentes colores en la fuente para facilitar su identificación y comparación a simple vista: hemos usado el color negro para el léxico disponible de Galicia, el verde para los vocablos obtenidos de Tomé Cornejo, el rojo para el vocabulario frecuente de Davies y el azul para las nociones específicas y estructuras que ofrece el *PCIC* (*vid.* Anexo 4).

También para hacer visible y rápido el cotejo, hemos utilizado sombreados: con el sombreado gris marcamos las palabras repetidas en todas las columnas, con sombreado azul las unidades del *PCIC* para las que no hay ninguna equivalencia en el resto de obras, con sombreado rojo las palabras frecuentes del volumen de Davies que no aparecen en el resto de obras, con sombreado verde las unidades de Tomé Cornejo que no tienen correspondencia en los otros trabajos y con sombreado gris claro los vocablos del *Léxico disponible en el español de Galicia* que no se encuentran en ninguno de los otros tres proyectos.

Aunque llevemos a cabo un recuento cuantitativo, esto es, aunque valoremos cuántas unidades obtenemos en uno y otro texto, nuestro objetivo primordial es un análisis cualitativo, es decir, determinamos qué elementos léxicos encontramos en cada una de estas obras en concreto, así como las correspondencias que observamos entre ellas. Identificar y cotejar estas obras nos permite comprobar las convergencias y divergencias entre vocabularios, es decir, descubrir qué elementos de los que conforman el léxico disponible y de los que están en las primeras posiciones por su índice de disponibilidad se encuentran o no en el *PCIC* y, a la vez, si hay algún parecido con los núcleos temáticos que Davies creó a partir de su corpus de frecuencias. Nos interesa también comparar los resultados entre los estudios de disponibilidad léxica, pues en uno se trabaja con los 16 centros de interés tradicionales del Proyecto Panhispánico, mientras que en la nueva propuesta de la investigadora Tomé Cornejo los centros de interés se modifican y amplían. Para ello, en ocasiones

cotejamos los dos listados de disponibilidad por separado, o bien tenemos en cuenta uno u otro a la hora de ver su parecido con el *PCIC*. En el caso de Tomé Cornejo, contamos además con su propuesta de selección de vocabulario, en la que, siguiendo procedimientos diferentes a los nuestros, ha comparado sus listas de léxico disponible con el *PCIC*, organizando el vocabulario en función de las nociones específicas y subapartados que este último texto contempla y viendo qué vocabulario disponible podría encajar en ellos. En algunas ocasiones hemos valorado también este modo de operar, para observar si el parecido era mayor y para obtener datos de los índices de disponibilidad que le sirven de corte con su método.

Partimos, además, de que el *PCIC* es un referente básico ya asentado entre la comunidad investigadora y docente, y de que la clasificación de los temas y sus epígrafes son un modo útil de organizar la enseñanza de ELE. No obstante, nos cuestionamos si se obtiene un inventario parecido usando otro tipo de métodos objetivos y explícitos, tanto para reflexionar sobre las unidades que se han incluido en los niveles de *PCIC* como para implementarlo. No olvidemos que el *PCIC* no deja de proponer listas abiertas e incompletas, que sientan unas bases que debemos enriquecer y adaptar en cada caso. Así, aunque no se explica cómo han sido escogidas las diferentes nociones, antes de la exposición del inventario de nociones específicas, en la Introducción del apartado 9, se pueden leer unas líneas que manifiestan:

No cabe duda de que esta manera de parcelar y clasificar el mundo es arbitraria. De hecho, en algunos casos una misma unidad léxica aparece en dos temas, pues no son taxonomías de límites discretos. Se trata de una presentación que ha permitido incorporar los contenidos léxicos necesarios para cumplir los objetivos de cada nivel, pero que debe consultarse, en todo caso, como *una lista de carácter orientativo y no como un resultado final e inamovible. A este respecto, el inventario debe considerarse abierto en los distintos sentidos que se exponen a continuación.*

Las entradas —exponentes aislados o series de exponentes— que se presentan en los distintos epígrafes del inventario han sido seleccionadas con el objetivo de dar una idea del tipo de unidades léxicas que el hablante debe conocer en cada uno de los niveles, pero no son una lista cerrada. En función del análisis que se haga de las necesidades y expectativas de los alumnos, así como de las características de cada situación particular de enseñanza, podrán añadirse nuevas entradas de exponentes en uno u otro epígrafe, de manera que pueda darse una respuesta adecuada a los requerimientos de un currículo o programa determinado. Por otra parte, debe entenderse que *las series de exponentes que se*

presentan en cada una de las entradas son también abiertas, es decir, admiten otros elementos que pueden completar la serie del modo que sea conveniente a la situación y a las circunstancias pedagógicas. En este sentido, deberían aparecer puntos suspensivos al final de prácticamente todas las series de exponentes. Por razones de presentación, para no hacer molesta la lectura, se ha optado en este caso por no incluir los puntos suspensivos, si bien habrá de tenerse en cuenta el carácter no cerrado de las especificaciones que se presentan. [PCIC, Centro Virtual Cervantes. La cursiva es nuestra]

En cuanto a la propuesta de Tomé Cornejo (2015: 440), la propia investigadora pone a prueba su método de selección del vocabulario llevando a cabo dos acciones⁶⁸:

- un incremento del número de centros de interés que obliga a la realización de varias pruebas sobre muestras de características sociológicas similares para no alargar en exceso la duración de la encuesta.
- la utilización de centros de interés de gran amplitud semántica y naturaleza heterogénea que susciten la evocación de diferentes clases de palabras y de diversas agrupaciones.

No obstante, el procedimiento de Tomé Cornejo para comparar su léxico disponible con el que contempla el *PCIC* no puede considerarse demasiado objetivo, ya que elabora listados buscando las palabras de un tema concreto dentro de las respuestas de todo el centro de interés, independientemente de si están en el puesto 4 o en el 40 o de su índice de disponibilidad, de su aparición y su frecuencia. Lo que se hace en realidad es buscar equivalencias entre este listado y los subtemas del *PCIC* sin tener en cuenta los cálculos matemáticos de su índice de disponibilidad, o los porcentajes de frecuencia. A las agrupaciones temáticas del trabajo de Davies (2006) se les podría hacer una crítica similar, aunque a diferencia de aquella, este indica la frecuencia que tiene cada palabra, y lo hace con elementos concretos, no con todo su vocabulario. De todas formas, ambos métodos resultan poco válidos a nivel científico.

⁶⁸ Ambas tareas, las realiza teniendo como punto de referencia básico los temas y subtemas del *PCIC*, dado que esta obra constituye en la actualidad el punto de partida de toda iniciativa relacionada con el diseño y desarrollo de cursos y materiales de español como lengua extranjera (Tomé Cornejo 2015: 440). Aunque ello le supuso ampliar los núcleos temáticos a 47 y distribuirlos en 4 modelos diferentes de encuesta, como hemos explicado en otro capítulo.

Sea como fuere, entendemos que esta estrategia de comparación le sirve a Tomé Cornejo como primera aproximación en el cotejo de ambos listados para probar la validez de sus nuevos centros de interés para la selección léxica. La propia autora presenta su trabajo como una propuesta piloto y comenta que ha ido haciendo modificaciones en función de los resultados que ha obtenido, sin que estos sean definitivos. En este sentido, para nuestro objetivo han sido unos datos muy útiles, pues así podemos llevar a cabo una selección de vocabulario más completa, con diferentes aportaciones y puntos de referencia. Compartimos las conclusiones a las que esta estudiosa ha llegado:

En cuanto al tipo de vocabulario obtenido, no ha de entenderse como un inventario alternativo al del *PCIC*. Los listados difieren porque se han elaborado con distintas metodologías, y los vocablos disponibles son el resultado de una tentativa, que no pretendía ofrecer resultados definitivos. El carácter piloto de la encuesta ha propiciado, por ejemplo, la aparición de respuestas atípicas, que se eliminarían si se aumentara el tamaño de la muestra y se llevara a cabo una buena estratificación previa. [Tomé Cornejo 2015: 446]

En cualquier caso, un primer acercamiento revela que el grado de coincidencia entre los listados varía también en función del área temática que se considere, como así se muestra en la tabla siguiente. En ella exponemos los datos sobre las unidades totales que hemos contabilizado en los cuatro repertorios dentro de cada tema. Incluimos, además, las similitudes entre ellas en la columna de la derecha:

Obras	Léxico Disp. López Meirama	Léxico Disp. Tomé Cornejo	Léxico Frec. Davies	Nociones específ. PCIC (A1- A2 y B1)	Coincide ncias entre los cuatro repertori os
Temas					
1. El cuerpo humano	39	43	58	58	20
2. La ropa	33	36	19	47	5
3. Partes de la casa (sin muebles)	47	45	∅	112	8
4. Los muebles de la casa					
5. Alimentos y bebidas 6. Objetos colocados en la mesa para la comida 7. La cocina y sus utensilios	89	110	49	171	17
8. La escuela: muebles y materiales	32	56	∅	129	10
10. La ciudad 11. El campo 13. Trabajos del campo y del jardín 14. Los animales	95	139	107	218	20
12. Medios de transporte	22	25	28	103	4
15. Juegos y distracciones	18	26	21	104	0
16. Profesiones y oficios	31	66	81	116	4

TABLA 19. Número de palabras contabilizadas en los núcleos temáticos de López Meirama (2008) y Tomé Cornejo (2015) con un índice de disponibilidad \geq a 0,1, dentro de las 5.000 más frecuentes en Davies (2006) y las pertenecientes a los niveles A1-A2 y B1 en el PCIC. Número de unidades coincidentes entre los cuatro repertorios

El cuerpo humano

Las coincidencias totales entre las cuatro obras ascienden a 20 ítems. Entre las dos obras de léxico disponible (López Meirama y Tomé Cornejo) y el

léxico frecuente (Davies) hay 27 términos iguales, aunque si solo contamos uno cualquiera de los léxicos disponibles la cantidad es mayor. Entre los dos léxicos disponibles hay 31 equivalencias. Las palabras diferentes, es decir, las que solo se ofrecen en uno de los repertorios, suman 27 en el *PCIC*, 12 en Davies, 5 en Tomé Cornejo y 4 en el léxico disponible del español de Galicia.

La ropa

Existen pocas similitudes en este tema: 5 entre las cuatro obras, 9 entre las dos de léxico disponible y el léxico frecuente, 8 entre el léxico frecuente y el *PCIC*. Las divergencias, es decir, las palabras que no encuentran equivalencia en el resto de volúmenes, son mayores: 22 en el *PCIC*, 12 en el léxico disponible de Tomé Cornejo, 5 en el léxico disponible de López Meirama y 4 en el léxico frecuente.

Partes de la casa (sin muebles) y Los muebles de la casa

Muy pocas similitudes: hay 8 casos entre los tres y 9 entre los dos léxicos disponibles. En cambio, hay 66 palabras que únicamente están en el *PCIC*, 19 que solo están en el recuento de Tomé Cornejo y 18 en el de López Meirama. Hay pues, muchas diferencias entre ellos. Resalta el número total de ítems que se registran en el *PCIC* (112) frente a los del léxico disponible (47 y 45).

Alimentos y bebidas, Objetos colocados en la mesa para la comida y La cocina y sus utensilios

En general se observan muchas diferencias. Solo hay 17 casos de palabras que coinciden en las 4 obras. Las semejanzas entre las dos obras de léxico disponible son considerables, suman 47 casos. Entre una de las obras de léxico disponible y la de Davies suman 28. Las palabras que no tienen equivalencia en las otras obras son 103, una cifra muy alta. Suman 22 en el caso de López Meirama y 29 en el de Tomé Cornejo. En Davies hay 15. La mayor distancia está entre el *PCIC* y el resto de obras, por lo tanto.

La escuela: muebles y materiales

Se observa gran divergencia entre las dos obras de léxico disponible y el *PCIC*. En las primeras el número de palabras es mucho menor (32 y 56) mientras que en el *PCIC* tenemos 129. De estas, hay 91 sin correspondencia, frente a las 10 unidades que son coincidentes entre las tres obras. Entre los dos léxicos disponibles solo hay 19 correspondencias. Las palabras sin equivalencia son 17 en Tomé Cornejo y 9 en López Meirama.

La ciudad, El campo, Trabajos del campo y del jardín y Los animales

En un total de casi 400 palabras solo se observan 20 coincidencias entre las cuatro obras. Entre los dos léxicos disponibles en el tema *La ciudad y el campo* hay 24 palabras diferentes de un total de 50 palabras en el léxico disponible del español de Galicia y 6 vocablos sin equivalencia de 33 en el de Tomé Cornejo. En general, en el *PCIC* no tienen correspondencia 131 palabras, en el léxico frecuente 36, en el de López Meirama 23 y en el de Tomé Cornejo 42. Sobresale el grado de divergencia del *PCIC* con el resto, como estamos viendo en otros centros de interés.

Medios de transporte

Las similitudes entre los cuatro trabajos son 4 casos. La mayor distancia se observa entre el *PCIC* y las otras tres, pues la diferencia de términos totales es muy elevada. Además, de las 103 nociones del *PCIC* no tienen equivalencia 83 casos. En la obra de López Meirama ocurre en 9, en la de Tomé Cornejo en 17 y en el léxico frecuente en 13 casos.

Juegos y distracciones

Llama la atención que en este caso no haya ninguna coincidencia entre los cuatro volúmenes y solo existen 3 entre las dos obras de léxico disponible.

Profesiones y oficios

Entre todos los trabajos hay solo 4 coincidencias. Ascenden a 16 similitudes si tenemos en cuenta solo tres volúmenes. Entre los dos trabajos de léxico disponible hay 25 equivalencias en total. Las palabras sin correspondencia

en ellas son 3 en el caso de López Meirama y 19 en el de Tomé Cornejo. Las grandes divergencias están en el *PCIC* y en el léxico frecuente de Davies, pues de 116 términos hay 97 palabras sin equivalencia y de 81 contabilizamos 50 diferencias, respectivamente.

Parece interesante subrayar la complementariedad entre las fuentes que hemos elegido, pues, al tratarse de métodos diversos para la obtención del léxico, observamos que hay pocos solapamientos que afecten a las unidades disponibles y frecuentes. Globalmente, debido al criterio acumulativo de las nociones específicas, encontramos gran número de vocablos, que contrasta con las cantidades más reducidas —en la mayor parte de los centros de interés menos de la mitad—. Con todo, la complementariedad de estos recuentos es patente, desde el momento en que las coincidencias son también muy reducidas. La conclusión es la conveniencia de conjugar estos repertorios: obtenemos unidades muy disponibles y muy frecuentes que no están contenidas en los ejemplos de las nociones específicas del *PCIC* e indudablemente sería bueno considerar en un vocabulario que se dice fundamental. Por otra parte, dentro del léxico temático (lo que también llamamos palabras léxicas) obtenemos vocablos diferentes según procedan de los diccionarios de disponibilidad o de los listados de frecuencia; por tanto, también en esta parcela podemos hablar de complementariedad. Por último, entre los dos repertorios de léxico disponible que hemos tenido en cuenta, también es productiva la fusión, desde el momento en que ambos aportan unidades en exclusiva, debido, seguramente, a la diferencia en los grupos de informantes, alumnado del último curso de Secundaria en López Meirama (2008) —de acuerdo con las pautas del PPHDL— y estudiantes de primer curso de universidad (Filología) en Tomé Cornejo (2015). Nótese que los resultados son ligeramente más cuantiosos en general en los listados de Tomé Cornejo (2015).

8.3. Aportaciones para obtener un vocabulario fundamental del español

En secciones anteriores (*vid. supra* § 7.6.1) nos hemos centrado en analizar las definiciones que se han ofrecido de los léxicos fundamentales y hemos aportado lo que, a nuestro modo de ver, debe ser un vocabulario

fundamental del español (*vid. supra* § 8.1). Recordemos que para nosotros este concepto se puede abordar desde dos puntos de vista diferentes, uno analítico y otro funcional, aunque ambas visiones pueden converger si nos centramos en un nivel umbral (B1), en el cual un usuario de la lengua debe desenvolverse con eficacia en situaciones comunicativas cotidianas tratando temas habituales. Esto nos sitúa en el grado mínimo de dominio que se debe alcanzar para que un hablante pueda considerarse “independiente”.

Como hemos venido explicando, una de las virtudes de elaborar un vocabulario fundamental es la posibilidad de usarlo a la vez como método de selección léxica. Por eso nuestro trabajo adopta una perspectiva más amplia, sin desviar el foco de su objetivo concreto. En consecuencia, antes de afrontar la elección de las obras con las que alcanzar el vocabulario fundamental y su posterior selección y fusión de palabras, hemos realizado una revisión de los distintos métodos de selección léxica que se contemplan en la bibliografía sobre lexicografía y en el ámbito de la enseñanza de lenguas (*vid. supra* § 3.3 y 3.4.). Esta tarea nos ha hecho recorrer caminos variados, como el estudio de fuentes lexicográficas (los corpus de frecuencias, listas de palabras, etc.) (*vid. supra* § 5.1., 5.2. y 5.3.), las publicaciones que arrojan datos sobre la comprensión léxica por parte de estudiantes de una lengua extranjera (umbral mínimo de comprensión) (*vid. supra* § 4.4.), las cantidades de palabras que los hablantes usan a diario y aquellas que son necesarias para realizar las tareas fijadas en los distintos niveles de dominio de la lengua, así como el análisis de diferentes tipos de vocabulario (*vid. supra* § 4.3. y 4.5.), o bien trabajos que tratan cuestiones léxicas relacionadas con la lingüística aplicada y las teorías sobre la enseñanza de lenguas extranjeras (*vid. supra* § 2.1. y 2.2.). Ha sido después de estas incursiones en diversas áreas, que van desde la enseñanza de lenguas hasta la lingüística del corpus, cuando hemos afrontado la creación del vocabulario fundamental del español.

En los siguientes apartados describiremos la metodología seguida para conseguir este objetivo, así como la bibliografía con la que hemos trabajado y los procesos desarrollados.

8.3.1. Fuentes usadas

A) Proceso de selección de las obras de vocabulario básico

El primer paso consistió en hacer una valoración global de los diferentes vocabularios básicos y diccionarios de léxico disponible de las últimas décadas. El objetivo era seleccionar apenas una obra de cada tipo de vocabulario para probar un método de selección y fusión de sus unidades léxicas. Como se refleja en el Capítulo 5, consideramos todas las propuestas relevantes de vocabulario básico del español desde los inicios del siglo XX hasta la actualidad, con lo que llegamos a contemplar más de una veintena de obras. Con este recorrido, al mismo tiempo, hemos conocido los orígenes, evolución y características de este tipo de trabajos, además de apreciar los distintos usos y diferencias entre los corpus textuales, los listados de frecuencias y los vocabularios básicos. Al final obtenemos una selección de aquellas obras que por su modo de proceder o críticas positivas recibidas merecían una mayor atención: García Hoz (1953), Juilland y Chang-Rodríguez (1964), Díaz Castañón (1977), Morales (1986), Alameda y Cuetos (1995), Ávila Muñoz (1999), Sebastián, Cuetos, Martí y Carreiras (2000) y Davies (2006).

En algún caso tuvimos que desechar la idea de tomarlas como base de nuestro estudio por tratarse de trabajos ya obsoletos, como ocurre con los volúmenes más antiguos. Las obras de Buchanan (1927), Eaton (1940), Rodríguez Bou (1952), García Hoz (1953) y Juilland y Chang-Rodríguez (1964) se basan en textos de la década de los 50 o fechas anteriores. Aunque el trabajo de Juilland y Chang-Rodríguez gozó de un gran prestigio por su buen hacer, como demuestra el hecho de que se haya tomado como modelo para obras de este tipo en otras lenguas, las fechas de los textos que conforman su corpus y el año de su publicación lo dejan fuera de nuestra consideración. Además, hoy por hoy el número de palabras de las que parte para su recuento, menos de un millón, resulta insuficiente. A lo que hay que añadir que se centra solo en la lengua escrita. Estos primeros proyectos seleccionan solo textos escritos, por lo que la lengua representada no era el español general. El trabajo de Morales (1986) es también parcial, ya que solo contempla el español de Puerto Rico, así como el de Ávila Muñoz (1999), que se centra en el español hablado de Málaga.

En realidad, esta búsqueda nos llevó a observar la escasez de vocabularios básicos actuales en nuestro idioma que retraten el español de la forma más completa posible, pues el proyecto que pretendía suplir este vacío vio frustrada su realización (Benítez Pérez, Hernández Cabrera y Samper Padilla 1995). Finalmente, optamos por considerar la obra de Davies (2006), pues además de ser más reciente, incorpora una gran pluralidad de textos, dominios y registros. Unido a esto, uno de los formatos en los que se pueden consultar sus datos es electrónico, lo que facilita enormemente su manejo. Las líneas introductorias del volumen especifican las fórmulas matemáticas y los cálculos que se han llevado a cabo para hallar el índice de frecuencia de las 5.000 unidades que contabiliza, además de otra información relevante como su rango, el registro, etc. Hemos descrito en detalle las características de esta obra en el Capítulo 5 (§ 5.2). La versión más reciente del *Corpus del Español* es del año 2016. Amplía el corpus a 2.000 millones de palabras, por lo que es 100 veces mayor que la sección inicial del siglo XX que había en la antigua versión. Con esto se consigue aumentar los resultados en las búsquedas. Las facilidades de consulta, búsqueda y comparación de unidades están mejoradas. Es posible comparar frecuencias de palabras, sintagmas y construcciones sintácticas entre los 21 países hispanohablantes incluidos en el corpus. Los textos, además, son más recientes (tomados de los últimos tres o cuatro años).

Otra obra que obtuvo nuestra atención fue *El léxico de frecuencia del español hablado en la ciudad de Málaga* (Ávila Muñoz 1999). En este caso, a pesar de ser un trabajo sobre la lengua hablada y localizado únicamente en un punto concreto de la geografía española, se debe resaltar el hecho de que trabaje con un criterio de lo que debe ser la unidad básica más actual y plural, ya que considera las unidades léxicas pluriverbales y no solo las palabras gráficas. En este sentido, este libro es de consulta recomendada para conseguir aislar algunas locuciones mediante su frecuencia, mientras no se publiquen obras más completas, con rigor metodológico y técnicas más avanzadas en el etiquetado y procesamiento de los datos.

Precisamente, una de las dificultades con las que nos hemos encontrado, además de la falta de obras de este tipo entre las que poder escoger, ha sido la

imposibilidad de seleccionar unidades pluriverbales, más en concreto colocaciones. Si bien un estudio minucioso sobre el tipo de unidades léxicas que debemos considerar en el proceso de enseñanza-aprendizaje de una lengua nos llevó a la conclusión de que las colocaciones son un elemento importante desde un primer momento, también nos encontramos con el problema de que ningún diccionario de frecuencias tenía en su base la lematización de estas multipalabras, por lo que no conseguimos seleccionarlas teniendo en cuenta cuáles son más usuales y cuáles menos. Acometer un proceso de selección que se centre también en las colocaciones conlleva un esfuerzo de investigación considerable incluso en los corpus en los que se permite la búsqueda de más de una palabra usando el criterio de las coapariciones, como el CORPES o el *Corpus del Español* de Davies. Por consiguiente, este aspecto queda fuera de nuestro objetivo, aunque reconocemos la importancia de acometerlo a posteriori para quien quiera implementar el Vocabulario Fundamental que aquí ofrecemos. De todos modos, sí hemos considerado oportuno citar algunas fuentes y criterios a la hora de escoger aquellas colocaciones y multipalabras adecuadas para los niveles iniciales.

En conclusión, la obra escogida para acometer la selección del léxico básico del español ha sido la de Davies (2006). Asimismo, ofrecemos vías de selección de las colocaciones y las multipalabras en general, aunque no las incluyamos en nuestro vocabulario por completo.

B) Proceso de selección de las obras de vocabulario disponible

Para el léxico disponible la elección resulta más fácil, ya que a raíz del PPHDL existen muchos diccionarios de léxico disponible que se han llevado a cabo con los mismos patrones. En este sentido, cualquier obra de las más recientes que siga estos criterios generales sirve para realizar una prueba de selección y fusión. En un primer momento, decidimos acometer nuestro proyecto con el *Léxico disponible en el español de Galicia* de López Meirama (2008). Uno de los motivos de la elección fue la página web con la que cuenta el proyecto Dispogal⁶⁹, pues consultar el diccionario en línea facilita enormemente la

⁶⁹ Accesible en <http://gramatica.usc.es/dispogal/>

accesibilidad a los datos y su manejo. En realidad, la difusión electrónica de este diccionario de léxico disponible es algo innovador e inusual, pues en la mayor parte de las veces los volúmenes se publican en papel y están muy poco difundidos. En escasas ocasiones, forman parte de alguna tesis o proyecto de investigación con acceso en un formato de documento de texto. Por supuesto, para la elección del léxico disponible del español de Galicia también fue decisiva la calidad y representatividad del repertorio.

Como más adelante comentaremos, las primeras pruebas para escoger el vocabulario de cada centro de interés las llevamos a cabo con un criterio que habían seguido ya algunos investigadores anteriores cuando compararon diferentes estudios de léxico disponible. Estos estudios, centrados sobre todo en el cotejo de diferentes sintopías o de pruebas hechas a nativos frente a otras con extranjeros, solían servirse de un límite de 0,1 de índice de disponibilidad como punto de corte en cada lista de los diferentes centros de interés. Al llevar a cabo estas calas, salió a la luz uno de los primeros problemas: este criterio, establecido de forma arbitraria, dejaba a veces fuera palabras que en otros estudios de disponibilidad estaban entre las de mayor o igual índice de disponibilidad a 0,1 o al contrario. Debido a esto, decidimos considerar varios diccionarios de léxico disponible de lugares dispares de la geografía española, y optamos por incluir los vocabularios de Alicante (Martínez Olmos 2015), Salamanca (Gallos Camacho 2003) y Ceuta (Ayora Esteban 2006). En archivos *Excel* colocamos las palabras escogidas en paralelo para poder observar diferencias y similitudes, y así conseguir incluir en nuestro grupo de palabras escogidas aquellas que estaban rozando la frontera del 0,1 en el léxico disponible de Galicia. Pero más adelante optamos por abandonar esta vía al encontrar un criterio de selección del léxico disponible diferente, por lo que volvimos a centrarnos en el trabajo de López Meirama y a no considerar los otros tres.

No obstante, cuando seleccionamos el léxico disponible y establecimos los diferentes núcleos temáticos que un vocabulario fundamental debería tener, fuimos conscientes de que los centros de interés tradicionales resultan claramente insuficientes y limitados para este fin. Del mismo modo que al estudiar los trabajos del léxico frecuente del español observamos las carencias

existentes en ese campo, llegamos a esa misma conclusión en el ámbito de la disponibilidad léxica. Como hemos comentado en diferentes momentos a lo largo de la tesis, una de las grandes limitaciones de las investigaciones en disponibilidad léxica es la falta de actualización y ampliación de la nómina de centros de interés, así como su aplicación a la enseñanza de una lengua extranjera. Tenemos en cuenta que en la actualidad contamos con documentos de referencia para ello, como son el *MCERL* y el *PCIC*, en los que la lista de temas es más numerosa y plural que las 16 nociones con las que se ha trabajado en casi todos los proyectos del PPHDL. Conscientes de este problema, acogimos la reciente propuesta de la investigadora Tomé Cornejo sobre el procesamiento del léxico disponible con la esperanza de solucionarlo (Tomé Cornejo 2015)⁷⁰. Aunque solo como propuesta piloto, el estudio afronta esta cuestión y propone 47 nuevos centros de interés, que, además de guiarse por los 16 tradicionales, se ajustan a las nociones específicas del *PCIC*. Por eso, con las listas de léxico disponible de Tomé Cornejo conseguimos llevar a cabo la selección léxica de forma más eficaz, pero, al mismo tiempo, cotejamos sus resultados con los diccionarios de léxico disponible anteriores. Otra cuestión es que así podemos probar y aplicar nuestros criterios de selección léxica en dos trabajos de disponibilidad léxica diferentes, lo que nos ofrece una idea más firme de su validez.

Así, una vez que ya decidimos las obras de léxico frecuente y disponible con las que experimentar un proceso de selección y fusión de ambos vocabularios, pasamos a determinar cuántas unidades léxicas necesitamos para cada nivel de lengua y definimos cuáles son los parámetros para escogerlas y fundirlas. Estas son las cuestiones que tratamos en el siguiente epígrafe.

8.3.2. Metodología y criterios de selección

En esta sección explicamos cuántas unidades léxicas escogemos, cuáles en concreto y cómo lo llevamos a cabo.

⁷⁰ Ya hemos visto que su aportación, una tesis doctoral defendida en el curso 2014-2015 en la Universidad de Salamanca sobre el procesamiento y la aplicación del léxico disponible a la enseñanza de ELE, marca un antes y un después en esta línea de investigación.

8.3.2.1. Criterios de selección del vocabulario básico

La mayor parte de los vocabularios básicos establecen una nómina de palabras que varía entre las 5.000 y las 9.000 aproximadamente. ¿Quiere esto decir que son las unidades que debemos enseñar los estudiantes extranjeros? Esta es una de las primeras preguntas que surgen al construir una propuesta de léxico para la docencia de una lengua extranjera: cuántas palabras es recomendable escoger para cada nivel de lengua. Las respuestas a estas cuestiones no son simples, sobre todo teniendo en cuenta que nuestro trabajo se centra en la lengua española, en la que apenas existen estudios que analicen estos temas. En español, más allá de los corpus de frecuencias y vocabularios básicos, no hay, que sepamos, investigaciones sobre la cobertura léxica textual o la capacidad de comprensión oral y escrita por parte de aprendices de la lengua. Por eso, en un primer momento los diferentes diccionarios de frecuencias léxicas, tanto del español como de otros idiomas, pueden ayudar a dilucidar algunas cuestiones como el número de elementos léxicos que existen en una lengua, la cantidad de palabras que usa un hablante medio a diario, las palabras que se suelen clasificar como frecuentes frente a las académicas y poco usuales, etc. Sin embargo, como muestran los Capítulos 3 y 4 de este trabajo, para obtener datos más concretos y estudios más recientes hemos recurrido a la lengua inglesa. En ella, la bibliografía sobre estas cuestiones es considerable y nos ofrece datos valiosos sobre lo que los aprendices de una lengua son capaces de hacer con eficacia en función del tipo y la cantidad de vocabulario que conocen. Igualmente iluminadoras resultan las propuestas de clasificación que Nation establece en sus listas de frecuencia a partir del BNC, *British National Corpus* (Nation 2001), o las que ofrece Coxhead (2000) en el *Academic Word List (AWL)*. Estas listas de palabras, además de resultar útiles para posteriores calas sobre cobertura textual y tipos de textos, son altamente productivas aplicadas a investigaciones más concretas sobre conocimiento léxico y comprensión lectora (Llinares *et al.* 2008). También lo son si necesitamos medir el umbral léxico de comprensión lectora y comprensión oral necesario para que estas se produzcan de forma eficaz, o la relación entre el conocimiento léxico y las inferencias de

vocabulario⁷¹. La relación entre todas estas cuestiones se puede representar gráficamente de la siguiente forma:

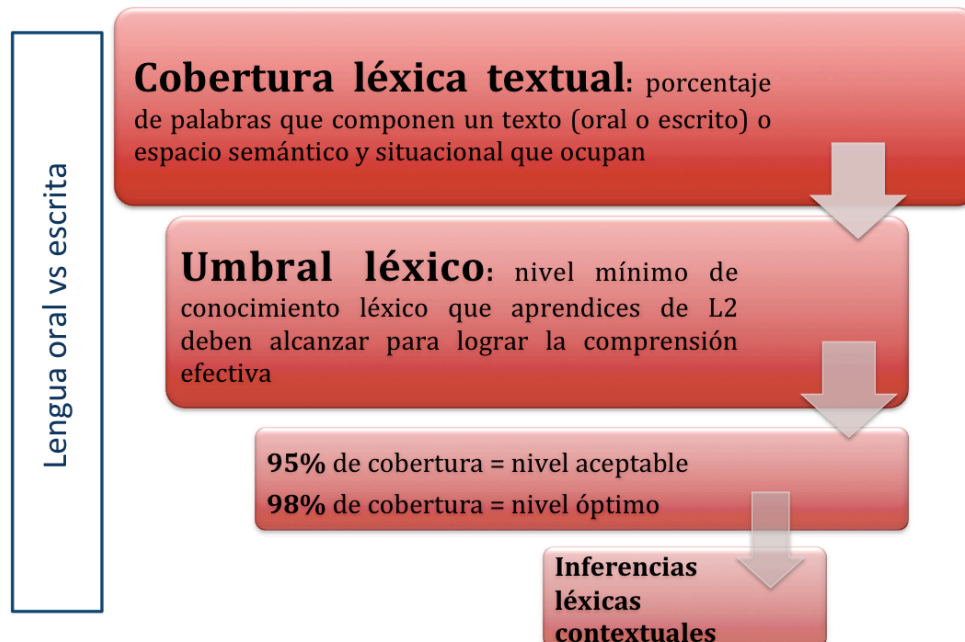


IMAGEN 8. Relación entre la cobertura léxica textual y los umbrales de comprensión léxica.

Las conclusiones de la mayor parte de los trabajos consultados se pueden resumir afirmando que un aprendiz que conoce las primeras 3.000 palabras más frecuentes del inglés y a la vez está familiarizado con un amplio inventario de nombres propios, puede alcanzar el 95% de cobertura léxica de un texto. Esto es lo que se considera el “umbral léxico mínimo” para alcanzar un nivel “aceptable” de comprensión lectora en una L2. Sin embargo, para poder alcanzar un nivel “óptimo” de comprensión, que se estima alrededor del 98% de cobertura textual, las investigaciones sugieren que el aprendiz debe dominar las primeras cinco listas o bandas de frecuencia del *British National Corpus*, es decir, unas 5.000 palabras de alta frecuencia. A esto hay que sumar que las inferencias léxicas no se producen si estos niveles de comprensión no se alcanzan.

A partir de estos estudios, y poniendo en relación estos niveles de comprensión con las destrezas que el *MCERL* establece para los diferentes

⁷¹ Laufer & Sim 1985; Hu & Nation 2000; Nation 2006; Laufer 1989, 1992, 1997a, 1997b; Nation 1990, 2001, 2006; Laufer & Ravenhorst-Kalovsky 2010; Coxhead & Ravenhorst-Kalovsky 2010; Cartaya Febres 2011; etc.

niveles, intentamos calcular cuántas palabras frecuentes son necesarias para cada estadio, si bien nuestro objetivo último es el nivel B1. El resumen de nuestros cálculos se muestra en una tabla de elaboración propia en la que señalamos el número de palabras necesarias, en relación con la cobertura léxica que alcanzan, el nivel de comprensión que se consigue y los niveles del *MCERL* (*vid. supra* § 4.5.2.). Según estos resultados, que ofrecemos por extenso en el Capítulo 3, para el nivel B1 son necesarias entre 3.000 y 5.000 palabras, para conseguir una cobertura de entre 89% y 95%, con la que se puede obtener un nivel de comprensión aceptable (umbral mínimo de comprensión). No obstante, se debe tener en cuenta que estos cálculos pueden variar según nos centremos en la lengua oral o escrita, ya que para esta última la relación es de 5.000 palabras para un 95% y de 8.000 – 9.000 palabras para un 98% cobertura léxica, mientras que en la oralidad las cifras son de 3.000 palabras para un 95% y de 5.000 para un 98% (*vid. supra* § 4.4., 4.4.2. y 4.5.2.).

Gracias a estos cálculos logramos definir qué número de palabras frecuentes necesitamos seleccionar para un nivel B1 a partir de la obra de Davies (2006): entre 3.000 y 5.000. Si bien estas afirmaciones se deben hacer con prudencia. Téngase en cuenta que los estudios anteriores establecen las cantidades a partir del inglés y no contemplan el concepto de léxico disponible.

Por nuestra parte, consideramos una nómina de las 3.000 primeras unidades más frecuentes de la obra de Davies y de las palabras de sus vocabularios temáticos que están entre las 5.000 más frecuentes.

8.3.3.2. Criterios de selección del vocabulario disponible

Aunque contamos con escasa bibliografía en la que se haya acometido la selección del vocabulario disponible de forma total y práctica, sí que existen algunas aportaciones que van en esta línea, como ya hemos señalado (*vid. supra* § 3.4.2.). Hay casos en que la selección se debía al intento de comparar listados de disponibilidad léxica de diferentes autores y lugares. La mayor parte de ellos ha tomado como referencia el índice de disponibilidad y el porcentaje de aparición. Recordemos que el índice de disponibilidad es la cifra a partir de la cual se ordenan las palabras en cada centro de interés y la aparición añade información

sobre el número de personas que han nombrado cada vocablo. En general, todas estas propuestas han establecido como corte para la selección de términos un índice de disponibilidad de 0,1 o similar y un porcentaje de aparición del 30% o 25%. Por ello, en un primer momento probamos estos criterios para seleccionar el léxico disponible. Sin embargo, las dudas que surgen con estos parámetros son de diversa índole. Por una parte, al establecer un parámetro rígido para todos los centros de interés y los diferentes proyectos de disponibilidad léxica, hay palabras que, aun rondando el límite de 0,1, quedan recogidas en unos listados y fuera de otros. Para intentar corregir estas limitaciones consideramos varios estudios de disponibilidad léxica de lugares dispares de la geografía española (Galicia, Alicante, Salamanca y Ceuta). Seleccionamos los vocablos con índice de disponibilidad léxica igual o superior a 0,1 en todos ellos, comparamos los resultados de unos y otros e intentamos abarcar un mayor número de palabras gracias a este procedimiento. No obstante, seguíamos estableciendo un parámetro de forma totalmente arbitraria, sin tener en cuenta que hay muchos centros de interés en los que se obtienen resultados diferentes en función de los informantes y del lugar donde se haya llevado a cabo. No existen propuestas que sustenten estos criterios a partir de un estudio previo de su comportamiento y de los resultados que se obtienen, ni tampoco que definan argumentos claros sobre cuáles son los parámetros adecuados. Por otra parte, tampoco se han puesto estos límites de corte en relación con los niveles de lengua del *MCERL*, ni se han comparado con las unidades que el *PCIC* ofrece para cada estadio.

Al considerar para nuestro trabajo el diccionario de léxico disponible obtenido por Tomé Cornejo (2015), pudimos tener en cuenta otra propuesta de selección léxica del vocabulario disponible, en este caso puesto en relación con un B1 del *MCERL* y cotejado con las nociones específicas que el *PCIC* contemplaba hasta ese nivel. Con todo, el método de selección que se sigue en este proyecto es más una forma improvisada de poder comparar sus resultados con el *PCIC* que una apuesta firme de selección del léxico disponible, como ya hemos comentado y como la propia autora reconoce. Así, aunque puede parecer una ventaja el hecho de que no fije un límite estable para todos los centros de interés, en realidad tampoco se centra en estudiar la cuestión de la selección. Los pasos que

sigue para cotejar sus listas de palabras con el *PCIC* son más una cuestión práctica, pues no esgrime argumentos a favor de este proceso ni toma en consideración el índice de disponibilidad, ni la aparición, ni la frecuencia. En su caso limitó la elección de las palabras disponibles al número exacto de unidades que se incorporan en los niveles A1, A2 y B1 del *PCIC*. Luego seleccionó aquellas que se corresponden con el tema y el subapartado en concreto, sin mostrar las cifras de su índice de disponibilidad, su frecuencia o aparición. No obstante, como hemos comprobado al analizar sus datos, fue escogiendo las unidades de mayor a menor índice de disponibilidad, es decir, si tenía que seleccionar 29 elementos de *El cuerpo humano* en la lista de palabras disponibles, debido a que este era el número de exponentes léxicos en el *PCIC* desde el A1 hasta el B1, entonces escoge los elementos de mayor a menor disponibilidad. De todos los elementos del centro de interés, se queda solo con aquellos que, según su parecer, cumplen el requisito de estar dentro de esa categoría en concreto, que en este ejemplo es *El cuerpo humano*. En este caso sería fácil llevar a cabo esta acción, porque distinguir entre partes del cuerpo y acciones resulta simple, ya que estas últimas pertenecen a la clase de los verbos o participios, mientras que las primeras son sustantivos. Además, la mayor parte de estos nombres están en las primeras 30 posiciones por su índice de disponibilidad, por lo que la coincidencia es alta. En cambio, en la mayor parte de los centros de interés y epígrafes del *PCIC*, distinguir en qué apartados se deben encajar las unidades puede resultar más complicado: no se observan tales equivalencias y la selección de los vocablos del léxico disponible se debe acometer en posiciones dispersas y distantes. Por este motivo, estos criterios resultan aleatorios y subjetivos.

Por nuestra parte, creemos que se debe tener en cuenta el índice de disponibilidad y/o el porcentaje de aparición, ya que a estas cifras se ha llegado de una forma empírica, por lo que no debemos obviarlos limitándonos a emplear criterios puramente subjetivos. Con todo, también creemos importante optar por una vía que tenga en cuenta el comportamiento y las características de cada centro de interés, así como las unidades léxicas concretas que los forman. Por consiguiente, apostamos por una fórmula que cumpla los requisitos de no ser aleatoria o arbitraria, pero tampoco rígida.

La solución la encontramos al analizar los centros de interés como un conjunto de elementos, observando su comportamiento a través de los datos del índice de disponibilidad de sus vocablos de una forma global. Al fijarnos en la evolución del número de palabras que obtenían un alto índice de disponibilidad frente a las que se acumulaban en las posiciones más bajas de las listas, percibimos que las cantidades eran siempre opuestas, formando parte de dos polos distantes. Así, frente a las primeras posiciones, en las que solo conseguimos aislar un número muy reducido de elementos, los que obtienen un índice de disponibilidad bajo son la mayor parte del conjunto total. Entre ellos siempre existe una gradación de elementos hasta que llegamos a un punto del conjunto en que la cifra de unidades se dispara, pasando a ser muchas más. Esto es lo que ocurre cuando observamos el número de palabras que tienen una disponibilidad alta o baja, ya que son pocas en un inicio y muchas al final, como veremos en los histogramas. Pero, si observamos la evolución del índice de disponibilidad, entonces la tendencia es a caer en picado, debido a que su disponibilidad comienza a disminuir muy rápidamente en determinado punto⁷². Si lo representamos gráficamente, la tendencia de la cola que obtenemos en las imágenes es igual para todos los centros de interés, aunque adaptada a los datos con los que contamos en cada caso. Esto es lo que se puede observar en el Gráfico 1, en el que representamos la evolución del índice de disponibilidad de los vocablos del centro de interés *El cuerpo humano* extraídos de Tomé Cornejo (2015):

⁷² Nótese que en los Gráficos 1 y 2 representamos el índice de disponibilidad de las palabras, mientras que en los gráficos siguientes (los que denominamos histogramas de frecuencia) contabilizamos el número de palabras que obtenemos con alta disponibilidad, media y baja disponibilidad. En el primer caso, las palabras que están a la izquierda son las más disponibles, mientras que en los histogramas, las palabras que están a la izquierda del gráfico son las de menor índice de disponibilidad, pues se acercan al grado 0 de disponibilidad.

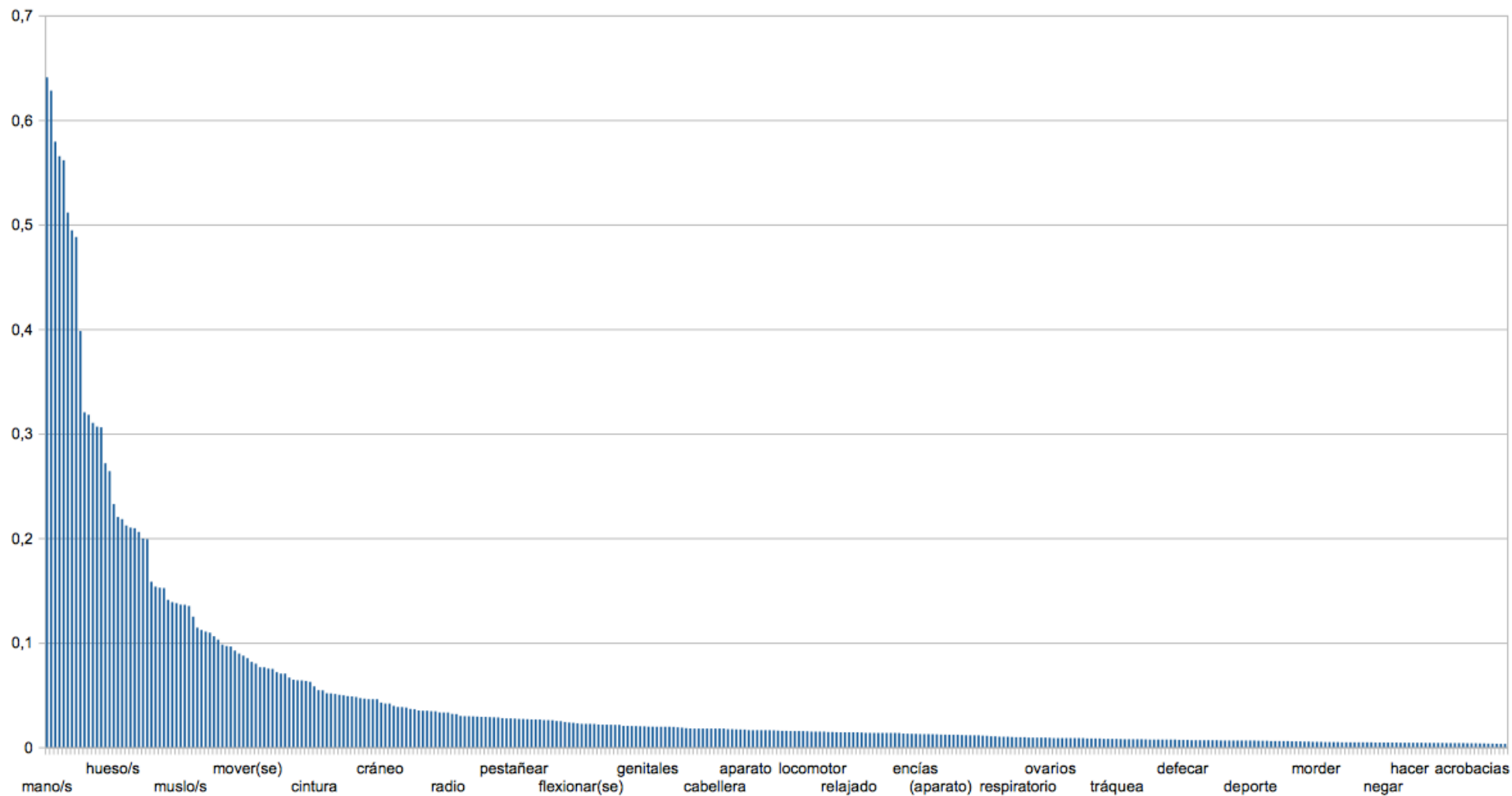


GRÁFICO 1 . Evolución del ID de los vocablos del CI *El cuerpo humano* extraídos de Tomé Cornejo (2015)

En el eje de ordenadas (también llamado *y*, el vertical) se reflejan los índices de disponibilidad de cada vocablo, mientras que en el de abscisas (o *x*, el horizontal) se muestran las palabras que obtienen esos valores. Como se puede observar, a medida que el índice de disponibilidad se aproxima a cero, la cola de la imagen se estabiliza. Si creamos el mismo tipo de gráfico para cada uno de los centros de interés, independientemente del trabajo de léxico disponible que escojamos y del número de palabras que se obtengan, el espectro y la tendencia de la cola es siempre esta, como se puede apreciar en este gráfico sobre los datos de López Meirama del CI *Calefacción e iluminación* (vid. Gráfico 2). En él se ve cómo, aunque se trate de un centro de interés diferente y con un conjunto total mucho mayor debido a que la muestra de informantes era más numerosa, el perfil del gráfico es exactamente el mismo.

Al comparar ambos gráficos también obtenemos otra información: aunque la muestra sea mucho mayor y contemos con un número total de palabras mayor, lo que aumenta es la cantidad de palabras con baja disponibilidad, pues la tendencia en las primeras posiciones de la lista es la misma.

Estos hechos nos llevaron a considerar los centros de interés desde una nueva perspectiva: representándolos en histogramas de frecuencia, tomando como base su índice de disponibilidad⁷³. Como es sabido, el histograma es un gráfico en el que se agrupan todos los elementos en diferentes conjuntos y se contabilizan cuántos se obtienen en cada uno de ellos. El objetivo es observar, en términos numéricos, la tendencia de estos conjuntos para intentar detectar *clusters*, es decir, asociaciones entre ellos. O bien, por el contrario, descubrir qué

⁷³ No debemos confundir esta “frecuencia” con la “frecuencia” que indican los diccionarios de léxico disponible, porcentaje de frecuencia relativa o absoluta. En el caso de los histogramas, estos son la representación gráfica de la distribución de la frecuencia de una variable cuantitativa continua (en nuestro caso, el índice de disponibilidad). Lo que hacemos es agrupar los datos en “clases” o “conjuntos” y contamos cuántos elementos observamos (*frecuencia absoluta*) en cada una de ellas. En algunas variables (*variables cualitativas*) las clases están definidas de modo natural. Por ejemplo si hablamos de sexo, se suelen considerar solo dos clases: mujer y varón. Si las clases son de un grupo sanguíneo, partimos de cuatro: A, B, AB, O. En las variables cuantitativas, como en nuestro caso, las clases hay que definir las explícitamente (*intervalos de clase*). Nosotros hemos trabajado con histogramas de 10 clases y de 20, y hemos calculado los intervalos que se correspondían con estas divisiones.

cortes o separaciones se producen en su línea de tendencia. La ventaja frente a otros métodos es que no se necesita un *software* especial o fórmulas matemáticas creadas *ad hoc* para este fin, ya que la realización de histogramas es algo simple, popular y que se puede hacer con cualquier paquete de *Office* que permita trabajar con hojas de cálculo, como puede ser el conocido *Excel*.

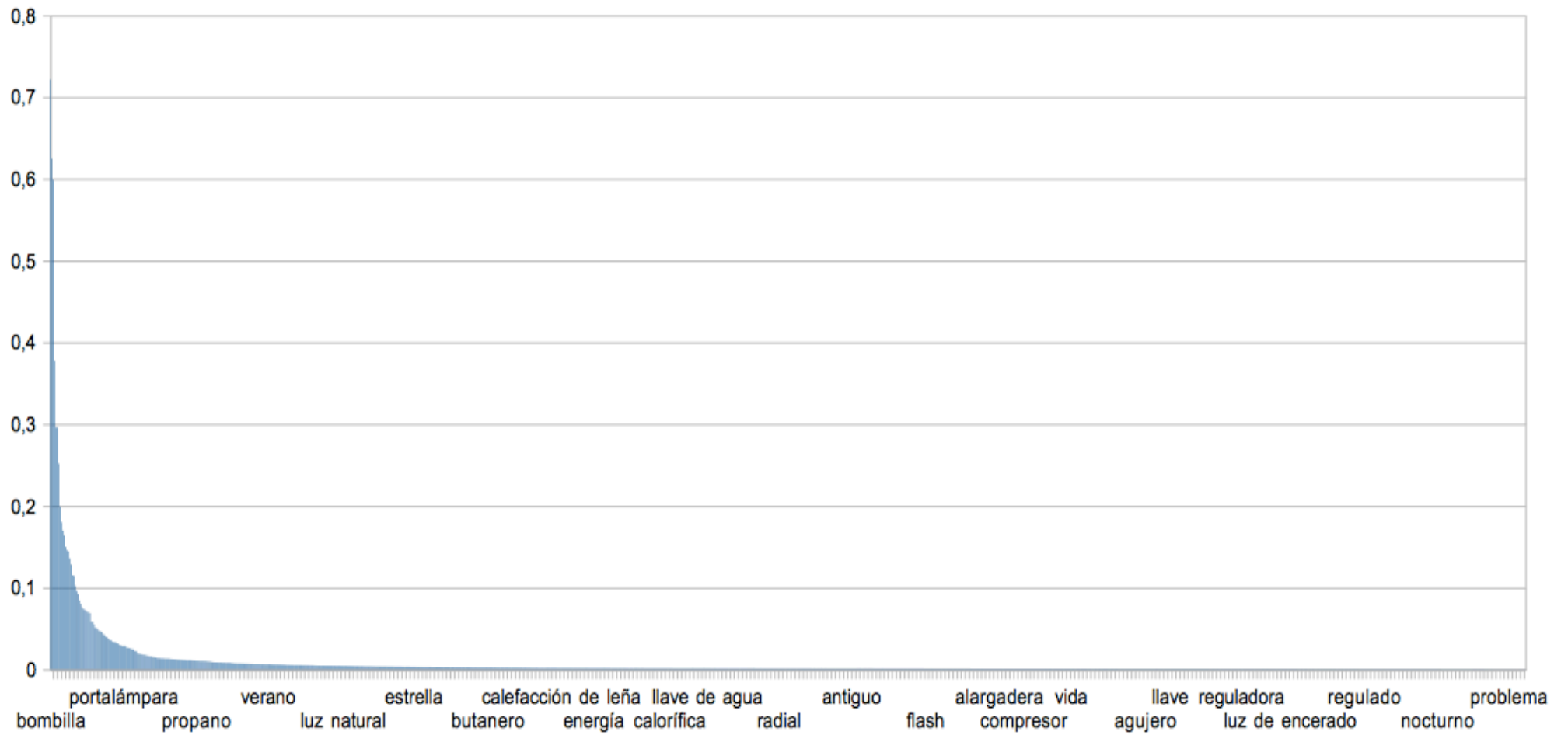


GRÁFICO 2. Evolución del ID de los vocablos del CI *Calefacción e Iluminación*, López Meirama (2008).

Para nuestro estudio, partimos de la hoja de cálculo de *Openoffice*. Con él realizamos histogramas de frecuencias a partir del índice de disponibilidad de los 16 centros de interés del proyecto de léxico disponible del español de Galicia de López Meirama (2008)⁷⁴ y de los 47 centros de interés del estudio de Tomé Cornejo (2015)⁷⁵. El número total de histogramas y sus resultados se pueden observar en los Anexos 5-6 y 7-8. Para llevar a cabo este trabajo ha sido necesario decidir en cuántos conjuntos es más operativo secuenciar la lista de palabras. Después de hacer diversas pruebas, optamos por histogramas de 10 y de 20 conjuntos, ya que los primeros nos ofrecen una visión más global, pero los segundos permiten establecer cortes y gradaciones con un mayor nivel de precisión y detalle. Una forma de probar si este método estadístico resulta eficaz para nuestro objetivo es realizar la misma operación en centros de interés similares de diferentes estudios, a la espera de obtener resultados paralelos. Para ello, podemos examinar como ejemplos los CI *El cuerpo humano* y *Profesiones y oficios*, ya que se mantienen en los dos estudios que hemos considerado. En general, podemos afirmar que en los casos en que en ambos proyectos se pregunta exactamente por lo mismo, el gráfico presenta idéntico aspecto. Ocurre esto con los CI *El cuerpo humano* y *Profesiones y oficios*, sobre todo en este segundo.

⁷⁴ Recordemos que los 16 centros de interés usados en el PPHDL, a los que llamamos tradicionales por no haber variado con respecto a los originales de los pioneros franceses, son:

1. El cuerpo humano, 2. La ropa, 3. Partes de la casa (sin incluir muebles), 4. Los muebles de la casa, 5. Alimentos, 6. Objetos colocados en la mesa para la comida, 7. La cocina y sus utensilios, 8. La escuela: muebles y materiales, 9. Calefacción, iluminación y medios de airear un recinto, 10. La ciudad, 11. El campo, 12. Medios de transporte, 13. Trabajos del campo y del jardín, 14. Los animales, 15. Juegos y distracciones, 16. Profesiones y oficios.

⁷⁵ Los 47 centros de interés de la nueva propuesta de Tomé Cornejo son los siguientes:

1. El cuerpo humano: partes, acciones y posturas, 2. El aspecto físico y la personalidad, 3. El ciclo de la vida: nacimiento, reproducción y muerte, 4. Sentimientos y sensaciones, 5. Para identificarse (documentación, datos y objetos personales), 6. Las relaciones personales (familiares y no familiares), 7. Actos sociales y celebraciones, 8. Alimentación y nutrición, 9. Preparar una comida, 10. En el restaurante, 11. En el bar, 12. El sistema educativo: etapas, personal y centros de enseñanza, 13. En clase, 14. Profesiones y oficios, 15. El mundo laboral, 16. Actividades de ocio y tiempo libre, 17. Actividades artísticas, 18. La información y la publicidad, 19. Comunicarse por teléfono o carta, 20. Informática y nuevas tecnologías, 21. Buscar una vivienda, 22. La casa y las tareas domésticas, 24. En el banco, 25. El sistema sanitario: personal y centros, 26. En comisaría, 27. Servicios sociales y solidaridad, 28. La ciudad y el campo, 29. En la tienda de ropa, 30. Hacer la compra, 31. En el médico, 32. Acciones que realizamos todos los días, 33. En el baño o aseo, 34. Conducir un coche, 35. Economía y hacienda, 36. Sectores de producción (primario y secundario), 37. La ciencia, 38. Organización política y social, 39. En un juicio, 40. El ejército, 41. En el cine o en el teatro, 42. Creencias religiosas y espiritualidad, 43. El universo y los viajes espaciales, 44. La geografía y los accidentes geográficos, 45. El tiempo atmosférico, 46. Animales y plantas, 47. Medioambiente y desastres naturales.

Sus gráficos son casi iguales (*vid.* Gráficos 3 y 4, histogramas en este caso): con 3 *clusters* bien diferenciados, que se pueden ampliar a 4 si tenemos en cuenta la variación entre la columna 8 y la 9 (empezando siempre a contar por la derecha, pues, recordémoslo, es donde están las palabras más disponibles). En el primer centro de interés, *El cuerpo humano*, el primer grupo lo forman las primeras columnas de la derecha hasta el primer conjunto en el que obtenemos 0 unidades. Estas 4 primeras columnas representan las escasas palabras que tienen un índice de disponibilidad alto, hasta que se observa una primera quiebra en la lista de palabras, ya que no obtenemos ningún ejemplo de vocablo disponible entre el índice de disponibilidad 0,323125 y el índice de 0,38734 en Tomé Cornejo (*vid.* Gráfico 3) y entre los índices 0,33537 y 0,402416 en López Meirama (*vid.* Gráfico 4). EL número de palabras que estarían entre este grupo de vocablos de mayor disponibilidad varía de 11, en el primer caso, a 7 en el segundo. En el Gráfico 3 se ve que en las primeras columnas el aumento de las palabras es gradual, con 3 elementos cada una, hasta la bajada en la cuarta columna. En cambio, en las primeras posiciones del Gráfico 4 solo vislumbramos 1 elemento, hasta que se produce una pequeña concentración de 6 vocablos en la tercera columna. En la quinta columna hay una quiebra, ya que no se advierten elementos en ese conjunto, y a partir de ahí un aumento leve y progresivo en las siguientes 3 columnas. Después hay un aumento un poco más elevado, justo antes de que se produzca la gran concentración de palabras en el último grupo. Esta última columna representa la gran mayoría de unidades en términos globales y son las que más se aproximan al punto 0, que es el nivel de disponibilidad mínimo.

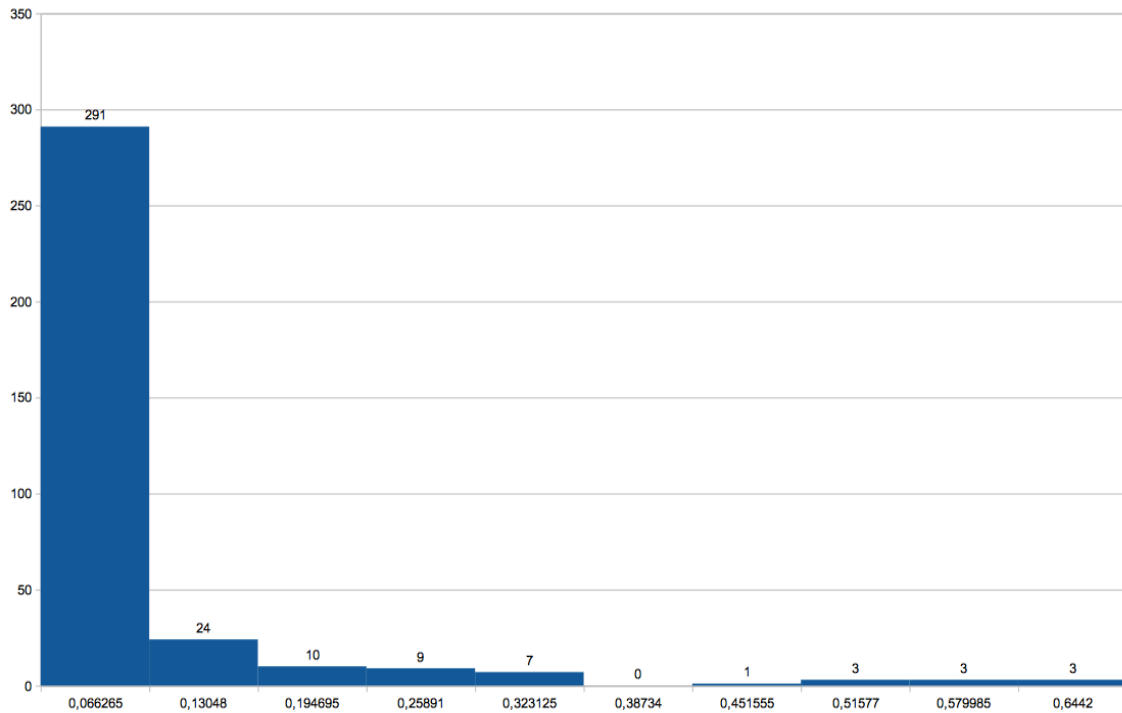


GRÁFICO 3. Histograma de 10 conjuntos a partir de la frecuencia de palabras según su ID en el CI *El cuerpo humano* de Tomé Cornejo (2015).

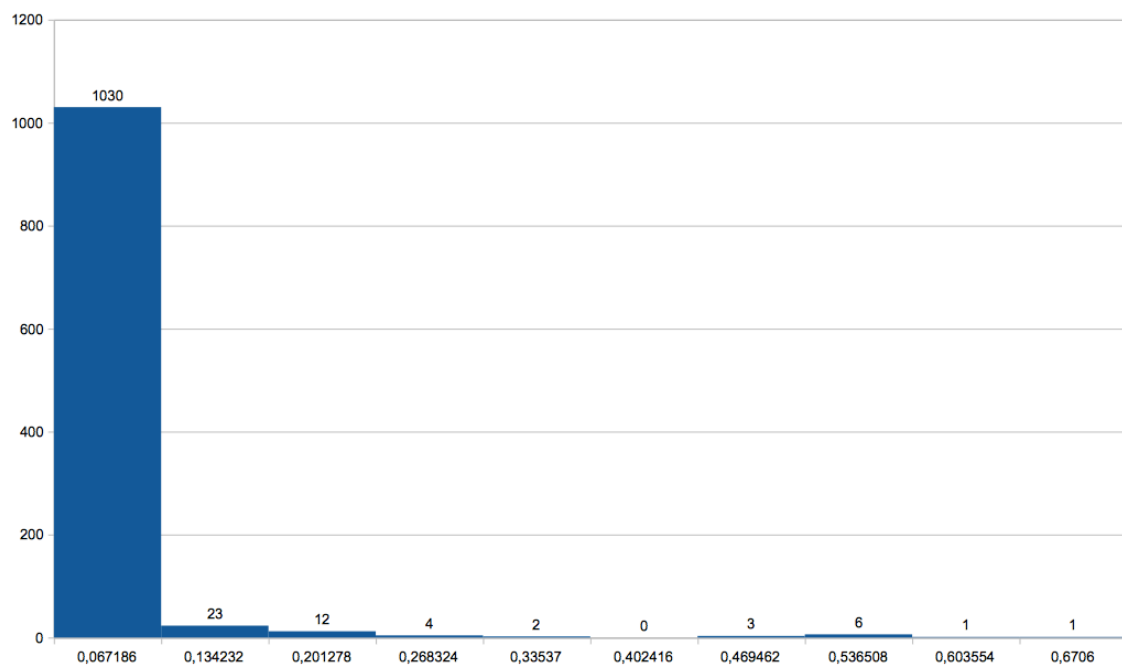


GRÁFICO 4. Histograma de 10 conjuntos a partir de la frecuencia de palabras según su ID en el CI *El cuerpo humano* de López Meirama (2008).

Si en lugar de fijarnos en las cantidades de palabras de cada grupo nos centramos en su línea de tendencia, su evolución es también parecida entre ambos trabajos.

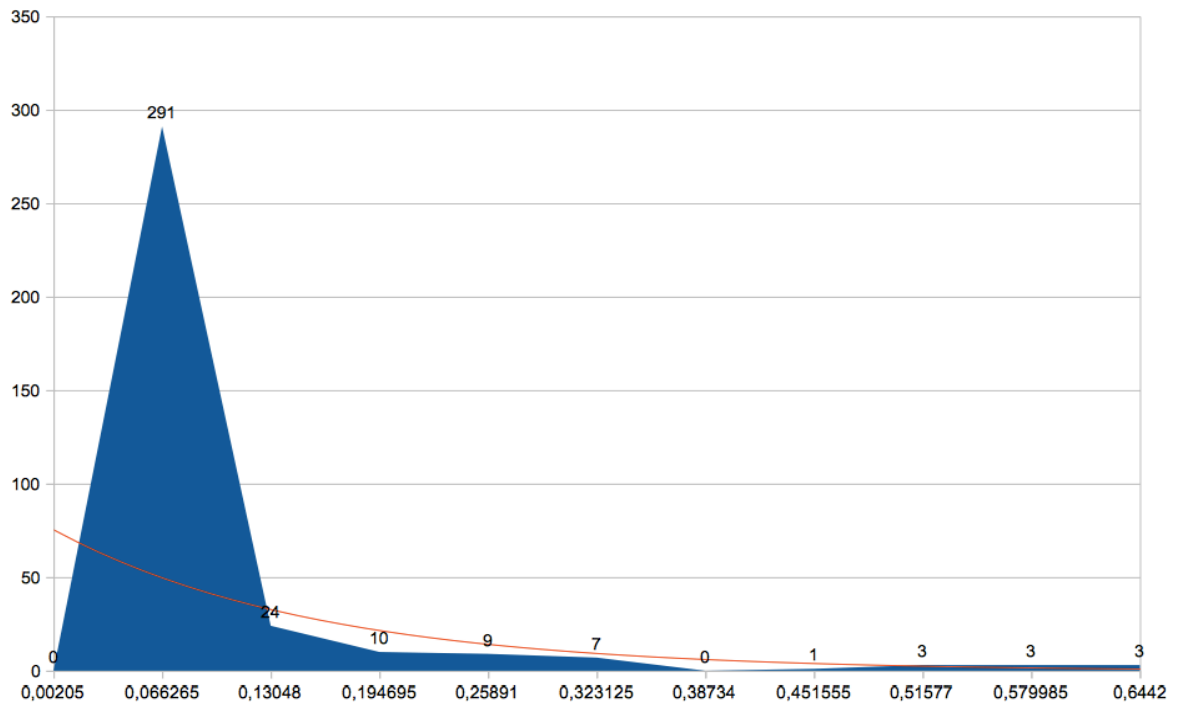


GRÁFICO 5. Histograma de 10 conjuntos con su línea de tendencia a partir de la frecuencia de palabras según su ID en el CI *El cuerpo humano* de Tomé Cornejo (2015).

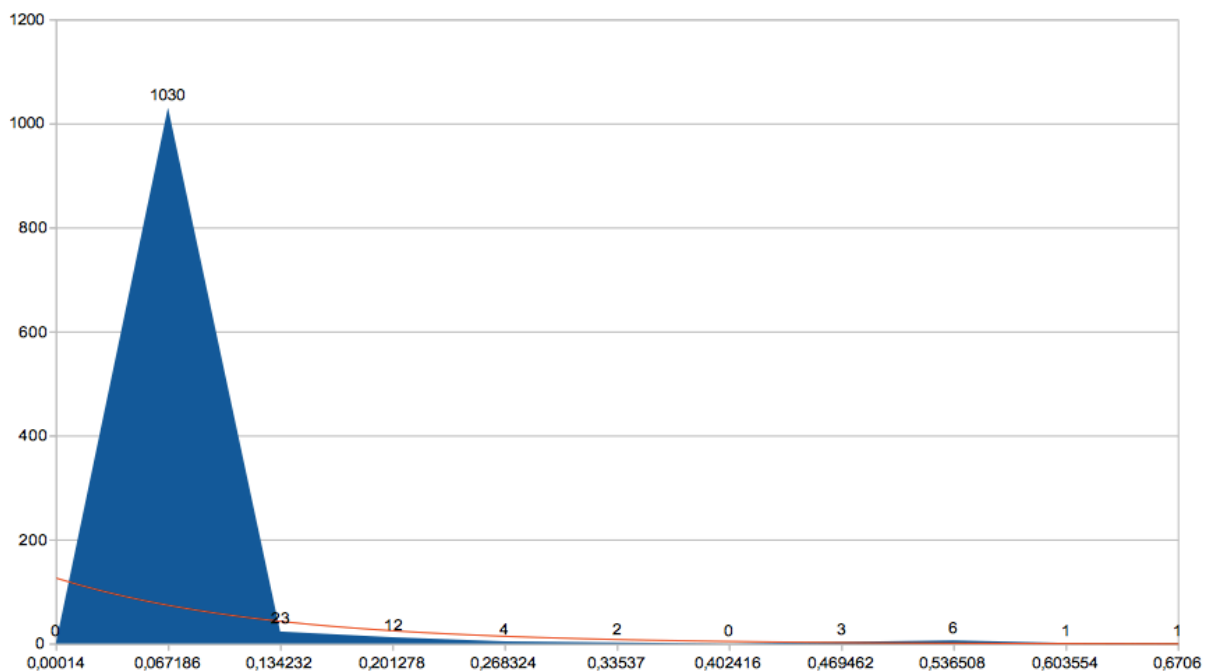


GRÁFICO 6. Histograma de 10 conjuntos con su línea de tendencia a partir de la frecuencia de palabras según su ID en el CI *El cuerpo humano* de López Meirama (2008).

Para los Gráficos 7 y 8, que representan los datos del CI *Profesiones y oficios*, la similitud es todavía mayor, como se puede observar:

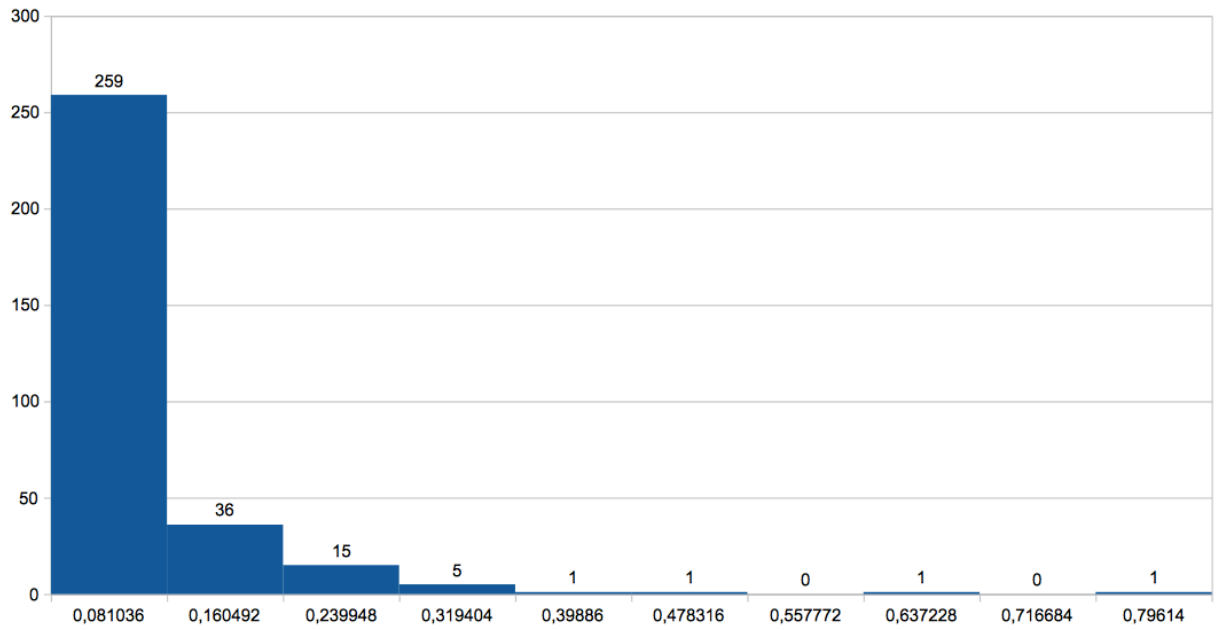


GRÁFICO 7. Histograma de 10 conjuntos a partir de la frecuencia de palabras según su ID en el CI *Profesioness y oficios* de Tomé Cornejo (2015)

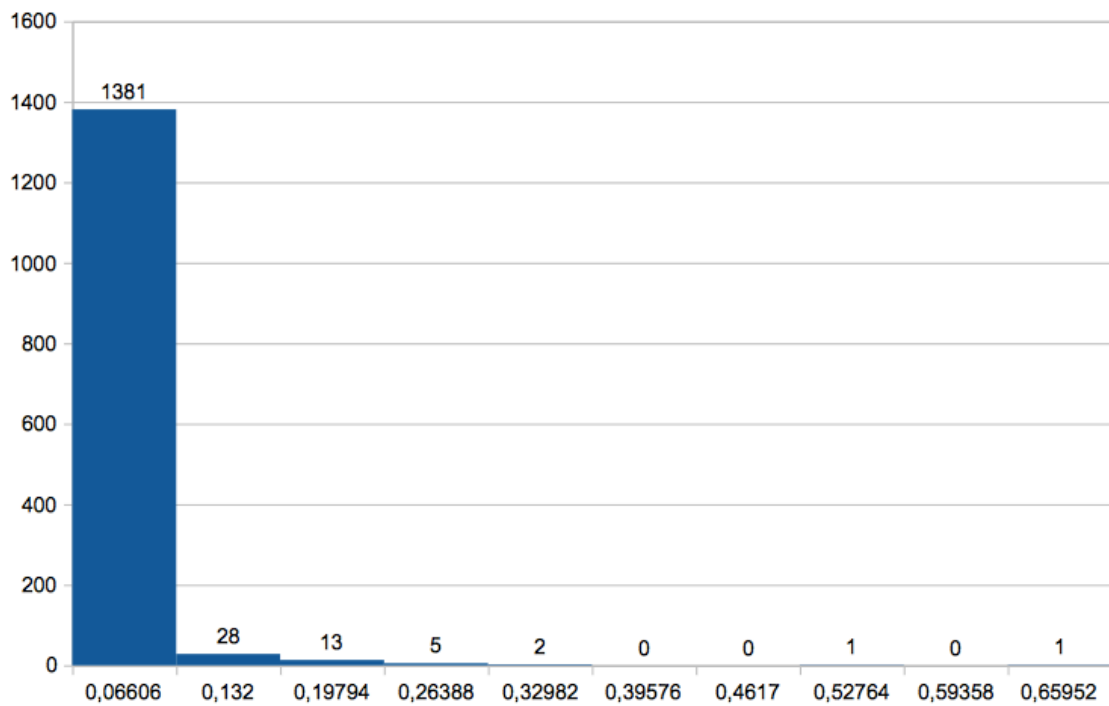


GRÁFICO 8. Histograma de 10 conjuntos a partir de la frecuencia de palabras según su ID en el CI *Profesioness y oficios* de López Meirama (2008)

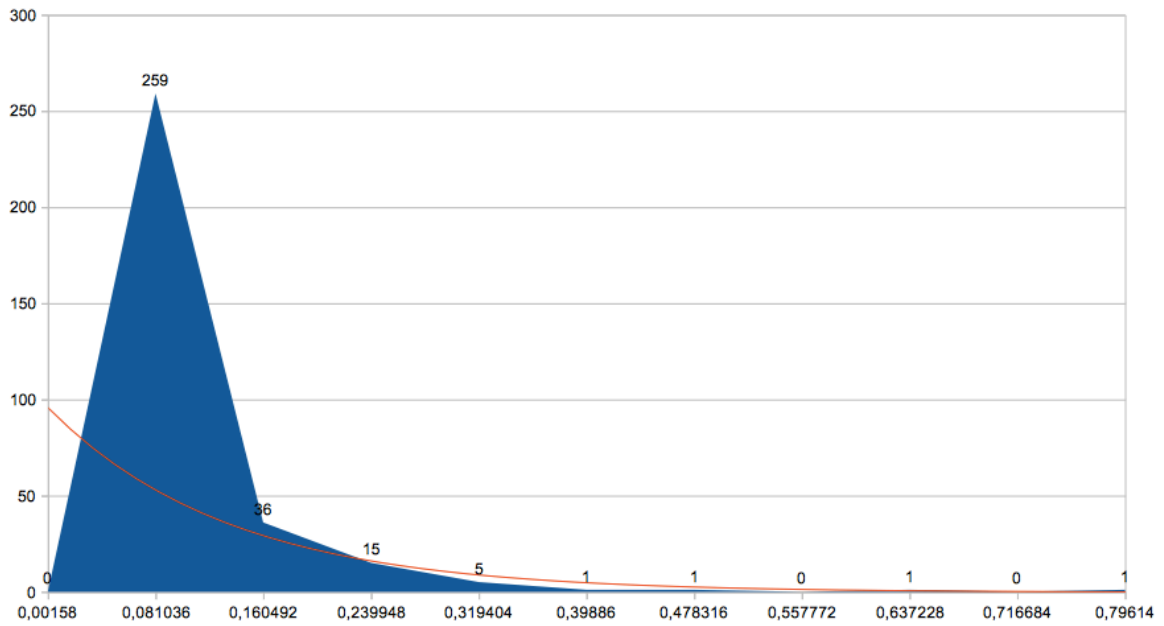


GRÁFICO 9. Histograma de 10 conjuntos con su línea de tendencia a partir de la frecuencia de palabras según su ID en el CI *Profesion y oficios* de Tomé Cornejo (2015).

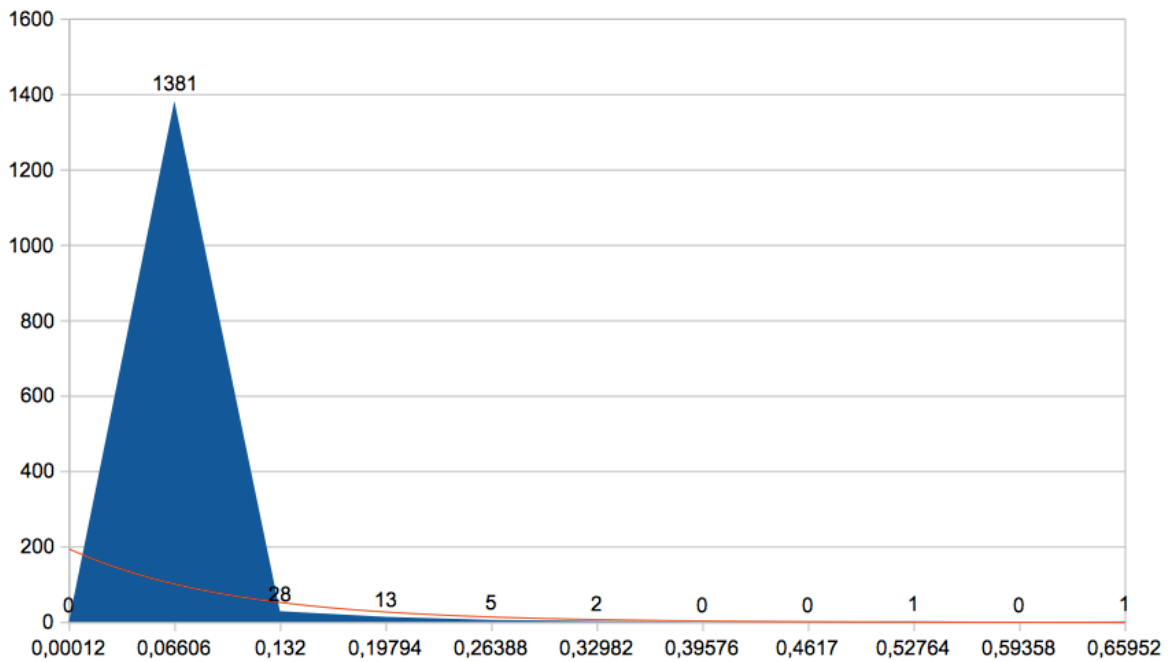


GRÁFICO 10. Histograma de 10 conjuntos con su línea de tendencia a partir de la frecuencia de palabras según su ID en el CI *Profesion y oficios*, López Meirama (2008).

En ambos estudios el índice de disponibilidad de las primeras posiciones es casi el mismo, 0,6442 y 0,6706, respectivamente. En el CI *Profesiones y oficios* hay una mayor diferencia, 0,79614 y 0,65952, aunque probablemente se acortaría si el número de informantes con el que se trabajase fuese el mismo. Lo que sí consideramos relevante mencionar es el hecho de que en todos los estudios de léxico disponible las palabras con mayor índice de disponibilidad obtienen cifras muy similares a estas. A veces incluso menores, pues hay casos en los que se acercan más al 0,5 de índice de disponibilidad, que se puede considerar como el punto intermedio entre 1 y 0. Lo que queremos resaltar con esto es que con las encuestas de léxico disponible no siempre obtenemos palabras con un alto grado de disponibilidad, lo cual se observa de forma más drástica en unos centros de interés que en otros. Estas ocasiones en las que los resultados son menos satisfactorios invitan a usar de la disponibilidad léxica con mayor prudencia. Es necesario tenerlo en cuenta cuando proponemos esta línea de investigación como un medio de selección del vocabulario, pues debemos ser conscientes de los puntos débiles y aspectos a mejorar de su metodología, del mismo modo que valoramos sus virtudes. Así pues, no cabe duda de que es un método altamente eficaz para conseguir un vocabulario temático, pero debemos pensar que usada como único método de selección no resulta tan útil y satisfactoria. En este sentido, creemos que se puede recurrir a la disponibilidad como una fórmula de selección léxica para un tipo de unidades concretas, que unidas a las palabras más frecuentes, por ejemplo, consiguen completar y enriquecer aquellos vacíos que estas dejan.

En conclusión, y aunque con pequeñas variaciones, estas tres o cuatro asociaciones entre las columnas de los histogramas se observan claramente en todos los centros de interés de ambos estudios. Los gráficos son casi iguales, con 3 *clusters* diferenciados, que se pueden ampliar a 4 si tenemos en cuenta el leve salto entre la columna 8 y la 9. Así pues, la tendencia que se aprecia a primera vista es clara:

1. Palabras más disponibles: las primeras columnas con menos de 10 unidades. Son las que más se acercan al 1 (el nivel de máxima disponibilidad).

2. Palabras de disponibilidad media-alta: un segundo grupo de columnas intermedias, en las que la hay un pequeño aumento gradual de unidades (que suelen estar entre 5 y 20).
3. Palabras de disponibilidad media: la penúltima columna, que suele sobrepasar los 15 elementos pero nunca los 40. Normalmente el índice de disponibilidad de estas palabras es menor a 0,1.
4. Palabras de baja disponibilidad: la última columna, en la que se acumulan todos los vocablos restantes de ese centro de interés. Son los que tienen un índice de disponibilidad que más se aproximan a 0 (la mínima disponibilidad posible).

Esta secuenciación en conjuntos se puede poner en relación con los niveles de lengua del *MCERL*. A nuestro modo de ver, las palabras de tipo 1 están en relación con los niveles A1-A2, las de tipo 2 son buenas candidatas para el nivel B1, mientras que las de tipo 3 corresponderían al nivel B2. Las palabras de tipo 4 son las que ostentan una disponibilidad muy baja, por lo que las reservamos para el nivel C del *Marco*. Una vez llegamos a estas conclusiones, estudiamos todos y cada uno de los histogramas de cada centro de interés de los dos proyectos de disponibilidad para establecer los límites de estos cuatro niveles de lengua. Tenemos como objetivo principal localizar el corte para el nivel B1, pues son esas las palabras que nos interesan en este trabajo.

El procedimiento fue siempre el mismo, analizar los histogramas de todos los centros de interés de ambas autoras y delimitar los 4 niveles referidos, en función de la estructura de sus conjuntos, tomando como referencia las agrupaciones y los cambios de frecuencia. Nótese que hemos definido los límites para cada caso, es decir, aunque las tendencias y los cortes sean muy similares, no los extrapolamos para el resto. En cada centro de interés de los diferentes diccionarios de léxico disponible se deben observar cuáles serían los límites adecuados, aunque se espere que estos sean muy parecidos si el centro de interés tiene la misma estructura. Esto quiere decir que se puede aplicar el mismo procedimiento, pero los márgenes no tienen por qué ser idénticos, se adaptarán en cada ocasión en función del análisis de los datos.

Sin embargo, seguimos fijando un corte rígido entre los diferentes conjuntos de palabras, por lo que hay algunos vocablos de las listas que rozan los límites y que pasan a formar parte de un grupo u otro por una diferencia pequeña. No creemos que esta cuestión sea un gran problema a la hora de escoger las palabras, pues se puede tomar estos límites como aproximativos, de tal modo que escogeremos los términos fronterizos si los consideramos interesantes por su índice de disponibilidad, su frecuencia y su porcentaje de aparición, o bien porque creemos que pueden ser de utilidad para el contexto educativo en el que vayamos a trabajar. En este sentido, los cortes no tienen por qué ser drásticos en los casos en los que se observa que no hay una separación clara entre conjuntos en los gráficos, sino una tendencia gradual. Aun así, podemos buscar una solución para esta cuestión basándonos también en cálculos matemáticos, que nos ofrezcan una forma más científica y menos intuitiva de hacer separaciones entre elementos que parecen estar más diluidos que distantes.

Buscando respuestas para estas cuestiones hemos encontrado algunas propuestas recientes que nos parecen sumamente interesante. Nos referimos a las propuestas lanzadas por el profesor Ávila Muñoz en conjunto con otros colegas de investigación del Proyecto CONSOLEX sobre la disponibilidad léxica en la ciudad de Málaga (Villena Ponsoda, Ávila Muñoz & Sánchez Sáez 2010). A raíz de este estudio, sus autores trabajaron para lograr una mejor organización de los términos y aportar soluciones que iban en la línea de una selección más precisa y objetiva de las unidades léxicas ajustándose a los niveles de aprendizaje de las lenguas extranjeras. Para ello exploraron las posibilidades que la psicolingüística y la lingüística cognitiva ofrecen a la hora de analizar el comportamiento del léxico disponible y el lexicón mental, al igual que había hecho Tomé Cornejo para ofrecer avances sobre los centros de interés. Estos paradigmas teóricos proporcionan nuevas herramientas para estudiar la naturaleza del léxico disponible a la luz de algunas de sus teorías. Así, Ávila Muñoz propone la teoría de los prototipos y los conceptos de “accesibilidad” y “compatibilidad” para interpretar el índice de disponibilidad de los vocablos (Ávila Muñoz 2014 y 2016).

Coincidimos con este investigador, en su afirmación de que, partiendo de la teoría de los prototipos, el índice de disponibilidad es un excelente indicador del grado de prototipicidad que los vocablos alcanzan dentro de cada centro de interés. Para demostrar esto, estudia el comportamiento de la “red léxica” de los informantes a partir de sus respuestas en las encuestas de disponibilidad léxica. Sabemos, además, que el léxico de los individuos tiende a ser similar al resto de personas de su comunidad por motivos culturales y ambientales, a pesar de que muestre también diferencias. Según su hipótesis, se podría interpretar que las listas producidas por cada informante corresponden, en su forma, a una representación del acceso a su léxico particular.

Cada centro de interés se puede entender como un prototipo, creado a partir del concepto que lo determina, de tal modo que cuando un individuo se somete a la pruebas de disponibilidad, accede a su red léxica a partir del prototipo del núcleo nocional (Ávila Muñoz 2016). A continuación, va recorriendo esta red léxica mientras se va alejando progresivamente del centro del prototipo. El grado de prototipicidad de los vocablos se pone en relación con un concepto usual en las encuestas de asociación de palabras a estímulos, como es el de “accesibilidad”. Así, mediante las encuestas se estudia el acceso léxico que los informantes muestran con respecto a un campo nocional. La disponibilidad de las unidades léxicas de cada centro de interés se corresponde con el concepto de “accesibilidad”, y los resultados finales que se muestran en los diccionarios de léxico disponible deben ser entendidos como su categorización conceptual colectiva. Los vocablos que alcanzan el índice de disponibilidad más alto forman la categorización conceptual colectiva del estímulo que se ofrece como centro de interés. A medida que el índice de disponibilidad de los vocablos disminuye, los elementos léxicos se van alejando del núcleo prototípico del centro de interés. El índice de disponibilidad está también asociado con su porcentaje de aparición, es decir, con el número de individuos que han nombrado esa palabra en cuestión. La relación entre ambos es la siguiente: las palabras más disponibles son también las nombradas por un mayor número de hablantes. Además, según este autor, a medida que decrece el índice de disponibilidad, los vocablos se alejan del centro del estímulo y los elementos

léxicos pasan a formar parte de lo que denomina la “línea de incertidumbre”. Este es el límite en el que el sujeto empieza a dudar de si las palabras que ofrece corresponden exactamente al estímulo propuesto.

Por otra parte, si el concepto de centro de interés se asocia al de prototipo o a un “campo de experiencia compartido”, la accesibilidad puede entenderse como el nivel de compatibilidad de las palabras con la idea que el centro de interés representa. Así, podemos medir la pertenencia de los elementos a un conjunto (o centro de interés) como la compatibilidad que estos guardan con el concepto representado por este conjunto. Contar con un índice matemático que mida el grado de compatibilidad de los vocablos con el campo nocional con el que se asocian nos permite seleccionarlos de manera objetiva y secuenciarlos en función de su nivel de adecuación a cada nivel de aprendizaje. Su propuesta permite a Ávila Muñoz distinguir varios niveles de compatibilidad del léxico dentro de un centro de interés, cuya secuenciación se basa en el “grado de compatibilidad”, que varía entre el 1 (compatibilidad absoluta) y el 0 (incompatibilidad absoluta). Las palabras que se acerquen más al 1 pertenecen al núcleo del centro de interés y son las más prototípicas o accesibles para los hablantes, mientras que las que se acerquen más a 0 forman parte de la periferia de ese concepto y son poco accesibles.

Ávila Muñoz muestra los resultados de esta investigación en un artículo reciente en que se ofrecen estas explicaciones a las que nos hemos referido y se ilustran con algunos de los resultados obtenidos (Ávila Muñoz 2016). Su objetivo es proponer marcas objetivas de corte por niveles para el conjunto del vocabulario de cada centro de interés, sin que estas dependan de la percepción subjetiva de cada persona que aborda la investigación. Al mismo tiempo, estas marcas relacionan el grado de compatibilidad de cada elemento con la valoración del conjunto. Un vez demarcado el léxico nuclear, lo coteja con los niveles que le son asignados en el *MCERL* y en el *ACTFL. Performance Descriptors for Language Learners* (2012). De hecho, este grado de prototipicidad o compatibilidad se puede equiparar con los 4 niveles de palabras que antes hemos discernido en nuestros histogramas de frecuencia.

La gran dificultad que hemos encontrado es que, en su trabajo, se explican los conceptos y resultados pero no se muestran las operaciones que se han realizado para llegar a ellos, como ya hemos comentado en el Capítulo 6 (§ 6.2.1). Sí se afirma que la propuesta es un procedimiento de selección léxica a partir de la teoría de los “conjuntos difusos”, aplicándola al ámbito de la disponibilidad léxica. Para ello la referencia que se ofrece es la de Zimmermann (2001), quien expone los diferentes modelos matemáticos de la teoría de los conjuntos difusos. Según Ávila Muñoz, de esta obra se extraen las herramientas para determinar el valor de compatibilidad (mediante los modelos FEV, *Fuzzy Expected Value*, o su variación WFEV, *Weighted Fuzzy Expected Value*) y establecer con ellos los límites de caracterización de los valores de pertenencia, así como los parámetros para identificar elementos muy característicos o poco característicos.

No obstante, hemos tenido acceso a dos trabajos anteriores en los que podemos observar una explicación más completa y donde se incluyen algunas operaciones matemáticas que se han llevado a cabo para obtener estos resultados (Ávila Muñoz & Sánchez Sáez 2010; Villena Ponsoda, Ávila Muñoz & Sánchez Sáez 2014). En ellos vemos como, en realidad, el modelo matemático empleado ha sido la fórmula de Zipf-Mandelbrot. Esta fórmula exponencial fue creada en los años 40 por G. K. Zipf, lingüista de la Universidad de Harvard, y modificada posteriormente por el matemático Benoît Mandelbrot. Se ha aplicado sobre todo a los estudios de corpus lingüísticos o de frecuencias, ya que ayuda a determinar la frecuencia de aparición de las palabras y su distribución. Esta ley afirma que un pequeño número de palabras son utilizadas con mucha frecuencia, mientras que un gran número de ellas son poco empleadas.

La fórmula resultante se puede expresar como:

$$T = \frac{K}{N}$$

Donde **K** es una constante y **N** es la posición que ocupa el término en la lista.

Villena Ponsoda, Ávila Muñoz y Sánchez Sáez (2014: 8) explican “que en un corpus de lenguaje natural la frecuencia de cualquier término es inversamente proporcional a su posición en la ordenación en frecuencia de aparición”, y lo formulan matemáticamente así:

$$m(t) \propto \frac{1}{k}$$

Donde **k** es la posición del término en la lista enunciada por el hablante.

Esta ley se puede aplicar a cualquier lengua con los mismos resultados y se ha popularizado en otros muchos ámbitos. Una de las cuestiones básicas que aporta es la relación entre los porcentajes 80% y 30%. Aplicado a la frecuencia de las palabras, esto significa que con un 30% de los elementos más frecuentes (palabras en este caso) se cubre el 80% de todo el conjunto de los textos. Aunque esta ley ha sido muy difundida, como se puede observar es demasiado simple y poco precisa, motivo por el que el matemático Mandelbrot intentó mejorarla.

Así pues, en la propuesta de estos autores (Ávila Muñoz & Sánchez Sáez 2014: 148-154), el grado de compatibilidad de las palabras es calculado mediante esta fórmula aplicada a los resultados que obtienen en las pruebas de disponibilidad para cada hablante. A continuación los resultados individuales son sumados mediante la fórmula:

$$a + b - a . b$$

Con ella se logra obtener lo que se puede llamar el “espectro léxico”, que representa los niveles de compatibilidad de las palabras de cada centro de interés.

En general, los argumentos y propuestas lanzadas por estos autores resultan de gran interés para nuestro estudio por diferentes motivos. En primer lugar, aunque la metodología usada para describir el comportamiento de las palabras dentro de los centros de interés sea diferente, en ambos casos partimos

de una observación de la estructura de los conjuntos para aislar el vocabulario más y menos disponible. En nuestro caso, en lugar de aplicar la fórmula de Zipf-Mandelbrot a los resultados individuales y sumarlos luego con otra operación matemática, hemos preferido el análisis de los datos en su conjunto a través de histogramas de frecuencias. Su manejo resulta más práctico, están muy extendidos a este tipo de análisis estadísticos, pues no es necesario usar cálculos matemáticos, ya que el propio programa se encarga de realizar gran parte de los cálculos. Además, reducimos el margen de error que conlleva el uso de una fórmula demasiado simple y genérica. En segundo lugar, si bien hemos usado procedimientos diferentes, los resultados a los que se llega van en la misma línea: un pequeño grupo de palabras que se distinguen claramente por su alta disponibilidad; un grupo numeroso de elementos que está en el polo opuesto y que se caracterizan por su baja disponibilidad, y una zona de transición en la que se encuentran las palabras de disponibilidad media, que aumentan levemente de modo gradual. La teoría de los prototipos y los conceptos de accesibilidad y compatibilidad son también aplicables a nuestros datos. En este sentido la propuesta de Ávila Muñoz nos sirve como referente para comparar resultados y validar en cierta medida el método que hemos seguido, ya que los resultados obtenidos son iguales de modo global, aunque se hayan hallado mediante procedimientos diferentes.

Por otro lado, el conocimiento de la teoría de conjuntos difusos (*Fuzzy Sets*) nos ha sido de gran ayuda. En un primer momento, al conocer estos modelos matemáticos intentamos aplicar sus procedimientos a los conjuntos de palabras (centros de interés) desde una primera instancia, es decir, sin recurrir a histogramas. El objetivo era distinguir algún tipo de organización interna o agrupaciones de palabras en los centros de interés. Para ello fue necesario recurrir al uso de un *software* especial creado a partir de algunos de los algoritmos de la lógica difusa⁷⁶. No obstante, el uso de estos programas no arrojó datos relevantes, ya que en realidad la organización del léxico disponible de cada

⁷⁶ Agradecemos el uso de este *software* así como toda la ayuda prestada para la aplicación de la lógica difusa a nuestro trabajo ofrecida por el profesor de la Universidade do Porto Vladimiro Miranda, director del centro de investigación INESC TEC <http://www.inesctec.pt/>. Su asesoramiento resultó fundamental para poder contemplar esta opción en nuestro estudio.

centro de interés muestra dos polos bien diferenciados y un conjunto de elementos graduales entre ambos⁷⁷.

Por este motivo usamos la lógica difusa *a posteriori*, como complemento de los histogramas de frecuencias a la hora de solucionar la cuestión de los cortes rígidos en los conjuntos de palabras. Principalmente resulta útil en la zona central de nuestros gráficos, en la que, como hemos comentado, la línea de evolución de los centros de interés no muestra cambios claros, sino graduales. Esta “zona gris” de los conjuntos de palabras se puede demarcar como una zona de transición, en la que los elementos no muestran un grado de pertenencia clara a un único conjunto, sino que están muy próximos a ambos, por lo que podrían encuadrarse en uno u otro en función de los límites que fijemos. Estas zonas a las que nos referimos suelen coincidir precisamente con los marcos que identificamos con el nivel B1, por lo que hallar una solución satisfactoria refuerza el trabajo que llevemos a cabo. En este sentido, aunque en la práctica no hemos realizado las mismas operaciones y procedimientos indicados por Villena Ponsoda, Ávila Muñoz y Sánchez Sáez (2010), Ávila Muñoz y Sánchez Sáez (2014) y Ávila Muñoz (2016), sus apuntes teóricos sobre los beneficios de aplicar la teoría de conjuntos difusos a la secuenciación de las palabras disponibles nos resultan altamente interesantes, por lo que los hemos seguido en nuestra propuesta en un segundo momento.

A continuación explicaremos qué pasos hemos seguido para secuenciar las palabras en los histogramas, determinar los límites de los niveles de lengua y por último aplicar una partición difusa (*fuzzy partition*) para conseguir divisiones flexibles y no conjuntos estancos. Seguiremos con CI *El cuerpo humano* del estudio de Tomé Cornejo para ejemplificar nuestro modo de proceder.

El primer paso de nuestro método consiste en tomar los datos sobre la disponibilidad de todos los vocablos de un centro de interés y hacer con ellos un histograma de frecuencias. En los histogramas de los Gráficos 3 y 4 habíamos

⁷⁷ En cambio, el programa del profesor Miranda resulta útil para buscar estructuras en las que existan diferentes conjuntos con elementos que no pertenecen al 100% a ninguno de ellos, sino que comparten rasgos propios tanto de uno como de otro.

observado la estructura del CI *El cuerpo humano* en una organización de 10 conjuntos. En el Gráfico 11 se muestra el número de vocablos por cada índice de disponibilidad, agrupados ahora en 20 conjuntos. Con este número de agrupaciones conseguimos afinar con más detalles las divisiones en las zonas en las que el número de vocablos es más numeroso, es decir, allí donde se observa una mayor cantidad de vocablos disponibles. Estas partes coinciden con las palabras que obtienen un índice de disponibilidad menor, por lo que tenemos que pensar que se trata de los términos que se corresponden con los niveles más altos del *MCERL*, el intermedio alto y el avanzado (B2 y C).

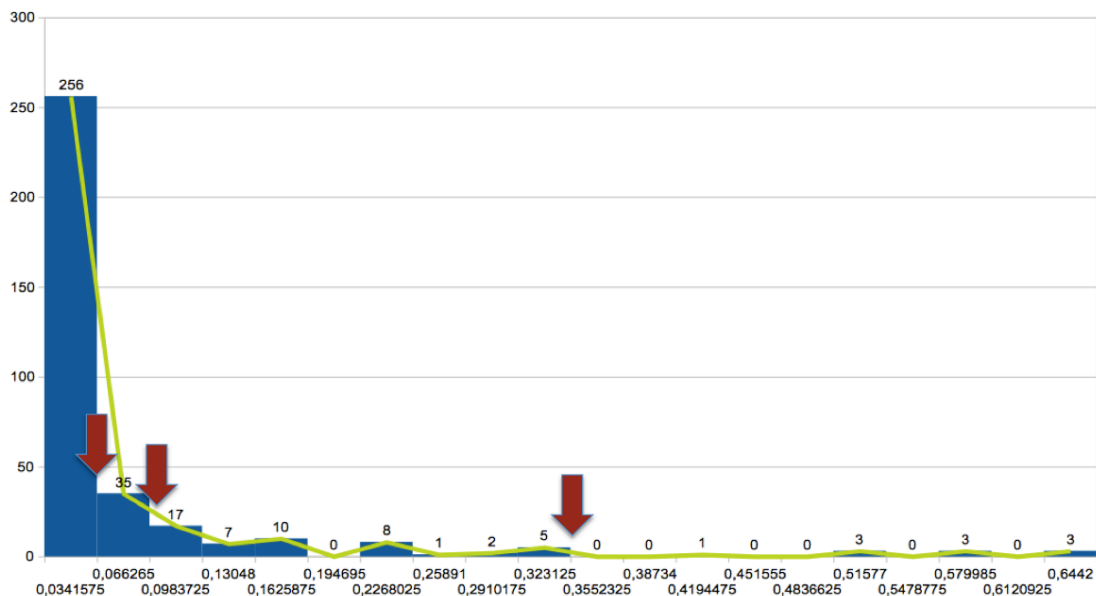


GRÁFICO 11. Histograma de 20 conjuntos con su línea de tendencia a partir de la frecuencia de palabras según su ID en el CI *El cuerpo humano* de Tomé Cornejo (2015).

Después de analizar estructura representada en el Gráfico 11, establecemos los cortes que se corresponden con los niveles de lengua. En este caso son los siguientes:

Palabras de los niveles A1-A2: todas las palabras hasta el índice de disponibilidad de 0,323, es decir, todas las palabras que están antes de ese conjunto en el histograma y que en el centro de interés tienen un índice de disponibilidad mayor a 0,323125.

Palabras del nivel B1: las palabras que están entre el índice de disponibilidad de 0,323125 y el índice 0,0983725.

Palabras de nivel B2: las palabras que están entre 0,0983725 y 0,0341575.

Palabras del nivel C: desde 0,0341575 hasta 0.

Una vez situamos estas divisiones en el histograma de frecuencias, podemos trasladarlos al gráfico que muestra el índice de disponibilidad de cada palabra del centro de interés sin agrupaciones de frecuencia:

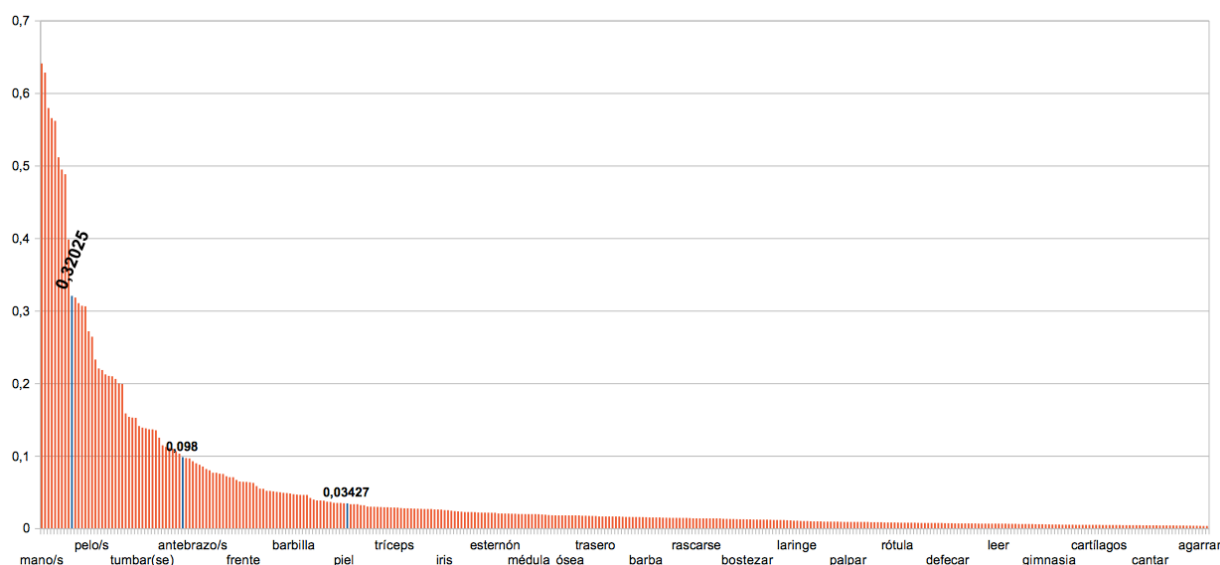


GRÁFICO 12. Límites de selección de las palabras que se corresponden con los niveles del MCER en el CI *El cuerpo humano*, Tomé Cornejo (2015)⁷⁸.

El paso siguiente consiste en trasladar estos parámetros a las palabras del centro de interés, observando cuáles están en un conjunto u otro. En el caso de *El cuerpo humano*, obtenemos estos niveles:

1. Las 11 primeras palabras para el nivel A1-A2, desde la palabra *ojo/s* hasta la palabra *uña/s*.
2. Para el B1, desde *cuello* hasta *pulmón/es*.
3. Para el nivel B2, desde *columna (vertebral)* hasta *olfato*.
4. Para el nivel C, desde *pezón/es* hasta el final de la lista de palabras.

⁷⁸ Recuérdese que en este tipo de gráfico se representa solo el índice de disponibilidad de las palabras de un centro de interés. En el eje de abscisas (x, el horizontal) se encuentran los vocablos, mientras que en el de ordenadas (y, el vertical) se indica el valor de su disponibilidad. Por este motivo, en este gráfico, al contrario que en los histogramas, las palabras menos disponibles se sitúan a la derecha y las más disponibles a la izquierda.

Nivel 1: Palabras muy disponibles, A1 y A2. Elemental

1	<i>ojo/s</i>	0,6432
2	<i>mano/s</i>	0,64063
3	<i>brazo/s</i>	0,628
4	<i>cabeza</i>	0,57916
5	<i>pie/s</i>	0,56525
6	<i>pierna/s</i>	0,56145
7	<i>nariz</i>	0,51132
8	<i>boca</i>	0,49429
9	<i>dedo/s</i>	0,48791
10	<i>oreja/s</i>	0,39802
11	<i>uña/s</i>	0,32025

TABLA 20. Palabras disponibles seleccionadas según su ID para el nivel elemental (A1-A2) dentro del CI *El cuerpo humano* de Tomé Cornejo (2015).

Nivel 2: Palabras bastante disponibles, B1. Umbral

12	<i>cuello</i>	0,31789
13	<i>rodilla/s</i>	0,31017
14	<i>hombro/s</i>	0,30655
15	<i>pecho</i>	0,3059
16	<i>codo/s</i>	0,27148
17	<i>pelo/s</i>	0,26394
18	<i>hueso/s</i>	0,23245
19	<i>tronco</i>	0,22019
20	<i>dientes</i>	0,21807
21	<i>sentar(se)</i>	0,21197
22	<i>espalda</i>	0,20994

23	<i>músculo/s</i>	0,20939
24	<i>muñeca/s</i>	0,20581
25	<i>tobillo/s</i>	0,19949
26	<i>cadera/s</i>	0,1989
27	<i>(estar) de pie</i>	0,15821
28	<i>labios</i>	0,1535
29	<i>cara</i>	0,15245
30	<i>estómago</i>	0,15215
31	<i>andar</i>	0,14085
32	<i>tumbar(se)</i>	0,13865
33	<i>lengua</i>	0,13761
34	<i>muslo/s</i>	0,1363
35	<i>corazón</i>	0,13618
36	<i>comer</i>	0,13495
37	<i>pestañas</i>	0,12477
38	<i>cejas</i>	0,11427
39	<i>extremidades</i>	0,11206
40	<i>ombligo</i>	0,11046
41	<i>cerebro</i>	0,10945
42	<i>saltar</i>	0,10604
43	<i>(estar) sentado</i>	0,10282
44	<i>pulmón/es</i>	0,098

TABLA 21. Palabras disponibles seleccionadas según su ID para el nivel umbral (B1) dentro del CI *El cuerpo humano* de Tomé Cornejo (2015).

Nivel 3: Palabras poco disponibles, B2. Avanzado

45	<i>columna (vertebral)</i>	0,0967
46	<i>levantarse</i>	0,0962
47	<i>antebrazo/s</i>	0,09232
48	<i>caminar</i>	0,08948
49	<i>torso</i>	0,08742
50	<i>mover(se)</i>	0,08503
51	<i>correr</i>	0,08152
52	<i>arteria/s</i>	0,07978
53	<i>barriga</i>	0,07644
54	<i>costillas</i>	0,07638
55	<i>riñón/es</i>	0,07513
56	<i>hablar</i>	0,0748
57	<i>articulación/es</i>	0,07162
58	<i>(estar) tumbado</i>	0,07031
59	<i>agacharse</i>	0,07031
60	<i>tripa/s</i>	0,06649
61	<i>venas</i>	0,06441
62	<i>frente</i>	0,06392
63	<i>pelvis</i>	0,06378
64	<i>bailar</i>	0,0633
65	<i>hígado</i>	0,06243
66	<i>cintura</i>	0,05807
67	<i>talón/es (de Aquiles)</i>	0,05445
68	<i>ver</i>	0,05429
69	<i>tocar</i>	0,05152
70	<i>pene</i>	0,05134

71	<i>respirar</i>	0,05082
72	<i>abdomen</i>	0,05
73	<i>tórax</i>	0,04955
74	<i>oír</i>	0,04876
75	<i>órganos</i>	0,04845
76	<i>intestino/s</i>	0,04794
77	<i>barbilla</i>	0,04683
78	<i>cúbito</i>	0,04632
79	<i>mirar</i>	0,04588
80	<i>llorar</i>	0,04585
81	<i>espinilla</i>	0,04583
82	<i>cráneo</i>	0,04253
83	<i>culo</i>	0,04168
84	<i>vértebra/s</i>	0,04165
85	<i>mejillas</i>	0,03941
86	<i>cruzar las piernas</i>	0,03832
87	<i>sangre</i>	0,03824
88	<i>tendón/es</i>	0,03788
89	<i>pechos</i>	0,03649
90	<i>peroné</i>	0,03624
91	<i>agachado</i>	0,035
92	<i>incorporarse</i>	0,03477
93	<i>esternocleidomastoideo</i>	0,03476
94	<i>piel</i>	0,03437
95	<i>olfato</i>	0,03427

TABLA 22. Palabras disponibles seleccionadas según su ID para el nivel avanzado (B2) dentro del CI *El cuerpo humano* de Tomé Cornejo (2015).

Nivel 4: Palabras casi nada disponibles, C. Competente

96	<i>pezón/es</i>	0,03309
97	<i>besar</i>	0,03308
98	<i>radio</i>	0,03306
99	<i>gemelo/s</i>	0,03178
100	<i>pensar</i>	0,03161
101	<i>bíceps</i>	0,02981
102	<i>cuádriceps</i>	0,0296
103	<i>(estar) erguido</i>	0,02951
104	<i>fémur</i>	0,02936
105	<i>vista</i>	0,02914
106	<i>estirarse</i>	0,02891
107	<i>tibia</i>	0,02889
108	<i>páncreas</i>	0,02862
109	<i>tríceps</i>	0,02849
110	<i>pantorrilla/s</i>	0,02832
111	<i>reír</i>	0,02775
112	<i>cabello</i>	0,02746
113	<i>glúteos</i>	0,02745
114	<i>pestañear</i>	0,02725
115	<i>tacto</i>	0,02692
116	<i>gusto</i>	0,0269
117	<i>abrazar</i>	0,02666
118	<i>húmero</i>	0,02644
119	<i>caer(se)</i>	0,02643
120	<i>encéfalo</i>	0,02643
121	<i>sentir</i>	0,02591
122	<i>(ponerse) de puntillas</i>	0,0258

123	<i>abdominales</i>	0,02566
124	<i>iris</i>	0,02497
125	<i>(sentarse) de rodillas</i>	0,02484
126	<i>nudillos</i>	0,02407
127	<i>beber</i>	0,02349
128	<i>aparato/s reproductor/es</i>	0,02324
129	<i>escribir</i>	0,02268
130	<i>flexionar(se)</i>	0,02224
131	<i>paladar</i>	0,02222
132	<i>encorvado</i>	0,02222
133	<i>pupilas</i>	0,02212
134	<i>sonreír</i>	0,02162
135	<i>(aparato) digestivo</i>	0,02145
136	<i>dormir</i>	0,02143
137	<i>ligamento/s</i>	0,02135
138	<i>derecho</i>	0,02124
139	<i>esternón</i>	0,02124
140	<i>estirado</i>	0,02031
141	<i>temporal</i>	0,02031
142	<i>escuchar</i>	0,0202
143	<i>reproducción</i>	0,02016
144	<i>nadar</i>	0,01992
145	<i>bazo</i>	0,01989
146	<i>genitales</i>	0,01951
147	<i>coxis</i>	0,0195
148	<i>doblado</i>	0,01941
149	<i>colocarse</i>	0,01941

TABLA 23. Palabras disponibles seleccionadas según su ID para el nivel competente (C) dentro del CI *El cuerpo humano* de Tomé Cornejo (2015).

Ilustramos a continuación estos mismos pasos con otro conjunto de unidades diferente, es decir, con otro centro de interés que tenga distintas características internas, debido a su naturaleza y los resultados obtenidos. Escogemos en este caso el de *Alimentación y nutrición*.

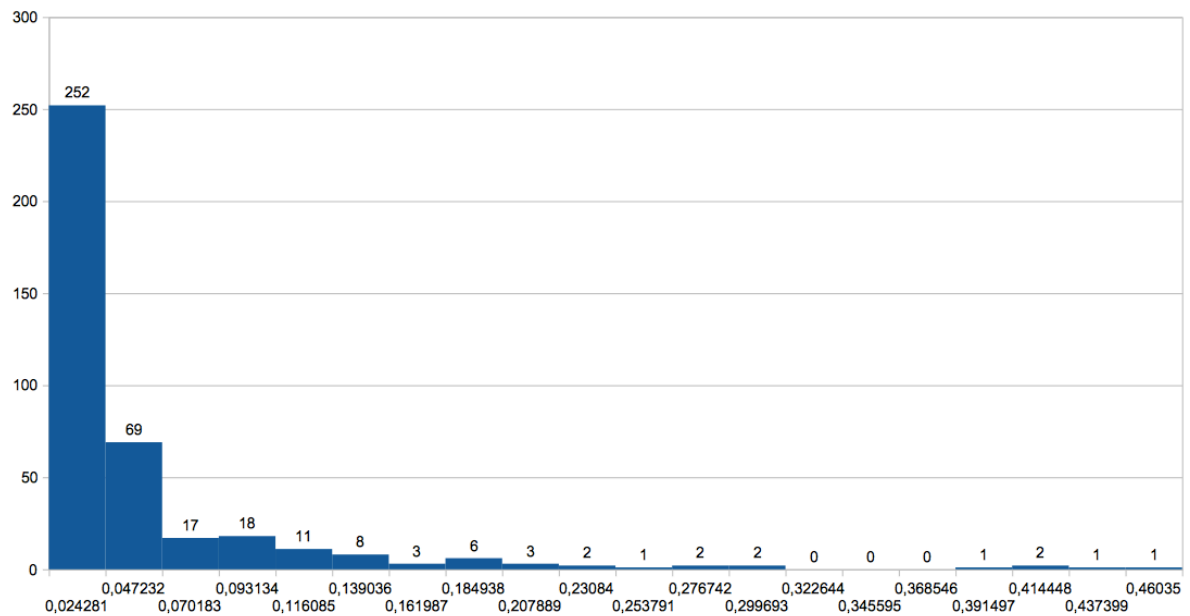


GRÁFICO 13. Histograma de 20 a partir de la frecuencia de palabras según su ID en el CI *Alimentación y nutrición*, Tomé Cornejo (2015).

En este caso las divisiones se sitúan del siguiente modo:

A1-A2 vocablos hasta el del ID 0,139036.

B1 vocablos comprendidos entre el ID 0,139036 y el ID 0,047232.

B2 vocablos comprendidos entre el ID 0,047232 y 0,024281.

C1 desde 0,024281 hasta 0.

Para evitar el problema de los límites rígidos, en un segundo momento de este proceso podemos aplicar una lógica de *fuzzificación* a la escala que hemos definido en cada histograma.

Como hemos dicho, la lógica difusa o *fuzzy* está basada en la teoría de conjuntos difusos (*Fuzzy Set Theory*⁷⁹) y difiere de los sistemas lógicos

⁷⁹ Para la definición, modelos matemáticos y algoritmos, así como sus aplicaciones, hemos recurrido a Zimmermann (2001).

tradicionales en su forma de entender y tratar los universos estudiados. Así, si la lógica tradicional trabaja con universos discretos, desde esta otra perspectiva nos encontramos ante universos "continuos", conjuntos *difusos* definidos por particiones con fronteras diluidas, en lugar de estrictas. Hay muchas diferencias entre un análisis de datos clásico (estático) y otro dinámico. En la lógica difusa, el valor de verdad de una proposición puede ser un subconjunto *fuzzy* de cualquier conjunto ordenado parcialmente, al contrario de los sistemas lógicos binarios, en los que el valor verdad solo puede asumir dos opciones: verdadero (1) o falso (0). En los sistemas lógicos multivalores, el valor verdad de una proposición puede ser un elemento de un conjunto finito, en un intervalo, o un álgebra *booleana*. Además, en la lógica difusa, los valores verdad son expresados lingüísticamente (por ejemplo: *cierto, muy cierto, no verdad, falso, muy falso, etc.*), y cada término lingüístico es interpretado como un subconjunto *fuzzy* del intervalo unitario. Otra diferencia es que en los sistemas lógicos binarios los predicados son exactos (por ejemplo: *igual, mayor que, etc.*), mientras que en la lógica *fuzzy* los predicados son nebulosos (por ejemplo: *alto, bajo, etc.*). En los sistemas lógicos clásicos, el modificador más utilizado es la negación, en cambio en la lógica *fuzzy* son posibles una variedad de modificadores de predicados (por ejemplo: *mucho, bastante, más o menos, poco...*). Estos modificadores son esenciales en la generación de términos lingüísticos (por ejemplo: *muy alto, más o menos, aproximadamente, etc.*) con los que denominamos las unidades que estamos estudiando. En los sistemas lógicos clásicos existen solamente los cuantificadores existenciales (\exists) y universales (\forall). Por el contrario, la lógica *fuzzy* admite, como decimos, una amplia variedad de cuantificadores (por ejemplo: *poco, varios, usualmente, frecuentemente, en torno a cinco, etc.*). La probabilidad, en el contexto de la lógica clásica, es un valor numérico o un intervalo; en la lógica difusa existe la opción adicional de emplear probabilidades lingüísticas (por ejemplo: *probable, altamente probable, improbable, etc.*).

La lógica difusa aporta un modo de razonar que se basa en aproximaciones y no en exactitudes, pero esto no quiere decir que no sea igualmente riguroso. Al contrario, tiene en cuenta la falta de exactitud para describirla y analizarla desde el conocimiento.

La tecnología que deriva de la lógica difusa ha generado aplicaciones y productos en diversas áreas, en cuestiones en las que era más conveniente una visión flexible para analizar los datos. Como área pionera puede mencionarse el control de procesos industriales. Otras que también han bebido de esta línea son la monitorización de pacientes en medicina, por ejemplo durante la anestesia. También en el mantenimiento de máquinas que dependen del Estado o en los índices de las acciones, para examinar la evolución de sus precios.

En nuestra propuesta para la clasificación de las palabras disponibles de cada centro de interés, además de tomar en consideración este tipo de variables, hemos estudiado un enfoque alternativo a lo que podría llamarse *clustering* tradicional, el llamado *clustering* difuso o *fuzzy*, y lo ponemos en relación con los niveles de lengua del *MCERL*. La realidad es que en muchas ocasiones un elemento (sector) puede pertenecer a distintos grupos, aunque con diferentes grados de posibilidad, tal y como se ha subrayado desde esta línea de investigación. En este sentido, el *clustering* difuso flexibiliza las agrupaciones efectuadas y otorga a cada elemento un grado de pertenencia a uno u otro conjunto, lo que enriquece los análisis posteriores. Dicho de otra forma, al seguir estos postulados los conjuntos con límites poco claros se secuencian llevando a cabo una partición difusa. Determinamos el grado de pertenencia relativa que sus elementos tienen con respecto a ambos grupos a los que pueden pertenecer e intentamos establecer en cuál de ellos es más adecuado encuadrarlos.

Los algoritmos de *clustering* son una herramienta eficaz para extraer información de datos en bruto. Son métodos de aprendizaje no supervisado. Su objetivo es dividir los datos en grupos (*clusters*). Básicamente, se trata de obtener los grupos de un conjunto de datos no etiquetados. Se puede expresar matemáticamente de la siguiente forma:

$$X = \{x(1), \dots, x(m)\} \text{ es partir } X \text{ en } C \text{ subgrupos } \{1 < C < m\}$$

de tal manera que cada uno de los grupos represente una subestructura de \mathbf{X} .

Uno de los métodos más conocidos y ampliamente utilizado es el *fuzzy-c means*. El algoritmo *fuzzy-c means* es un algoritmo de *clustering* difuso, cuyas

aplicaciones se llevan a tareas de clasificación de conjuntos. Este algoritmo asigna pertenencias entre 1 y 0 a los elementos, las cuales están inversamente relacionadas con la distancia entre dichos elementos y los centros de los *clusters*. Debido a esto, la mayoría de los algoritmos de *clustering*, incluyendo el *fuzzy-c means*, imponen una restricción que se llama condición probabilista, por la cual la suma de las pertenencias, o probabilidades de que un elemento pertenezca a todos los grupos, debe sumar uno. Estas funciones de pertenencia o distribuciones de probabilidad obtenidas no son capaces de distinguir entre elementos que podrían pertenecer a cualquiera de los grupos, normalmente llamados *inliers*, o los elementos que podrían no pertenecer a ninguno de los grupos, conocidos como *outliers*.

El sentido de nuestra secuenciación, por lo tanto, se encuadra en esta lógica de la Teoría de Conjuntos Difusos (Zimmermann 2001). En concreto, nos interesa la denominada *fuzzy-c partition* (Bezdek 1981: 26, Zimmermann 2001: 290-429). Siguiendo a Bezdek (1981: 26), una *fuzzy-c partition* se define a grandes rasgos porque la suma de las funciones de pertenencia asociadas a cada ítem es siempre igual a 1.

En la definición que ofrece Zimmermann se especifica que \mathbf{X} , \mathbf{Vcn} y \mathbf{C} se consideran parte de una *fuzzy-c partition* si satisfacen las siguientes condiciones (Bezdek 1981: 26, Zimmermann 2001: 290):

1. $\mu_{ik} \in [0, 1] \quad 1 \leq i \leq c, 1 \leq k \leq n$
2. $\sum_{i=1}^c \mu_{ik} = 1 \quad 1 \leq k \leq r$
3. $0 < \sum_{k=1}^n \mu_{ik} < n \quad 1 \leq i \leq c$

Zimmermann añade que todas las matrices que satisfacen esas condiciones se llaman \mathbf{Mfc} . Las condiciones 2 y 3 simplemente indican que el total de los miembros de un elemento se normaliza en 1 y que los elementos no pueden pertenecer a más clústeres de los que existen. El grado de pertenencia de un elemento al conjunto corresponde a un valor que está dentro del intervalo considerado. No podemos estar seguros de qué valor es, por lo tanto se le provee

de un intervalo de confianza. Así, se asignan intervalos de pertenencia para cada elemento a todos los clústeres. Debemos ser conscientes de que cuanto mayor sea la amplitud del intervalo, mayor será la incertidumbre que tengamos respecto a la pertenencia real de ese elemento a ese clúster.

Zimmermann representa esta división a nivel gráfico del siguiente modo:

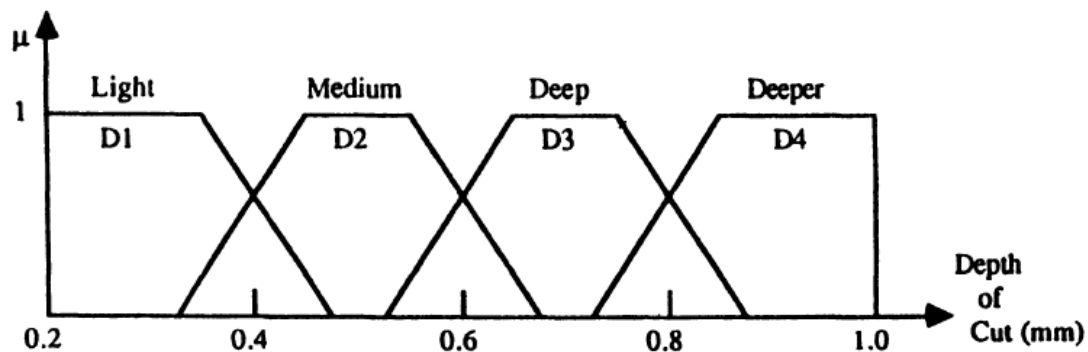


GRÁFICO 14. Representación de una *Fuzzy-C partition* y las etiquetas lingüísticas de los *clusters* (Zimmermann 2001: 304).

Lo que aportamos es una visión de clasificación de las palabras disponibles permitiendo flexibilizar la interpretación de los grupos e informando sobre matices que consideramos interesantes. Aunque tal forma de clasificar no establece grupos rígidos, se puede llevar a cabo una solución que resulta más idónea en estos casos: seleccionar los valores más altos de las funciones de pertenencia y asignar cada sector a uno de los grupos encontrados previamente. Tales grupos se han caracterizado adecuadamente, como hemos comentado al estudiar los centros de interés como un conjunto de aquellas palabras más prototípicas o menos prototípicas en función de su grado de disponibilidad. Así, el primer grupo, el de las palabras que consiguen mayor disponibilidad, es el de las palabras más prototípicas. El último grupo, el mayoritario, se corresponde con las que menos se acercan al prototipo que expresa el centro de interés. En el grupo central están aquellas que son bastante disponibles o poco disponibles y que experimentan un desarrollo gradual, ascendente en número de elementos y descendente en nivel de disponibilidad o prototipicidad. Para llevar a cabo estos pasos, nuestra propuesta es seguir las indicaciones y los algoritmos explicados por Bezdek (1981) y Zimmermann (2001: 290-306).

En nuestro caso, proponemos usar los histogramas de frecuencia para establecer cuatro grupos y usar etiquetas lingüísticas para denominarlos. Hemos optado por las etiquetas: *muy disponibles*, *bastante disponibles*, *poco disponibles*, *casi nada disponibles*. Los límites que hemos definido en los histogramas son los puntos que nos sirven de anclaje para cada *cluster*, los que nombramos con la letra C. A partir de ellos se puede determinar el centro de cada *cluster* o grupo y establecer una gradación de elementos en el intervalo de corte. Este intervalo demarcado supone la frontera difusa entre un conjunto y otro.

Gracias a este modelo, podemos asignar un valor a cada elemento que está en esta horquilla de tal modo que conseguimos cuantificar en qué medida se relaciona con el punto intermedio de los dos conjuntos a los que se puede asignar. Estos intervalos u horquillas, que suponen la periferia de cada conjunto, evolucionan gradualmente con respecto a los centros. Gráficamente lo podemos representar de la siguiente manera:

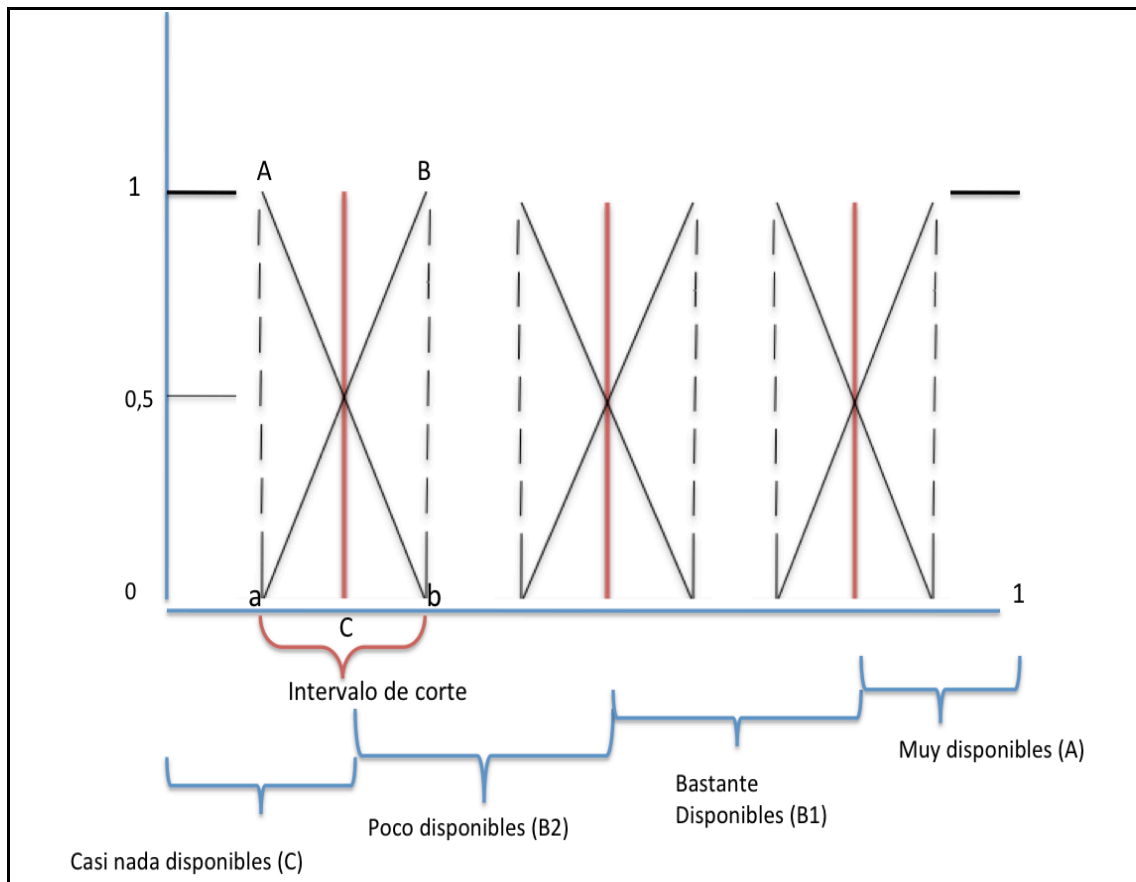


GRÁFICO 15. *Fuzzy C partition* de las palabras disponibles y etiquetas lingüísticas acuñadas para cada grupo.

En la parte inferior de la imagen observamos cuáles son los 4 conjuntos que hemos establecido y las etiquetas escogidas para cada uno de ellos. Las palabras *Muy disponibles* se corresponden con las más prototípicas con respecto al concepto que enuncia el centro de interés, y consideramos que se encuadran en un nivel elemental de la lengua. Las líneas rojas (puntos denominados **C**) indican los bordes escogidos para cada conjunto, establecidos a partir de los histogramas correspondientes. Lo que denominamos “intervalo de corte” es la zona en la cual hay elementos que pueden pertenecer tanto a un grupo como al otro. La distancia entre el segmento **a** y **C** es siempre la misma que hay entre **C** y **b**. Así mismo, este intervalo de corte tiene las mismas medidas en el resto de conjuntos del histograma. La recta que une los puntos **A** y **b** y la que une el segmento **B** con **a** tienen el mismo valor, que es siempre 1. Su punto intermedio, donde se cruzan, es por lo tanto 0,5. Con esto podemos establecer que los elementos que estén entre **b** y **C** pueden pertenecer al conjunto de las palabras *Poco disponibles* o bien al grupo de las *Casi nada disponibles*, en función de nuestros intereses. Sin embargo, podemos graduar y expresar matemáticamente en qué medida pueden pertenecer a un grupo y qué nivel de pertenencia tienen con respecto al otro. Con esto podemos escoger en qué grupo incluimos los elementos dudosos sin caer en la arbitrariedad. Estas decisiones obtienen un rigor científico avalado por los algoritmos que sustentan el modelo matemático de la lógica difusa.

La progresión de los elementos se comporta como una transición entre ambos conjuntos. Cuanto más cerca está un elemento de **a**, más probabilidades tiene de pertenecer al conjunto de las unidades *Casi nada disponibles*. Al contrario, cuanto más próximo está de **b**, la probabilidad de que pertenezca al grupo de las *Poco disponibles* aumenta. Si un elemento se sitúa en **C**, en este caso diremos que tiene las mismas probabilidades de pertenecer a un conjunto o al otro.

Quien opte por llevar a cabo estos procedimientos que aquí proponemos debe tener en cuenta varias cuestiones. En primer lugar, hay decisiones que debe tomar cada investigador, es decir, usando histogramas de frecuencias hemos

establecido los conjuntos que consideramos oportunos, así como sus cortes, pero han sido resultados escogidos por nosotros. No obstante, gracias al método estadístico de los histogramas, hemos podido analizar cada centro de interés como un conjunto de elementos y observar sus características, por lo que esas decisiones no han sido tomadas aleatoriamente.

Igualmente, si se desea aplicar el modelo de la lógica difusa que hemos presentado para la selección de las unidades lingüísticas de cada centro de interés, debemos ser conscientes de que las decisiones sobre cuáles son los puntos de corte y cuál es la zona difusa de transición (intervalo de corte) también queda a elección del responsable de cada investigación. Sin embargo, y dada la naturaleza de las palabras que conforman los centros de interés, hemos considerado que esta forma de proceder es más apropiada para este caso que los métodos que ofrecen resultados rígidos y absolutos. Con esto, no estamos optando por una fórmula matemática que nos dé respuestas categóricas y fijas, sino que aplicamos un modelo que describe matemáticamente el comportamiento de los centros de interés, permitiendo que sea quien lleve a cabo cada estudio concreto quien lo aplique en función de sus necesidades e intereses. Es, por lo tanto, un método matemático adaptable a las diferencias, que se puede usar en todos los casos, pero que no nos ofrece unas cantidades estáticas por las que optar para dividir las unidades, sino que nos proporciona un método a seguir. Es igualmente riguroso, pues sus criterios se expresan con algoritmos y sus argumentos son científicos, lo cual avala las elecciones por las que se opte. De hecho, los centros de interés no funcionan igual en todos los casos, ni tan siquiera dentro de un mismo estudio. Al contrario, hay centros de interés en los que las palabras más disponibles superan por muy poco el índice de disponibilidad de 0,5, que podríamos considerar como el punto intermedio. Es decir, las palabras que están más arriba en la lista obtienen una disponibilidad muy baja, con lo que aplicarles la etiqueta de *Muy disponibles* no resulta del todo satisfactorio. Así pues, si ya parece complicado clasificarlas en el primer grupo de los que hemos definido, más difícil sería proponer unas cifras fijas para demarcar conjuntos. Por lo tanto, no creemos que esta forma de proceder, que hasta ahora

había sido la mayoritaria a la hora de seleccionar las palabras disponibles, sea la más adecuada.

Este aspecto también nos lleva a reflexionar sobre la validez de algunos de los centros de interés, en la medida en que sus datos alcanzan valores muy bajos en general y la mayor parte de los vocablos están más cerca de 0 que de 1 (que es el intervalo con el que funciona el índice de disponibilidad). Esto realmente nos indica que en general son palabras poco disponibles, por lo que usar este método de selección resulta dudoso. Después de llevar a cabo estas consideraciones, también somos más conscientes de hasta dónde podemos beneficiarnos de la disponibilidad léxica y en qué ocasiones debemos explorar otras vías.

8.3.4.3 Selección de las multipalabras

Como venimos diciendo a lo largo de este trabajo, uno de los problemas con los que se encuentran los docentes de una segunda lengua es la nivelación léxica, es decir, decidir qué palabras se enseñan en un nivel inicial y cuáles en un nivel intermedio. Quizá la cuestión de qué unidades léxicas enseñar en un nivel avanzado resulte más fácil de discernir, mientras que la dificultad se incrementa cuando tratamos el tema de la combinación entre palabras; lo que solemos llamar expresiones pluriverbales, o, más comúnmente, colocaciones. En los capítulos iniciales hemos abordado ya la importancia de las combinaciones léxicas, especialmente para la traducción y en el aprendizaje de una LE/L2, por lo que sabemos que tal importancia está ampliamente reconocida hoy en día en las investigaciones sobre el desarrollo de la competencia léxica y la enseñanza del vocabulario. Somos conscientes, además, de que aprender una palabra significa mucho más que conocer su significado, pues debemos adquirir otros aspectos como su uso, su concordancia y su coocurrencia con otras palabras (Nation 2001).

A pesar de ello, consideramos que existen muy pocos recursos que nos ofrezcan información sobre qué colocaciones enseñar, cuáles son las más

frecuentes o cuáles debemos introducir en cada nivel. En la mayor parte de las lenguas esta falta de recursos es notable, algo que también observamos en español. Por el momento, contamos con *Redes. Diccionario combinatorio del español contemporáneo* (Bosque 2004), el primer diccionario combinatorio del español. Es innovador en su enfoque, pues en él los predicados se convierten en núcleo y son el principal objeto de estudio. Dos años más tarde sale a la luz el *Diccionario combinatorio práctico del español contemporáneo*, una versión práctica y reducida del anterior, que se ofrece como una obra más manejable (Bosque 2006). Aunque ninguno de ellos se centra específicamente en el ámbito de la enseñanza de una lengua extranjera, cosa que sí ocurre por ejemplo en inglés, lengua pionera en este tipo de estudios, desde la publicación de *Oxford Collocations Dictionary for Students of English* (Crowther, Dignen & Lea 2002). El español cuenta también con un recurso gratuito en línea, aunque por ahora dedicado únicamente a los nombres de sentimiento: el *DiCE, Diccionario de Colocaciones del Español*, herramienta que también aporta actividades destinadas al aprendizaje. Esta base de datos de colocaciones está fundamentada en la lexicología explicativa y combinatoria (Mel'čuk *et al.* 1995).

No obstante, la existencia de estos recursos resulta insuficiente a la hora de poder trasladar las colocaciones al aula, pues, al igual que ocurre con el resto del vocabulario, es necesario un trabajo de selección y secuenciación por niveles que resulta todavía más complejo en este caso. Así se han manifestado algunas voces, al escribir sobre la necesidad de un diccionario de colocaciones para aprendientes de ELE (Hiqueras García 2004b y 2005). La información que aporta el *DiCE* es altamente útil para esta labor, ya que ofrece información sobre la frecuencia de las colocaciones, de tal modo que podemos escoger las que son más usuales. En él encontramos las colocaciones más frecuentes de los nombres de sentimiento, así como sus derivados semánticos. Obtenemos información sobre la frecuencia de cada una de las palabras base contenidas en el diccionario (212 nombres de emoción). En este sentido, los diccionarios de colocaciones pueden jugar un papel muy interesante para la enseñanza de una L2 (Alonso Ramos 2008). Para hacerlo posible, los miembros que forman el proyecto de investigación COLOCATE (dirigido por Alonso Ramos desde el 1999 en la

Universidade da Coruña) han tenido que realizar una labor minuciosa de búsqueda en corpus textuales para realizar análisis cuantitativos de cada una de las combinaciones recogidas en el *DiCE*. Para obtener los datos emplearon uno de los corpus del español con más de 2.000 millones de formas, que está incluido en el sistema de búsqueda *Sketch Engine* (Kilgarriff *et al.* 2004a y 2004b).

En español no contamos con corpus anotados que tengan como unidad básica también las colocaciones, por lo que es necesario realizar los trabajos a partir de numerosas búsquedas de coapariciones en los corpus existentes. Lo ideal, sin duda, sería contar con una herramienta como el *DiCE*, pero del español general y con una cantidad de formas registradas de un volumen considerable, ya que en esta interfaz se pueden visualizar los resultados con la función de “ordenar por frecuencia”, con lo que obtenemos una lista de colocaciones que clasifica las distintas combinaciones de mayor a menor frecuencia. En el caso del *DiCE*, por ejemplo, si se buscan combinaciones partiendo de la palabra *miedo* para transmitir el sentido de “causar ~ en alguien”, obtenemos que es mayor la frecuencia de *dar miedo* o *meter miedo* que la de *sembrar miedo*, *inspirar miedo* o la expresión fusionada *amedrentar*.

Es por este motivo por lo que incluir las colocaciones en nuestra selección de palabras frecuentes excede los objetivos de nuestro trabajo. Sin embargo, somos conscientes de que hoy en día la importancia de las colocaciones es generalmente reconocida, especialmente en la traducción y en el aprendizaje de una L2, y aunque consideramos que las herramientas colocacionales son todavía escasas, no debemos dejar de nombrar aquellos recursos que ya existen, algunas herramientas que nos pueden servir como fuentes y apuntar criterios de selección. De esta forma, posibilitamos que los docentes y los investigadores pueda guiarse por ellos a la hora de escoger las colocaciones que llevar al aula de ELE. Queremos, además, potenciar esta tendencia y apoyar la enseñanza explícita de este tipo de unidades, pues es una carencia que se observa en general y que avalan otras investigaciones. Un estudio reciente que se centra en el tratamiento que se da a las colocaciones léxica en libros de texto de inglés y español como L2, muestra como resultado que casi hay una ausencia de criterios explícitos de selección del léxico y un desconocimiento acerca del concepto de colocación

léxica por parte del profesorado, lo que se traduce en la inexistencia de la enseñanza explícita de colocaciones léxicas en los libros de español. Esto supone ignorar los consejos que las investigaciones de la lingüística aplicada aportan. Al mismo tiempo, cuando se incluyen, se les da un tratamiento tradicional en su presentación y en su práctica. Estos datos son fruto de un análisis de 24 manuales, pertenecientes a tres niveles, junto con 208 cuestionarios distribuidos entre el profesorado de ambas lenguas, inglés y español (López Jiménez 2013). Otra autora que explica las ventajas de llevar al aula las colocaciones es Higuera García (2006). En varios de sus trabajos podemos encontrar información sobre esta y otras cuestiones relacionadas con las unidades pluriverbales.

Antes de aportar recursos, criterios y fuentes de información para la selección de colocaciones, empecemos por explicar muy brevemente a qué nos referimos con este término. En las últimas décadas, el foco de atención de los trabajos sobre vocabulario ha pasado de la palabra aislada a las combinaciones usuales que forman las palabras en su contexto natural de uso. Así se puede ver en los trabajos de Firth (1957), Halliday (1966) y Sinclair (1966, 1991) sobre las combinaciones léxicas en inglés. Estos autores han subrayado el vínculo entre el léxico y la gramática y han promovido estudios sobre unidades de significado superiores a la palabra gráfica no solo en inglés sino también en otros idiomas. El desarrollo de la lingüística de corpus también ha tenido un papel relevante en este replanteamiento, pues ha permitido identificar y analizar las colocaciones como aquellos grupos de palabras que coexisten con elevada frecuencia. La lingüística del corpus ha hecho posible el tratamiento computacional de grandes cantidades de textos orales y escritos que tenemos almacenados en formato digital, por lo que la cantidad y rapidez de textos tratados ha aumentado hasta cifras sin precedentes. En español estos estudios son más recientes, aunque también están empezando a recibir gran atención.

Se pueden considerar pioneros los de Alonso Ramos (1994) y Corpas Pastor (1996). La primera considera que la relación entre las palabras que forman una colocación consiste en que uno de los lexemas de la combinación, la base, selecciona al otro lexema, el colocativo. Corpas Pastor coincide con lo anterior en cuanto a la direccionalidad de la restricción léxica de las colocaciones

en su *Manual de fraseología española* (1996), donde define las colocaciones como “sintagmas libres, generados a partir de reglas, que presentan cierto grado de restricción combinatoria determinada por el uso”. Señala, así mismo, el carácter estable y prefabricado de estas unidades léxicas pluriverbales, y su fijación arbitraria condicionada por el uso repetido. En otros estudios posteriores se destaca también la tipicidad (Koike 2001) o bien la regularidad sintáctica y la falta de transparencia semántica (Zuluaga 2002). Sin embargo, como no existe una definición que esté aceptada de modo general por los diferentes estudiosos del tema, seguiremos un modelo operativo, que se basa fundamentalmente en la coocurrencia léxica restringida. Según esto, una colocación es un tipo de unidad fraseológica, que consta de dos unidades léxicas, una base y un colocativo. La aparición de una restringe la selección de la otra unidad. Así, en las colocaciones se da la máxima de la selección irrestricta, una característica más propia de los sintagmas libres: la base A es el núcleo semántico de la colocación y se selecciona libremente por su significado, mientras que el colocativo B es seleccionado en función de la base. La unidad léxica B puede significar ‘B’ normalmente o solo en combinación con A. En el primer caso, estamos por ejemplo ante el nombre de *loncha* que funciona como colocativo seleccionado por las bases *queso*, *jamón*, *chorizo* o *tocino* para expresar el significado ‘porción de’. En el segundo caso, también con el significado de ‘porción’, tenemos ejemplos como *diente* al ir con la base *ajo* (Alonso Ramos 2013). Koike ofrece en su estudio una tipología de las colocaciones partiendo de su estructura. En esta clasificación se distinguen 5 tipos de colocaciones simples (y sus subtipos) y otros 5 de colocaciones complejas (Koike 2001):

A. Colocaciones simples

1. Sustantivo + verbo

1.1. Sustantivo (sujeto) + verbo: *la lluvia amaina*

1.2. Verbo + sustantivo (compl. directo): *saciar el hambre*

1.3. Verbo + preposición + sustantivo: *viajar en avión*

2. Sustantivo + adjetivo

2.1. Atributo: *sol radiante*

2.2. Predicativo: *el hotel estaba completo*

3. Sustantivo + de + sustantivo: *diente de ajo*
4. Verbo + adverbio: *comer parcamente*
5. Adverbio + adjetivo/participio: *altamente cualificado*
6. Verbo + adjetivo: *sentirse cansado*

B. Colocaciones complejas

1. Verbo + locución nominal: *dar una importante prioridad*
2. Locución verbal + sustantivo: *llevar a cabo un proyecto*
3. Sustantivo + locución adjetival: *salud de hierro*
4. Verbo + locuc. adverbial: *mantener fuera del alcance de los niños*
5. Adjetivo + locución adverbial: *sordo como una tapia*

También se recurre a las diferencias entre lenguas para ejemplificar el carácter restringido e “idiomático” de las colocaciones: *dar un PASEO / faire une PROMENADE / fare una PASSEGIATA, FUMADOR empedernido / gros FUMEUR / FUMATORE incallito o diente de AJO / gousse d'AIL/ spiccio d'AGLIO* (Alonso Ramos 2013). Gracias al conocimiento que tiene un hablante nativo sobre el funcionamiento de su lengua, reconoce que no es normal en español la frase “*A Juan le introdujeron ganas de salir. Cuando llegó a la calle, albergó un chasco porque vio que llovía*”; mientras que una persona extranjera debe aprender que las combinaciones **introducir ganas* y **albergar un chasco* no son las idóneas en español, pero sí *entrar ganas* y *albergar esperanzas*. Por eso, la importancia de las colocaciones no se limita solamente a la descripción lingüística. En el campo de la enseñanza de lenguas extranjeras, algunas investigaciones han reconocido el valor del aspecto colocacional en la adquisición del vocabulario de la lengua meta. Un nativo sabe por intuición qué palabra combinar con otra para producir enunciados correctos, pero “el desconocimiento de las restricciones combinatorias de las palabras en una lengua separa al no nativo de las producciones del nativo, precisamente porque las colocaciones están fijadas en la norma” (Higueras García 2004b: 481).

La noción de colocación léxica y el marco cognitivo en el que aparece revoluciona la perspectiva didáctica de la enseñanza-aprendizaje de léxico e

incide en la importancia de llevar al aula no sólo las piezas léxicas sino también las reglas que rigen su combinación. El número total de combinaciones posibles de las palabras de la lengua española puede superar los cientos de miles, y no es posible ni necesario para un alumno aprenderlas todas. Resulta fundamental planificar su incorporación en las clases de ELE para decidir qué colocaciones enseñar en cada nivel. En español no contamos con una planificación de este tipo, pues la incorporación de las expresiones pluriverbales en el *PCIC* no es sistemática ni responde a criterios de selección objetivos.

En el ámbito de la enseñanza-aprendizaje del inglés como lengua extranjera, ya lo hemos comentado, sí existen herramientas que aportan más información sobre las colocaciones. Nos hemos referido a las ventajas que ofrece en esta lengua el *English Vocabulary Profile* (EVP), en el sentido en el que ofrece una organización del vocabulario basada en estudios de investigación previos y que tiene en cuenta los niveles del *Marco*. Esta herramienta, elaborada por la Universidad de Cambridge, ofrece la ventaja de asignar niveles a cada uno de los sentidos de cada lema. Con respecto al tema que nos interesa, su virtud es que, además de las 7.000 entradas clásicas, incorpora expresiones fraseológicas y lo que se llaman en inglés *phrases*, unidades pluriverbales de diverso tipo. Así, la búsqueda avanzada dispone de un filtro que selecciona solo *phrases*. Una vez seleccionados el tema y el nivel, se obtiene la lista de vocablos, así como los resultados de expresiones que nosotros tratamos como colocaciones, por ejemplo, *stage a show* (C2), *on air* (C1) o *horror film* (B1). También se incluyen otras expresiones fraseológicas como *the media* (B2), *the news* (B1) o casos de preposición regida por un verbo como *appear in/on/at* (B1), que tampoco aparecen en otros diccionarios de colocaciones, como el *Oxford Collocations Dictionary*.

Pero en el caso del español, por los motivos que se han ido presentando, el esfuerzo de planificación de la enseñanza de las multipalabras que se requiere reside en gran medida en los docentes. Existen manuales que incorporan colocaciones en sus unidades didácticas, así como alguna obra específica (por ejemplo De Prada, Salazar & Molero 2012), o algunas propuestas didácticas más puntuales en sitios web colaborativos de materiales de ELE. En los manuales, el

criterio para incorporar las colocaciones suele ser el tema o el vocabulario que se reúne en cada unidad, mientras que en las propuestas puntuales que encontramos en internet suelen estar centradas en áreas temáticas concretas.

Como también ocurre con el vocabulario en general, los criterios de la rentabilidad y utilidad suelen jugar un papel fundamental para seleccionar aquellas combinaciones léxicas que debemos enseñar. Sin embargo, qué entendemos por combinaciones rentables sigue siendo la gran cuestión pendiente de resolver. La bibliografía existente sobre este tema ofrece una serie de criterios que pueden resultar altamente útiles para conseguir este objetivo. En función de los autores, hay diferentes maneras de restringir el número de colocaciones que presentamos en el aula. Según consultemos a uno u otro podemos limitarnos:

- a las combinaciones más frecuentes y generalizadas (Luque Durán & Manjón Pozas 1998, Alonso Ramos 2012),
- a las que incluyan los sustantivos más relevantes (Lewis 1997b),
- a las que tengan un grado de cohesión media (Higueras García 2006),
- a las que pertenezcan a un área temática concreta (Higueras García 2006)
- o a las que ofrecen contrastes entre la lengua materna y la lengua meta del alumnado (Bahns 1993; Molero & Salazar 2013).

Si optamos por el criterio de la frecuencia, podemos por ejemplo recurrir al *DiCE* en aquellos casos en los que ya se contemplan en este diccionario o bien realizar un trabajo de búsquedas concretas en los corpus léxicos existentes, usando la opción de las coocurrencias. En este caso, la recomendación puede ser partir de las palabras con una frecuencia mayor y observar cuáles son los colocativos que suelen aparecer más veces con ellas como bases. A continuación, estas unidades léxicas se pueden añadir a la lista de palabras más frecuentes y ordenarlas igualmente por niveles y temas. Algunos investigadores defienden que es imprescindible implementar las listas de frecuencia con la información sobre locuciones, expresiones fraseológicas o colocaciones e incluso desambiguar semánticamente las unidades léxicas de los corpus antes de elaborar las listas de frecuencias léxicas (Alonso Ramos 2012; Sánchez 2000).

Casos como los que ya se han comentado de *vez* y *embargo* son un buen ejemplo de lo útil que resultaría contar ya con corpus textuales de este tipo, pues sabido es que la razón de que estos sustantivos estén entre los más frecuentes es que estas formas son recurrentes en locuciones como *a la vez*, *a veces*, *de una vez*, *de vez en cuando*, *tal vez*, etc. y como *sin embargo*. Realmente, sería mucho más útil conocer la frecuencia de todas estas locuciones con independencia del sustantivo que contienen. Por otro lado, para aprender el verdadero uso de palabras como *ducha* o *cepillarse* es necesario saber combinarlas con los verbos. Así, *ducha* coocurre con *darse*, *pegarse* y *tomar*; *cepillarse* es sinónimo de *lavar* cuando va con *dientes* pero no cuando va con *pelo* y la frecuencia en el corpus de *cepillarse* con *pelo* es diferente de la de *cepillarse* con *dientes*. En el caso de verbos usados muy comúnmente como soporte (también llamados verbos de apoyo), es más útil partir de los nombres que los seleccionan léxicamente, pues así se puede llegar a las expresiones *dar un paseo*, *hacer un viaje* o *tomar una decisión*. En conclusión, incluir esos cuatro verbos como prioritarios en el aprendizaje del vocabulario sin los nombres asociados no parece adecuado, porque muchas de las palabras frecuentes lo son porque son escogidas frecuentemente como colocativos de bases que son menos frecuentes.

No obstante, también encontramos aportaciones que se basan en una lógica diferente y proponen partir de algunos de los verbos soporte más comunes para observar las diferencias de uso en contraste con otras lenguas, o en concreto con la lengua materna de nuestros aprendices (Molero & Salazar 2013). En este caso, el criterio es seleccionar las colocaciones que nos resulta interesante enseñar desde una óptica contrastiva, considerando las interferencias de las lenguas maternas en los errores colocacionales que cometen los aprendices de ELE. Es lo que ocurre con colocaciones que presentan disimilitud entre la L1 y la L2, ya que la traducción directa puede llevar a errores, como **pagar atención* en lugar de *prestar* o **hacerse/convertirse en consciente* para decir *tomar conciencia de*. Una forma de decidir las colocaciones a las que daremos prioridad en la enseñanza es partir de la “fraseología contrastiva”, disciplina cuyo objetivo es “analizar las analogías y las divergencias de los

sistemas confrontados, mediante criterios que sirvan de base en la comparación interlingual” (Navarro 2007: 4-5).

Si antes defendimos la explotación de los corpus textuales para complementar las listas de frecuencias léxicas con las locuciones y colocaciones de las palabras más frecuentes, otra estrategia que podemos practicar es el uso de los corpus de aprendices. A falta de corpus que contemplen las unidades léxicas pluriverbales en su proceso de etiquetaje o de herramientas de nivelación del vocabulario, en español sí contamos ya con varios corpus de estudiantes de ELE. Dos herramientas de este tipo en español son:

- *Corpus Escrito del Español L2 (CEDEL2)*: parte de un proyecto de investigación de la Universidad Autónoma de Madrid y la Universidad de Granada que investiga el aprendizaje del español por parte de nativos de inglés. Cuenta con la participación de hablantes nativos de español para tomarlos como referencia. Es una gran base de datos con más de 730.000 palabras y más de 2.300 participantes (alrededor de 1.700 hablantes nativos de inglés que están aprendiendo español y sobre 660 hablantes nativos de español).
- *Corpus de Aprendices de Español como lengua extranjera (CAES)* (Rojo & Palacios 2014): es un conjunto de textos escritos producidos por estudiantes de español con diferentes grados de dominio lingüístico (niveles A1 a C1 del *Marco*), procedentes de seis L1: árabe, chino mandarín, francés, inglés, portugués y ruso. Fue lanzado en octubre de 2014.

Estas bases de datos se pueden usar como recurso para estudiar el empleo de las colocaciones que hacen los discentes, con el objetivo de extraer información que nos permita clasificarlas por escalones, observando las que ya dominan y las que todavía no controlan en cada nivel. Esta vía ha sido explotada por Alonso Ramos a partir del CEDEL2 (2012) para esbozar una primera tipología de errores colocacionales. Así, el etiquetado de un corpus de aprendices ha servido a los autores del *DiCE* para obtener datos sobre el uso real que los aprendices hacen de las colocaciones, que podemos comparar con la frecuencia

de uso que se observa en otros corpus textuales de hablantes nativos. Además, gracias al CEDEL2 o al CAES tenemos información de diferentes lenguas maternas, por lo que se puede tomar estas bases de datos como fuentes de información al respecto para luego aplicar el criterio de la lingüística contrastiva.

El equipo de investigación del proyecto de la Universidade da Coruña ColocaTe⁸⁰ también está llevando a cabo trabajos que arrojan información interesante sobre el nivel en que encuadrar las colocaciones con respecto a los 6 que contempla el *MCERL*. Para ello, realizaron un test de nivelación a 57 profesores de ELE en el que debían ordenar por niveles, según su criterio, 56 unidades léxicas del *DiCE* y 36 colocaciones que toman como base la palabra *alegría*. Para ello se les facilitó la palabra en contexto como por ejemplo, *alegría* referida a “sentimiento” en la frase: *Siento una **alegría** intensa*. Los resultados muestran que la tendencia general entre los docentes es incluir las unidades proporcionadas en niveles más bajos que los que el *PCIC* recomienda. Además, existe cierta correspondencia entre la frecuencia de las unidades léxicas, los niveles que les asignan los encuestados y los niveles de referencia del Cervantes. La conclusión es que a medida que la frecuencia disminuye, los niveles más altos aparecen con mayor regularidad.

⁸⁰ Se pueden observar los resultados completos de esta aproximación en el blog del grupo ColocaTe <http://collocate.blogspot.pt/>. Más en concreto encontramos la entrada con el título “Qué colocaciones debemos enseñar en cada nivel” en el link: <http://collocate.blogspot.pt/search?updated-max=2014-11-27T14:04:00%2B01:00&max-results=7&reverse-paginate=true>

Capítulo 9. Propuesta de vocabulario fundamental del español

9.1. Selección y organización del vocabulario

Para elaborar nuestra propuesta hemos seguido la metodología y procesos descritos más arriba. En concreto, hemos considerado las obras de léxico básico y léxico disponible descritas en anteriores apartados (*vid. supra* § 8.3.2.1. y 8.3.2.2.) y que resumimos a continuación.

Del léxico básico extrajimos las primeras 3.000 palabras más frecuentes del volumen de Davies (2006), así como las que en su obra se incluyen en los recuadros de vocabulario temático dentro de las 5.000. De su “thematic vocabulary list” hemos considerado solo las siguientes categorías: cuerpo, adjetivos de emoción, familia, comida, profesiones, deportes, ropa, medios de transporte, el tiempo, naturaleza y plantas, animales, colores. En su trabajo, Davies no muestra prácticamente las locuciones, colocaciones, etc. Lo que sí hace es indicar algunas de ellas junto a las definiciones de los términos que recoge. Por ejemplo, en la posición 180 de su lista aparece la palabra *embargo*. Evidentemente esta palabra no tendría una frecuencia tan alta si no fuera porque la usamos en la locución conjuntiva *sin embargo*. Lo que hemos hecho nosotros es buscar esta información dentro de las 3.000 palabras que hemos seleccionado y ofrecerla por separado, en el lugar indicado. En el caso de *sin embargo*, la añadimos a la información que ofrecemos sobre las conjunciones. Dejamos, asimismo, *embargo* en los sustantivos. Además, hemos enriquecido esta nómina con algunas locuciones, colocaciones y expresiones que obtuvimos en la obra mencionada de Ávila Muñoz (1999). Este autor registra algo más de 100 expresiones de este tipo en su trabajo sobre frecuencias léxicas en el habla de Málaga. Hemos escogido las más adecuadas y las añadimos en la categoría

correspondiente. La mayor parte de ellas son locuciones adverbiales, locuciones preposicionales y algunas locuciones verbales, conjuntivas e interjectivas.

Sobre el léxico disponible, contamos con las palabras que clasificamos hasta el nivel B1 en los histogramas de frecuencias de los estudios de López Meirama (2008) y Tomé Cornejo (2015). Para esta propuesta hemos usado solamente los límites establecidos en los histogramas, sin llevar a la práctica cálculos para ampliar o disminuir el número de unidades seleccionadas siguiendo el procedimiento de la lógica difusa que hemos expuesto más arriba. Tratándose de un trabajo que pretende probar los métodos de selección de forma global y con un volumen de palabras y centros de interés de tamaño tan elevado, hemos optado por esta vía más práctica, pues por una cuestión de espacio y tiempo, llevar a cabo este segundo paso supondría una labor de mayor calado a nivel matemático, lo que está fuera de los objetivos y extensión de esta tesis. No obstante, creemos que su uso puede resultar altamente útil en otros contextos y estudios, e intentaremos acometerlo con más profundidad en futuros trabajos.

El número de palabras que incluimos en el vocabulario fundamental del español que aquí proponemos está dentro de los límites que hemos establecido para el nivel B1: entre 3.000 y 5.000 unidades. Al acometer la fusión de los dos tipos de léxico y organizarlos temáticamente, no contabilizamos por separado las palabras que están repetidas en ambos.

En un primer momento unimos el léxico básico y el disponible en orden alfabético organizándolo en una tabla de tres columnas (*vid.* Anexo 1). En esta lista están todas las palabras seleccionadas dentro de los límites fijados previamente, las temáticas, las atemáticas, las gramaticales, etc. En la primera columna (en color negro) las extraídas de López Meirama (2008), en la segunda (en verde) las de Tomé Cornejo (2015) y en la tercera (en rojo) las de Davies (2006). De este modo se puede cotejar y observar cuáles se registran en los tres, cuáles en dos y cuáles en uno. Así se evidencian las celdas en las que hay palabras repetidas entre los diferentes trabajos. En el léxico disponible esto se debe a que están repetidas en más de un centro de interés, de tal modo que si

aparecen tres veces, están en tres de ellos, y así sucesivamente. En el léxico básico hay repeticiones, pero esta vez se puede apreciar que hay una palabra en mayúsculas y otra en minúsculas. Las palabras escritas con mayúsculas son las que se encuentran entre las 3.000 más frecuentes, las minúsculas son las que se pertenecen a los conjuntos temáticos que Davies ofrece y que hemos considerado hasta la frecuencia 5.000.

En un segundo momento hemos procedido a fusionar el vocabulario frecuente y el disponible y a intentar organizarlo. Para ello optamos por una estructuración en series, clases de palabras y núcleos temáticos. Así, en el Anexo 2 encuadramos los vocablos en series y clases de palabras y en el Anexo 3 ofrecemos los que hemos organizado temáticamente (*vid.* Anexos 2 y 3). Al igual que el vocabulario ordenado alfabéticamente, en este caso también consideramos las palabras gramaticales y las léxicas, pero hemos extraído en primer lugar las gramaticales, que forman series más o menos cerradas, como decimos. Son las siguientes: artículos determinados e indeterminados, determinantes y pronombres demostrativos, determinantes y pronombres posesivos, pronombres personales, relativos, pronombres interrogativos, determinantes y pronombres indefinidos, indefinidos distributivos, numerales cardinales, numerales ordinales, numerales multiplicativos, numerales partitivos, preposiciones, conjunciones, adverbios, verbos, comparativos y superlativos, adjetivos y nombres. Añadimos, además, la serie *colores*. Estas categorías deben enseñarse todas juntas, aunque se trabajen de forma progresiva, por eso las aislamos y las agrupamos. No rellenamos aquellas clases de palabras que no están completas con los primeros 3.000 términos más frecuentes, pues es evidente cuáles son los que faltan y resulta sencilla su recuperación. La información sobre qué vocablos de estas series son más usuales es relevante para la enseñanza, pues podemos incidir más en unos o en otros, repetir más los frecuentes en los ejemplos y actividades, etc.

Por otro lado consideramos las palabras temáticas o léxicas, es decir, las disponibles y las frecuentes que Davies agrupa en conjuntos en su “thematic vocabulary list”. Las fusionamos y organizamos según los temas que mostramos más abajo, en el apartado siguiente (*vid. infra* § 9.2.). Igual que en la lista

ordenada alfabéticamente, los colores indican de qué obra se han obtenido: negro para las de López Meirama, verde para las de Tomé Cornejo y rojo para las de Davies. En este vocabulario organizado por temas no hemos incluido todavía los verbos, adjetivos, nombres y adverbios más frecuentes entre las 3.000 unidades extraídas de Davies. Estos los ofrecemos en el documento de las series y clases de palabras, organizados en cuatro conjuntos: verbos, adjetivos, nombres y adverbios (Anexo 2). Consideramos que las unidades frecuentes de estas tres clases de palabras deben ser fundidas con las del vocabulario temático, por lo que acometeremos ese trabajo en el futuro. El motivo de que no lo hagamos aquí es que procesar todas estas unidades supone un esfuerzo de enorme calado, que está fuera de nuestros objetivos actuales. De todas formas, el modo de proceder en este caso sería considerar cada palabra por separado y observar en cuál o cuáles de los temas se podría incluir. Por ejemplo, si pensamos en *El cuerpo humano*, dentro del tema número 1. *Individuo: dimensión física, perceptiva y anímica*, consideraremos aquellos verbos, adjetivos y adverbios que se usan más comúnmente en relación con ellas, como *mirar, sentir, oír, hablar, ver, decir*, etc. Si optamos por *El aspecto físico y la personalidad*, otro de los núcleos temáticos de 1. *Individuo*, podemos valorar los verbos *crecer, nacer, morir* y los adjetivos *bajo, fuerte, parecido y serio*. En el caso de *El ciclo de la vida*, los vocablos serían *muerto, viejo y joven*. Algunos de estos ejemplos se podrían asociar a varios núcleos temáticos.

Además, un buen procedimiento para relacionar las palabras frecuentes con las disponibles dentro de los núcleos temáticos definidos es recurrir a las indicaciones sobre la frecuencia de sus coapariciones que los corpus nos muestran. Es decir, el *PCIC* nos indica cuáles son las situaciones comunicativas propias de cada nivel, por lo que debemos pensar qué nombres, verbos, adjetivos, adverbios y otras unidades usamos en cada una de las situaciones y cómo solemos combinarlos en la práctica los hablantes nativos. Así, a continuación podemos observar las muestras y los datos que los corpus ofrecen sobre el uso que alcanzan estas unidades léxicas combinadas. De este modo obtenemos información objetiva para corroborar si estas uniones que deducimos son usuales en la práctica o debemos decantarnos por otras.

Como ya hemos comentado, tanto en el caso de la lista alfabética como en el de la lista estructurada en series y temas, hemos sumado algunas multipalabras: locuciones, expresiones y algunas colocaciones cuya frecuencia de uso hemos obtenido a partir de Davies (2006) o de Ávila Muñoz (1999).

9.2. Propuesta de vocabulario fundamental por orden alfabético (Anexo 1)

Aunque se puede apreciar en el Anexo 1 nuestra propuesta completa de vocabulario fundamental organizado alfabéticamente, exponemos a continuación una muestra que se corresponde con los datos obtenidos en la letra A, pero en este caso sin distinción de colores.

LETRA A

Léxico disponible de López Meirama (2008)	Léxico disponible de Tomé Cornejo (2015)	Léxico frecuente de Davies (2006)
		A Prep to, at 5
		ABAJO Adv down, below, downward 631
		ABANDONADO Adj abandoned 2298
		ABANDONAR V to abandon, leave (a place 638
		ABANDONO Nm abandonment, desertion 2524
		ABARCAR V to encompass, comprise 2163
	abeja	abeja
		ABIERTO Adj open, unlocked 439
abogado-a	abogado/a abogado/a/os abogado/s abogado	abogado ABOGADO Nm/f lawyer 1680
abonar		
		ABORDAR V to board, get in/on 2273
abrigo	abrigo/s	abrigo ABRIGO Nm overcoat, shelter 2996
		ABRIL Nm April 1922
		ABRIR V to open 252
		ABSOLUTAMENTE Adv absolutely, 1181
		ABSOLUTO Adj absolute 574
		ABSORBER V to absorb, soak up 2056
		ABSURDO Adj absurd 2010
	abuelo/a/os/as	abuela ABUELA Nf grandmother 2408 abuelo ABUELO Nm grandfather 1740

		ABUNDANTE Adj abundant, plentiful 2316
		ABURRIDO Adj boring, bored 2813
	aburrimento aburrimento	
		ABUSO Nm abuse, misuse 2605
		ACÁ Adv here, around here 2230
		ACABAR V acabar de V: to have just Ved; 266
		ACADÉMICO Adj academic 2399
	acantilado/s	
		ACASO Adv by chance, maybe 1072
		ACCEDER V to consent, agree, accede 2130
		ACCESO Nm access, entry 1315
	accidente	ACCIDENTE Nm accident, irregularity 1098
	acción/es	ACCIÓN Nf action, act, deed 405
	aceite aceite/s	ACEITE Nm oil 2076
	acelerador	
	acelerar	
		ACENTO Nm accent, stress 2610
		ACEPTACIÓN Nf acceptance, approval 2758
		ACEPTAR V to accept 296
acera		
		ACERCA Adv about, concerning 1288
		ACERCAR V to come near 410
		ACLARAR V to clarify, clear up 1133
	acomodador	
		ACOMODAR V to be/make comfortable 2196
		ACOMPañAR V to accompany 512
	acondicionador	
		ACONSEJAR V to advise, counsel 1828
		ACONTECIMIENTO Nm event, happening 1125
		ACORDAR V to remember, remind 595
		ACOSTUMBRADO Adj accustomed 1595
		ACOSTUMBRAR V to be accustomed to 1628
		ACTITUD Nf attitude 644
		ACTIVIDAD Nf activity, action 393
		ACTIVO Adj active 1328
		ACTO Nm act, action 572
	actor/es actor/es actor actriz/es	actor ACTOR Nm actor 1705 actriz
		ACTUACIÓN Nf performance, acting 1924
		ACTUAL Adj current 676

		ACTUALIDAD Nf present (time) 2046
		ACTUALMENTE Adv now, at present 2362
	actuar	ACTUAR V to act 576
		ACUDIR V to attend, go to, frequent 1044
		ACUERDO Nm accord; de acuerdo: in 313
		ACUMULAR V to accumulate, amass 1301
	acusación/es	
	acusado/a acusado/s	
		ACUSAR V to accuse 1524
		ADAPTAR V to adapt 2087
		ADECUADO Adj adequate, suitable 948
		ADELANTAR V to move forward 1777
		ADELANTE Adv forward, further 516
	adelgazar	
		ADEMÁS Adv also, as well, besides 159
		ADENTRO Adv inside 2084
		ADIÓS Nm goodbye 2557
		ADMINISTRACIÓN Nf administration 2055
		administrador
		ADMINISTRATIVO Adj administrative 2852
		ADMIRABLE Adj admirable 2934
		ADMIRACIÓN Nf admiration 2448
		ADMIRAR V to admire 1682
		ADMITIR V to admit 985
	ADN	
	adolescencia	
	adolescente/s	ADOLESCENTE Nc adolescent 2645
		ADOPTAR V to adopt 1267
		ADQUIRIR V to acquire, get 603
		ADULTO Adj adult, grown-up 2685
		ADULTO Nm adult, grown-up 2200
		ADVERTIR V to notice, warn 887
		AÉREO Adj aerial, air 2975
	aeropuerto	aeropuerto AEROPUERTO Nm airport 2993
		AFÁN Nm eagerness, zeal, effort 2152
		AFECTAR V to affect 991
		AFEECTO Nm affection, fondness 2132
	afila(lápices)	
		AFIRMACIÓN Nf statement, assertion 2357
		AFIRMAR V to assert, affirm 978
	afluente/s	
		AFORTUNADAMENTE Adv luckily,

		fortunately 2712
		AFUERA Adv outside 1858
	agacharse	
		AGARRAR V to grasp, catch 2074
		AGENCIA Nf agency, branch 2886
	agencia inmobiliaria/s	
	agenda/s	
		agente AGENTE Nc agent 1806
	agnóstico/s	
	agobio	
		AGOSTO Nm August 1244
		AGOTADO Adj tired, used up, sold out 2861
		AGOTAR V to exhaust, wear out, run out 1919
	agradable	AGRADABLE Adj nice, pleasant 1724
		AGRADECER V to thank for 1681
		AGREGAR V to add, gather 1708
		AGRESIVO Adj aggressive 2242
agricultor agricultor	agricultor/es agricultor	agricultor
agricultura agricultura	agricultura	
agua	agua agua agua agua agua agua	agua AGUA Nf (el) water 194
		AGUANTAR V to endure, put up with 1895
		AGUDO Adj sharp, acute 2180
águila	águila	
	aguja/s	
	agujero/s negro/s	
		AH interj oh 1489
		AHÍ Adv there 189
		AHORA Adv now 85
		AHORRO Nm saving, thrift 2977
	ahorro/s	
		AIRE Nm air, wind, appearance 374
		AISLADO Adj isolated 1533
ajedrez		
		AJENO Adj not belonging to, detached 1175
		AJUSTAR V to adjust, tighten 1545
	Alá	
		ALA Nf (el) wing, flank 2260
alacena		
albañil	albañil albañil	
	albornoz	
	alcachofa (de la ducha)	
	alcalde	
		ALCANCE Nm reach, scope 1083
		ALCANZAR V to reach, catch up with 329

	alcohol	alcohol ALCOHOL Nm alcohol 2421
	alegre	ALEGRE Adj happy, lively, cheerful 2304
	alegría alegría	ALEGRÍA Nf joy, happiness 1429
		ALEJADO Adj remote, distant, 1894
		ALEJAR V to move away 1196
		ALEMÁN Adj German 963
	alfombra	
		ALGO Pron something, somewhat 101
		ALGUIEN Pron somebody, someone, 480
		ALGUNO Adj some, someone (Pron)50
		alimentaciónALIMENTACIÓN Nf feeding, food 2989
		ALIMENTAR V to feed, support 1609
	alimentos alimentos	alimento ALIMENTO Nm food, nourishment 1143
		ALLÁ Adv there, over there 324
		ALLÍ Adv there, over there 167
		ALMA Nf (el) soul 984
		ALMACÉN Nm warehouse, store 2455
		almuerzo
	alquilar	
	alquiler	
		ALREDEDOR Adv about, around, round 556
		ALREDEDOR Nm surrounding area, vicinity 1617
	altavoz/es altavoz/es	
		ALTERAR V to alter, modify 1809
		ALTERNATIVA Nf alternative 1800
	alto/a	ALTO Adj tall, high 185
	altruismo	
		ALTURA Nf height, altitude 578
	alumno/a/os alumno/s	ALUMNO Nm student, pupil 969
ama-o de casa	ama de casa amo/a de casa	
	amable	
		AMABLE Adj kind, nice, friendly 2409
		AMANTE Nc lover 1893
	amapola/s	
	amar amar	AMAR V to love 1486
		AMARGO Adj bitter, sour, painful 2710
		AMARILLO Adj yellow 1586
		AMBICIÓN Nf ambition 2440
		ambiente AMBIENTE Nm enviroment, atmosphere 683
		ÁMBITO Nm sphere, space 1547
		AMBOS Adj both 488
	ambulancia ambulancia/s	

	ambulatorio	
		AMENAZA Nf threat 2229
		AMENAZAR V to threaten 2239
		AMERICANO Adj American 1261
		AMIGA Nf friend (f) 1817 AMIGO Nm friend 262
	amigo/a/os/as amigo/s amigo/s amigos amigos amigos/as	
	amistad/es amistad amistad	AMISTAD Nf friendship 1232 AMISTAD Nf friendship 1232
		amo AMO Nm master, boss (m) 2787
	amor amor amor amor	AMOR Nm love 423
		AMPLIAR V to enlarge, increase, expand 1931
		AMPLIO Adj wide, ample, broad 763
	análisis	
		ANÁLISIS Nm analysis 1148
		ANALIZAR V to analyze 1048
	anarquía	
		ANCHO Adj wide 1720
		ANCIANO Adj elderly, aged 1912
	anciano/a/os	
	andar andar	ANDAR V to walk, function 536
	Android	
		ANÉCDOTA Nf anecdote 2028
	anemia	
	anestesista/s	
		ÁNGEL Nm angel 2248
		ÁNGULO Nm angle 2389
	angustia angustia	ANGUSTIA Nf anguish, distress 1957
animal	animal/es animal/es animales	ANIMAL Nm animal 497
		ANIMAR V to cheer up 1354
		ÁNIMO Nm zest, spirit, heart 1240
	aniversario	
		ANÓNIMO Adj anonymous 2499
anorak		
	anorexia	
		ANTE Prep before, in the presence of 236
antebrazo	antebrazo/s	
		ANTECEDENTE Nm antecedent 2567
		ANTERIOR Adj previous, preceding 370
		ANTERIORMENTE Adv previously 2719
		ANTES Adv before 114
	antibiótico/s	
	anticiclón	
		ANTICIPAR V to anticipate, advance 2366

		ANTIGUO Adj old, ancient, former 348
	antipático	
		ANUNCIAR V to announce, advertise 730
	anuncio/s anuncio/s (publicitario)	ANUNCIO Nm advertisement, announcement 2114
		AÑADIR V to add 1369
		AÑO Nm year 55
	año/s luz	
		APAGAR V to turn off, extinguish 2104
		APARATO Nm apparatus, device 1168
		APARECER V to appear 221
		APARENTE Adj apparent 2659
		APARENTEMENTE Adv apparently 1853
		APARICIÓN Nf appearance, apparition 1854
		APARIENCIA Nf appearance, aspect 1863
	apartamento apartamento/s	
		APARTAR V to move away, separate 1505
		APARTE Adv apart, aside, separately 1313
	apellido/s apellido/s	APELLIDO Nm last name 2167
		APENAS Adv hardly, barely 587
	aperitivo/s	
		APLICACIÓN Nf application 2263
	aplicaciones	
		APLICADO Adj applied 1827
		APLICAR V to apply 650
		APORTAR V to provide, contribute 1738
		APOYAR V to support, lean on 616
		APOYO Nm support, backing 830
	Apple	
		APRECIAR V to appreciate 1483
	aprender	APRENDER V to learn 422
		APRENDIZAJE Nm learning, apprenticeship 2675
		APRETAR V to squeeze, tighten, compress 2404
		APROBAR V to pass (test), approve 1443
		APROPIADO Adj appropriate 2422
		APROVECHAR V to take advantage of 612
		APROXIMADAMENTE Adv approximately 2256
		APROXIMAR V to come near, closer 2070

		APUNTAR V to point, note, aim 1361
	apuntes	
		AQUEL Adj that (over there) 107 AQUÉL Pron that one over there (m) 1390
		AQUELLO Pron that over there (n) 814
		AQUÍ Adv here 129
		ÁRABE Adj Arab 2433
arado	arado	
araña		araña
arar		
árbol árbol	árbol/es árbol/es árbol/es	árbol ÁRBOL Nm tree 833
arbusto		
	archipiélago/s	
archivador	archivador/es	
		ARCO Nm arch, arc, bow 2412
	arcoíris	
	ardilla	
		ÁREA Nf (el) area, zone 1602
		arena ARENA Nf sand 1635
		ARGENTINO Adj Argentine 1626
		ARGUMENTO Nm argument, plot 1279
	arma/s arma/s	ARMA Nf (el) weapon 832
		ARMADO Adj armed 1639
armario armario armario	armario/s armario/s	
		ARMONÍA Nf harmony 2529
arquitecto-a	arquitecto	arquitecto ARQUITECTO Nm/f architect 2808
	arquitectura	ARQUITECTURA Nf architecture 2516
		ARRANCAR V to start (machine), uproot 1268
		ARRASTRAR V to drag, pull 1134
		ARREGLAR V to fix, arrange 1351
		ARREGLO Nm arrangement 2607
		ARRIBA Adv up, above 544
		ARROJAR V to throw, fling 1846
		arroyo
arroz	arroz	arroz
	arte	ARTE Nm art, skill 567
arteria	arteria/s	
	articulación/es	
	artículo/s	ARTÍCULO Nm article, product, item 819
		ARTIFICIAL Adj artificial 2018
	artista/s	artista ARTISTA Nc artist, performer 928
		ARTÍSTICO Adj artistic 1598

		ASCENDER V to ascend, be promoted 1909
	ascenso/s	
		ASEGURAR V to assure, secure, insure 542
	asesinato asesinato	
	asesino/s	
		asesor
		ASÍ Adv like that 59
	asiento/s asiento/s	ASIENTO Nm seat, chair 2061
	asignatura/s	
		ASISTENCIA Nf attendance, assistance 2751
		ASISTIR V to attend 822
		ASOCIACIÓN Nf association 2123
		ASOCIAR V to associate with 1661
		ASOMAR V to appear, show, stick (out) 2383
		ASPECTO Nm aspect, appearance 385
		ASPIRACIÓN Nf ambition, aspiration 2589
	aspirador/a	
		ASPIRAR V to aspire, want 1552
	asteroide/s	
	astro/s	
	astronauta/s	
		ASUMIR V to assume, take on 1132
		ASUNTO Nm matter, issue, affair 518
		ATACAR V to attack 1393
		ATAQUE Nm attack, raid 1157
		ATAR V to tie (up) 2153
atasco		
	ateísmo	
	atención	ATENCIÓN Nf attention 441
	atender	ATENDER V to serve, attend to 697
		ATENTO Adj attentive, polite 2532
	ateo/s	
ático		
		atmósfera ATMÓSFERA Nf atmosphere, environment 2003
	átomo/s	
		ATRACCIÓN Nf attraction 2768
	atracó/s	
	atractivo/a	ATRACTIVO Adj attractive 2131
		ATRAER V to attract 1293
		ATRÁS Adv back, behind, ago 483
		ATRAVESAR V to cross 937
		ATREVER V to dare 1276
		ATRIBUIR V to attribute 1539
	ATS	
	aula/s	
		AUMENTAR V to increase 633

		AUMENTO Nm increase, rise 1873
		AUN Adv even (though), still 843 AÚN Adv still, yet 282
		AUNQUE Conj although, even though 117
	auscultar	
		AUSENCIA Nf absence 1321
		AUSENTE Adj absent 2562
		AUTÉNTICO Adj authentic 1223
		auto AUTO Nm auto, car 2259
autobús	autobús/es autobús/es autobús/es	autobús
	autoescuela	
		AUTOMÁTICO Adj automatic 2984
		automóvil AUTOMÓVIL Nm automobile 1578
	autónomo/s	AUTÓNOMO Adj autonomous 2761
	autopista	autopista
		autor AUTOR Nm/f writer, author 538
		AUTORIDAD Nf authority 718
	autovía	
	auxiliar/es	
	avance/s	AVANCE Nm advance, progress 1565
		AVANZADO Adj advanced 1856
		AVANZAR V to advance, progress 618
	ave/s	ave AVE Nm bird 2207
avenida	avenida/s	avenida AVENIDA Nf avenue 2876
		AVENTURA Nf adventure 1366
		AVERIGUAR V to investigate, find out 1983
avión	avión avión/es	avión AVIÓN Nm plane, airplane, aircraft 1283
avioneta		
		AVISAR V to inform, announce 2268
		AVISO Nm warning 2611
	avispa	
		AY interj oh no!, oh my! 2337
		AYER Adv yesterday 1112
	ayuda/s	AYUDA Nf help, aid 594
		ayudante AYUDANTE Nc helper, assistant 2901
		AYUDAR V to help 345
ayuntamiento	ayuntamiento ayuntamiento/s	
	azafato/a	
		AZAR Nm random, chance, hazard 2400
	azúcar azúcar/es	azúcar AZÚCAR Nm sugar 2033
		AZUL Adj blue 904

TABLA 24. Muestra del vocabulario fundamental organizado alfabéticamente. Letra A

9.3. Propuesta de vocabulario fundamental del español organizado en series, clases de palabras y temas (Anexos 2 y 3)

Después de obtener la nómina de palabras que a nuestro parecer forma el vocabulario fundamental del español, nos detuvimos en estructurarlo y organizarlo del mejor modo posible para que sea funcional y útil para su consulta. Evidentemente, aunque intentemos ofrecer un resultado lo más completo y acabado posible, nuestro objetivo no es presentar un producto final dedicado a la consulta directa por parte de profesorado o alumnado de ELE, sino ofrecer un vocabulario de referencia para docentes y creadores de materiales. Para estos últimos, la información sobre la frecuencia y la disponibilidad de las unidades léxicas será sin duda más valiosa.

A nuestro modo de ver, una vez que hemos obtenido la lista de palabras que forman el vocabulario fundamental, su fusión debe responder a criterios más prácticos que teóricos. Con esto queremos expresar que, si bien para llegar a esta nómina de palabras hemos operado con métodos estadísticos y matemáticos propios de un trabajo de investigación, para su unión hemos intentado adoptar la perspectiva de un docente de ELE. Nos situamos, por lo tanto, en el ámbito de la enseñan-aprendizaje de una lengua extranjera e intentamos aplicar una visión que surja de nuestra experiencia de la docencia y la creación de materiales. Creemos que la fusión de ambos tipos de vocabulario no entraña demasiados problemas a nivel teórico, por lo que debe responder a criterios más funcionales. Nuestro objetivo, pues, para diseñar la estructura de este listado ha sido responder a la pregunta: ¿qué vocabulario sería útil y práctico si nos situásemos en el papel de quien enseña lengua o diseña materiales?

Para dar una respuesta operativa ante este planteamiento, en un primer momento separamos las palabras gramaticales, que obtenemos gracias a la lista de Davies, y las agrupamos en series (*vid.* Anexo 2). Como hemos explicado más arriba, con esto tenemos información relevante sobre términos que, aunque no son de difícil selección, sí son usados en diferente grado por los hablantes. Así, en estas series incluimos solo las unidades más usuales, con la numeración que indica el puesto que ocupan entre las 3.000 más frecuentes en Davies (2006) para que esta información pueda ser tomada como referencia a la hora de primar

y repetir unas unidades u otras en la práctica docente de los niveles iniciales e intermedios. Podemos observar el ejemplo de la serie de las preposiciones:

Preposiciones

2	de	Prep of, from
5	a	Prep to, at
6	en	Prep in, on
12	por	Prep by, for, through
13	con	Prep with
15	para	Prep for, to, in order to
43	sin	Prep without
48	sobre	Prep on top of, over, about
54	hasta	Prep until, up to, even (Adv)
58	entre	Prep between, among
61	desde	Prep from, since
125	hacia	Prep toward, towards
172	contra	Prep against, opposite
214	bajo	Prep under, underneath
236	ante	Prep before, in the presence of
534	tras	Prep after, behind
257	según	Prep according to
896	mediante	Prep by means of

TABLA 25. Preposiciones del vocabulario fundamental según el orden de frecuencia de Davies (2006)

Además, Davies recoge las locuciones preposicionales: *en cuanto a*, *conforme a* y *con respecto a*. No podemos olvidar que hay locuciones preposicionales muy comunes, pero, como hemos indicado en el epígrafe dedicado a la selección de las multipalabras (*vid. supra* § 8.3.4.3.), hoy en día casi no es posible obtener datos sobre cuáles son las más frecuentes, aunque hemos apuntado modos de conseguirlo. Sí contamos con datos extraídos de la lengua oral en Málaga (Ávila Muñoz 1999), por lo que los sumamos a los resultados obtenidos en el léxico básico de Davies (2006), que, por otra parte, apenas registra locuciones. Las locuciones preposicionales que Ávila Muñoz (1999) indica como usuales y que hemos seleccionado son:

a base de (loc. preposicional)
a costa de (loc. preposicional)
por medio de (loc. preposicional)
a fuerza de (loc. preposicional)
respecto a (loc. preposicional) (Según el DL la forma esperada es

“con respecto a”)
a lo largo de (loc. preposicional)
a partir de (loc. preposicional) (Nota: no se registra como tal en el *DLE*)
a pesar de (loc. preposicional)
en vez de (loc. preposicional)
a través de (loc. preposicional)
al cabo de (loc. preposicional)
en lugar de (loc. preposicional)

Faltarían otras locuciones preposicionales que conocemos, pero de las que no tenemos datos sobre su frecuencia, por lo que quedan fuera de nuestra selección. Son ejemplos como: *debajo de, detrás de, enfrente de, a favor de, en medio de, en contra de, encima de, de acuerdo con, rumbo a, camino de, junto con, por delante de, junto a, antes de, con arreglo a, lejos de, a falta de, etc.* También hay algunas palabras que se usan con valor preposicional, como: *durante, excepto, salvo, incluso, más, menos...* A esto hay que sumar las dos contracciones posibles en español de preposición más artículo: *del* y *al*. Estas contracciones sí habría que considerarlas ya desde un nivel elemental.

Del léxico frecuente también hemos obtenido unidades frecuentes que no son palabras gramaticales, aunque algunas puedan ser de contenido más general o abstracto y no tan ligadas a un tema concreto como las disponibles. Se trata de los verbos, adjetivos, sustantivos y adverbios más comunes listados por Davies (2006) entre las primeras 3.000 palabras, que en un futuro procederemos a incluir y fusionar con las palabras léxicas, siguiendo criterios como el de uso real en las situaciones comunicativas que apunta el *PCIC*. Como hemos explicado en el apartado anterior (*vid. supra* § 9.1.), las incluimos al final del Anexo 2 organizadas en clases de palabras frecuentes.

Como hemos dicho, además de las 3.000 primeras unidades de la lista general de frecuencia de la obra de Davies, también hemos considerado las palabras de lo que este autor denomina “Thematic vocabulary list”. Para esas últimas nos decantamos por valorar tan solo las que estuvieran dentro de las 5.000 primeras. A pesar de que Davies distingue 30 categorías dentro de su “Thematic vocabulary list”, nosotros nos hemos decantado por no considerarlas todas, sino solo las siguientes: cuerpo, adjetivos de emoción, familia, comida,

profesiones, deportes, ropa, medios de transporte, el tiempo, naturaleza y plantas, animales, colores¹. Añadimos, además, locuciones, expresiones o colocaciones de las que hemos obtenido información objetiva sobre su frecuencia de uso, tanto en Davies (2006) como en Ávila Muñoz (1999), pues son datos altamente útiles en este tipo de unidades.

Junto a los verbos, adjetivos y sustantivos (*vid.* final de Anexo 2), un ejemplo de una clase de palabras no cerrada es la de los adverbios. En esta, la información que mostramos es la siguiente:

Adverbios (tiempo, lugar, modo, cantidad, afirmación, negación, duda y terminados en *-mente*)

10	no	Adv no
24	más	Adv. o adj. comparativo more
36	ya	Adv already, still
42	muy	Adv very, really
53	también	Adv also
59	así	Adv like that
70	sí	Adv yes
73	bien	Adv well
76	entonces	Adv so, then
83	tan	Adv such, as, too, so
85	ahora	Adv now
87	después	Adv after
90	siempre	Adv always, forever
98	menos	Adv. o adj. comparativo less, fewer
102	sólo	Adv only, just
114	antes	Adv before
126	cómo	Adv how?
129	aquí	Adv here
132	luego	Adv later, afterwards
146	casi	Adv almost, nearly
148	durante	Adv during, for (time)
151	nunca	Adv never, ever
159	además	Adv also, as well, besides
164	hoy	Adv today, nowadays
167	allí	Adv there, over there
174	dentro	Adv inside
189	ahí	Adv there
211	todavía	Adv still, yet
279	tampoco	Adv neither, nor, either
282	aún	Adv still, yet

¹ Escogemos aquellas que están relacionadas con los centros de interés del léxico disponible o con las nociones específicas del PCIC.

294	incluso	Adv including, even (Adv)
297	quizás	Adv perhaps, maybe
301	mal	Adv badly
324	allá	Adv there, over there
336	solamente	Adv only
337	bueno	Adv well...
347	través	Adv a través: across, over, through
378	junto	Adv together with, next to
391	cerca	Adv close, near
396	pronto	Adv soon, quick
416	realmente	Adv really, actually, in fact
421	dónde	Adv where?
436	encima	Adv above, on top, in addition
451	fuera	Adv out, outside, away
469	lejos	Adv far (away, off)
482	claro	Adv of course, clearly
483	atrás	Adv back, behind, ago
516	adelante	Adv forward, further
544	arriba	Adv up, above
556	alrededor	Adv about, around, round
580	precisamente	Adv precisely
587	apenas	Adv hardly, barely
591	simplemente	Adv simply, just
602	siquiera	Adv even (if)
631	abajo	Adv down, below, downward
662	totalmente	Adv totally, completely
684	especialmente	Adv especially
701	detrás	Adv behind
728	finalmente	Adv finally, at last
788	jamás	Adv never
820	debajo	Adv underneath, below
843	aun	Adv even (though), still
850	perfectamente	Adv perfectly
863	recién	Adv recently, just
866	peor	Adv worse
869	completamente	Adv completely
882	inmediatamente	Adv immediately
885	delante	Adv (in) front (of), ahead
915	justo	Adv just
924	salvo	Adv except (for), but
936	directamente	Adv directly, straight away
971	probablemente	Adv probably
977	exactamente	Adv exactly
994	seguramente	Adv surely, securely
1008	rápidamente	Adv rapidly, quickly
1046	tarde	Adv late
1072	acaso	Adv by chance, maybe
1106	prácticamente	Adv practically
1112	ayer	Adv yesterday
1181	absolutamente	Adv absolutely,
1200	temprano	Adv early

1235	mañana	Adv tomorrow
1288	acerca	Adv about, concerning
1313	aparte	Adv apart, aside, separately
1358	definitivamente	Adv definitely, finally
1365	profundamente	Adv profoundly, deeply
1407	verdaderamente	Adv truly
1409	únicamente	Adv only
1431	igualmente	Adv equally
1454	nuevamente	Adv anew, again
1463	generalmente	Adv generally, usually
1477	claramente	Adv clearly
1485	justamente	Adv just, exactly
1523	naturalmente	Adv naturally
1604	efectivamente	Adv in fact, actually, indeed
1629	fácilmente	Adv easily
1766	evidentemente	Adv evidently, obviously
1769	exclusivamente	Adv exclusively
1803	cuando	Adv when
1805	suficientemente	Adv enough
1853	aparentemente	Adv apparently
1858	afuera	Adv outside
1868	constantemente	Adv constantly
1944	menudo	Adv a menudo: often
1964	personalmente	Adv personally
1975	lentamente	Adv slowly
1980	posiblemente	Adv possibly
2084	adentro	Adv inside
2183	repente	Adv de + repente: suddenly
2212	normalmente	Adv normally
2226	principalmente	Adv principally, mainly
2230	acá	Adv here, around here
2256	aproximadamente	Adv approximately
2282	necesariamente	Adv necessarily
2340	plenamente	Adv fully
2352	sencillamente	Adv simply
2362	actualmente	Adv now, at present
2380	sumamente	Adv extremely, highly
2430	posteriormente	Adv afterwards
2475	últimamente	Adv lately, recently
2785	ciertamente	Adv certainly, indeed
2585	continuamente	Adv continuously
2818	seguida	Adv en seguida: right away, at once
2909	francamente	Adv frankly, openly
2804	libremente	Adv freely
2688	fundamentalmente	Adv fundamentally, basically
2707	enseguida	Adv immediately, at once
2541	enfrente	Adv in front (of)
2712	afortunadamente	Adv luckily, fortunately
2652	particularmente	Adv particularly, personally
2719	anteriormente	Adv previously
2891	previamente	Adv previously

2753	puramente	Adv purely, simply
2749	recientemente	Adv recently
2519	relativamente	Adv relatively
2974	ligeramente	Adv slightly, lightly
2587	estrictamente	Adv strictly

TABLA 26. Adverbios que forman parte del vocabulario fundamental según el orden de frecuencia de Davies (2006)

Davies indica como locuciones adverbiales: *cuanto antes, cuanto más, a través, a menudo, de repente, en seguida, a lo mejor, al igual, a lo largo y al revés.*

Ávila Muñoz registra las siguientes:

en medio (loc. adverbial)
en principio (loc. adverbial)
en realidad (loc. adverbial)
en resumidas cuentas (loc. adverbial coloquial)
en seguida (loc. adverbial, también como “enseguida” según el *DLE*)
de acuerdo (loc. adverbial)
de buenas a primeras (loc. adverbial)
de cuando en cuando (loc. adverbial)
de hecho (loc. adverbial)
a lo mejor (loc. adverbial coloquial)
a lo tonto (loc. adverbial coloquial)
a medias (loc. adverbial)
a menudo (loc. adverbial)
más bien (loc. adverbial)
más de la cuenta (loc. adverbial)
más o menos (loc. adverbial)
hoy en día (loc. adverbial) (Nota: el *DLE* registra esta locución como equivalente de “hoy día”)
poco a poco (loc. adverbial)
por ahora (loc. adverbial)
por cierto (loc. adverbial)
por desgracia (loc. adverbial)
por ejemplo (loc. adverbial)
por fin (loc. adverbial)
por lo menos (loc. adverbial)
por lo tanto (loc. adverbial)
por lo visto (loc. adverbial)
por si acaso (loc. adverbial)
por supuesto (loc. adverbial)
por tanto (loc. adverbial)
sin embargo (loc. adverbial)
sobre todo (loc. adverbial)
a escondidas (loc. adverbial)
grosso modo ((loc. adverbial)

a la vera (loc. adverbial)
tal vez (loc. adverbial, también se registra como “talvez” en el *DLE*)
de tarde en tarde (loc. adverbial)
de vez en cuando (loc. adverbial)
desde luego (loc. adverbial)
en absoluto (loc. adverbial)
en cambio (loc. adverbial)
en condiciones (loc. adverbial)
en contra (loc. adverbial)
en cuanto (loc. adverbial)
en definitiva (loc. adverbial)
en efectivo (loc. adverbial)
en fin (loc. adverbial)
en general (loc. adverbial) (Nota: el *DLE* registra esta locución como equivalente de “por lo general”)
de momento (loc. adverbial) (Nota: el *DLE* registra esta locución como equivalente de “por el momento”)
a rajatabla (loc. adverbial)
a veces (loc. adverbial)
ahora mismo (loc. adverbial)
al fin y al cabo (loc. adverbial)
al menos (loc. adverbial)
al revés (loc. adverbial)
de nuevo (loc. adverbial)
de pronto (loc. adverbial)
de puta madre (loc. adjetiva vulgar)

En lo que se refiere a las series de palabras, un ejemplo es la de los colores. La ofrecemos de este modo:

Colores

250	blanco	Adj white
259	claro	Adj clear
317	negro	Adj black
533	rojo	Adj red
878	verde	Adj green
888	oscuro	Adj dark, obscure
904	azul	Adj blue
2290	gris	Adj gray, gloomy

TABLA 27. Los colores más frecuentes según Davies (2006) que forman parte del vocabulario fundamental

Otras series que se podrían añadir son los días de la semana, los meses, estaciones, etc., aunque las series cerradas no entrañan dificultad en su selección y tratamiento.

Por otro lado, en el Anexo 3, reunimos la lista de palabras léxicas, que contiene tanto todas las que obtuvimos de la disponibilidad léxica como las palabras frecuentes que Davies agrupa por temas (recuérdese, solo las que se encuentran entre las 5.000 más frecuentes). El conjunto total de palabras léxicas del Anexo 3 se muestra ordenado por temas, basados en las nociones específicas del *PCIC*. Ofrecemos una muestra del vocabulario correspondiente al tema número 1 (*El individuo, 1.1. Dimensión física, perceptiva y anímica*, dentro del subtema *El cuerpo humano: partes, acciones, posturas*) en la Tabla 28. La estructura general de temas se expone más abajo en la Tabla 29.

1. Individuo, 1.1. Dimensión física, perceptiva y anímica

1.1.1. El cuerpo humano: partes, acciones, posturas

Partes del cuerpo

López Meirama (2008)	Tomé Cornejo (2015)	Davies (2006)
antebrazo	antebrazo/s	
arteria	arteria/s	
	articulación/es	
		barba
barriga	barriga	
bíceps		
		bigote
boca	boca	boca
brazo	brazo/s	brazo
		cabello
cabeza	cabeza	cabeza
	cadera/s	cadera
cara	cara	cara
ceja	cejas	ceja
cerebro	cerebro	cerebro
		cintura
codo	codo/s	codo
	columna (vertebral)	
corazón	corazón	corazón
	costillas	
		cráneo
cúbito		
cuello	cuello	cuello
culo		

dedo	dedo/s	dedo
diente	dientes	diente
		entrañas
espalda	espalda	espalda
		esqueleto
estómago	estómago	estómago
	extremidades	
fémur		
		garganta
hígado		hígado
hombro	hombro/s	hombro
hueso	hueso/s	hueso
intestino		
labio	labios	labio
lengua	lengua	lengua
mano	mano/s	mano
		mejilla
muñeca	muñeca/s	
músculo	músculo/s	músculo
	muslo/s	muslo
nariz	nariz	nariz
		nervio
ojo	ojo/s	ojo
ombligo	ombligo	
oreja	oreja/s	oreja
páncreas		
pecho	pecho	pecho
pelo	pelo/s	pelo
pene		
peroné		
pestaña	pestañas	
pie	pie/s	pie
piel		
pierna	pierna/s	pierna
pulmón	pulmón/es	pulmón
		puño
radio		
riñón	riñón/es	riñón
rodilla	rodilla/s	rodilla
sangre		sangre
		seno
tibia		
tobillo	tobillo/s	
	torso	
	tripa/s	
tronco	tronco	
uña	uña/s	uña
vena		vena
		vientre

Acciones	Posturas
agacharse	(estar) de pie
andar	(estar) sentado
caminar	(estar) tumbado
comer	
correr	
hablar	
levantarse	
mover(se)	
saltar	
sentar(se)	
tumbar(se)	

TABLA 28. Palabras del vocabulario fundamental dentro del subtema *El cuerpo humano* (1. El individuo, 1.1. Dimensión física, perceptiva y anímica).

9.4. Organización temática del vocabulario fundamental

Como hemos indicado en el apartado sobre la metodología (§ 8.3.2.), para seleccionar las unidades del vocabulario fundamental nos centramos en las que se suelen denominar palabras léxicas, porque son las que entrañan un mayor grado de dificultad. Las palabras gramaticales (artículos, pronombres, preposiciones, conjunciones, etc.) o aquellas que forman parte de series cerradas (como los colores, los días de la semana, los meses, las partes del día, etc.) son mucho más fáciles de determinar, por lo que su dificultad reside sobre todo en conocer las normas de uso y saber combinarlas de modo correcto. En este sentido, es altamente útil estudiar algunas de estas series, como las preposiciones, teniendo en cuenta sus combinaciones en colocaciones, locuciones, etc. Para ello, resultan de utilidad los criterios que hemos aportado en el capítulo anterior sobre la selección de las multipalabras (*vid. supra* § 8.3.2.3.). En todo caso, el conocimiento sobre la frecuencia de uso de términos que pertenecen a series es también necesario, pues tenemos que usar algunos de ellos en ejemplos o muestras de lengua de los niveles iniciales. De este modo, utilizaremos o repetiremos con preferencia aquellas que son más usuales, dejando para los niveles intermedios y avanzados las que obtienen un menor uso en los corpus textuales.

Todo lo anterior afecta también a la organización del vocabulario resultante, ya que las palabras gramaticales se agrupan en conjuntos o series y no entrañan grandes dudas o problemas a la hora de incluirlas en unas u otras. La mayor dificultad la encontraremos, al igual que en el proceso de selección, en la manera de agrupar las palabras léxicas o de contenido. Para este tipo unidades es necesario establecer una taxonomía de temas y luego decir cuáles están en uno u otro, o incluso en varios.

Así pues, para la organización de las palabras léxicas que forman parte del vocabulario fundamental, los temas que hemos considerado guardan relación directa con los que el *PCIC* establece. Con estos hemos relacionado los centros de interés de los léxicos disponibles elegidos (López Meirama 2008 y Tomé Cornejo 2015) así como las agrupaciones temáticas que Davies realizó. Hemos intentado encajar y fusionar toda esta información en la siguiente estructura de temas:

<p>1. Individuo:</p> <ul style="list-style-type: none"> 1.1. Dimensión física, perceptiva y anímica <ul style="list-style-type: none"> 1.1.1. El cuerpo humano: partes, acciones, posturas 1.1.2. El ciclo de la vida: nacimiento, reproducción y muerte 1.1.3. El aspecto físico y la personalidad 1.1.4. Sentimientos y sensaciones 1.1.5. Adjetivos de emoción 1.2. Identidad personal <ul style="list-style-type: none"> 1.2.1. Para identificarse: documentación, datos y objetos personales 1.2.2. Nacionalidades
<p>2. Relaciones personales</p> <ul style="list-style-type: none"> 2.1. Las relaciones personales (familiares y no familiares) 2.2. Actos sociales y celebraciones
<p>3. Alimentación</p> <ul style="list-style-type: none"> 3.1. Dieta y nutrición 3.2. Alimentos 3.3. Bebidas 3.4. Preparar una comida: Recetas 3.5. La cocina y sus utensilios 3.6. En el restaurante 3.7. En el bar 3.8. Objetos sobre la mesa

<p>4. Educación</p> <ul style="list-style-type: none">4.1. El sistema educativo4.2. En clase
<p>5. Trabajo</p> <ul style="list-style-type: none">5.1. El mundo laboral5.2. Profesiones y oficios5.3. En la oficina
<p>6. Ocio</p> <ul style="list-style-type: none">6.1. Actividades de ocio y de tiempo libre6.2. Deportes6.3. Juegos y distracciones
<p>7. Información y medios de comunicación</p> <ul style="list-style-type: none">7.1. La información y la publicidad7.2. Comunicarse por teléfono o carta7.3. Informática y nuevas tecnologías
<p>8. Vivienda</p> <ul style="list-style-type: none">8.1. Buscar una vivienda8.2. Las tareas domésticas8.3. Partes de la casa8.4. Los muebles8.5. Calefacción e iluminación
<p>9. Servicios</p> <ul style="list-style-type: none">9.1. En comisaría9.2. Servicios sociales y solidaridad
<p>10. Compras, tiendas y establecimientos</p> <ul style="list-style-type: none">10.1. En la tienda de ropa: la ropa10.2. Hacer la compra
<p>11. Salud e higiene</p> <ul style="list-style-type: none">11.1. El sistema sanitario11.2. En el médico11.3. En el baño o aseo

<p>12. Viajes, alojamiento y transporte</p> <p>12.1. Ir de viaje</p> <p>12.2. Conducir un coche</p> <p>12.3. Medios de transporte</p>
<p>13. Economía e industria</p> <p>13.1. En el banco</p> <p>13.2. Economía y hacienda</p> <p>13.3. Sectores de producción (primario y secundario)</p>
<p>14. Ciencia y tecnología</p> <p>La ciencia</p>
<p>15. Gobierno, política y sociedad</p> <p>15.1. Organización política y social</p> <p>15.2. En un juicio</p> <p>15.3. El ejército</p>
<p>16. Actividades artísticas</p> <p>16.1. Actividades artísticas</p> <p>16.2. En el cine o en el teatro</p>
<p>17. Religión y filosofía</p> <p>Creencias religiosas y espirituales</p>
<p>18. Geografía y naturaleza</p> <p>18.1. La ciudad</p> <p>18.2. El campo</p> <p>18.3. Los animales</p> <p>18.4. Las plantas</p> <p>18.5. La geografía y los accidentes geográfico</p> <p>18.6. Medioambiente y desastres naturales</p> <p>18.7. El tiempo atmosférico</p> <p>18.8. El universo y los viajes espaciales</p>
<p>19. Acciones/verbos</p> <p>Acciones que realizamos todos los días</p>

TABLA 29. Temas considerados para la organización del vocabulario fundamental temático

9.5. Observaciones sobre los datos de nuestro vocabulario fundamental

En el Capítulo 8 hemos explicado la metodología seguida para obtener el vocabulario fundamental del español (*vid. supra* § 8.3.2.) y en los apartados precedentes de este capítulo hemos expuesto muestras de los resultados, que se pueden consultar completos en los Anexos 1, 2 y 3.

Si observamos dichos resultados de modo global, a nivel cuantitativo el producto es un vocabulario fundamental de más de 4.000 unidades léxicas diferentes y menos de 4.500, lo cual está dentro de los parámetros que habíamos fijado previamente. En realidad, si tenemos en cuenta todas las formas comprendidas en él, la suma es mayor: 5.986. No obstante, como ya habíamos comentado, no contabilizamos por separado las palabras repetidas, ni dentro de un mismo repertorio, ni si se repiten en los tres, aunque sí creemos que tiene interés contemplar las repeticiones en nuestras listas. Al computar las cifras por separado, el número de términos frecuentes resultantes es de 3.434, teniendo en cuenta, a partir de Davies (2006) las 3.000 palabras más frecuentes y las 434 que hemos escogido de su “vocabulario temático”. El número de vocablos disponibles es de 2.449, contabilizando las unidades extraídas de la obra de López Meirama (2008) y de la de Tomé Cornejo (2015). A estos sumamos 103 multipalabras, la mayor parte de ellas locuciones, que hemos incorporado gracias a los pocos datos que indica Davies (2006) y, sobre todo, a los que aporta Ávila Muñoz (1999).

A nivel cualitativo, hay conclusiones que afectan a la globalidad de los listados, algunas de las cuales guardan similitud con las impresiones que apuntamos en el Anexo 4, en el que cotejamos el léxico disponible, el léxico frecuente y las nociones específicas del *PCIC*. Por una parte, el léxico básico obtenido es más que el léxico disponible, si bien aquel incorpora muchas palabras gramaticales y de baja carga semántica, como era de esperar. Por otro lado, el léxico disponible resultante del proyecto de Tomé Cornejo es mucho mayor en número de unidades y más variado. Este es el resultado esperado,

teniendo en cuenta que los centros de interés que distingue esta investigadora son más numerosos que los que maneja López Meirama, quien sigue las pautas del PPHDL.

Si analizamos los datos con más detalle, llegamos a otras observaciones interesantes. Así, aunque son menos los vocablos que contabilizamos de López Meirama, hay muchos de ellos que no se encuentran en el otro léxico disponible, el de Tomé Cornejo, ni en el léxico frecuente de Davies. Veamos, como ejemplo, las primeras letras del alfabeto. En la letra A de nuestro vocabulario organizado alfabéticamente, de un total de 33 palabras que obtuvimos de López Meirama (2008), no encuentran equivalencia en las otras dos obras las 11 siguientes: *abonar, acera, ajedrez, alacena, anorak, araña, arar, arbusto, atasco, ático, avioneta*. En la letra B esto ocurre 15 veces de un total de 39 unidades —aunque una de ellas es una abreviatura de otra palabra que sí aparece (*bus, autobús*)—. Son las siguientes: *balcón, baloncesto, banqueta, bañador, barca, batidora, bíceps, billar, bistec, bodega, bombilla, braga, brisca, buhardilla, bus*. En la letra C, para acabar con la ejemplificación, el resultado de las diferentes es 27 de 94 palabras en total. Las que no encuentran equivalencia son: *cafetera, cafetería, calculadora, caldera, calefactor, calentador, canario, cartabón, cavar, cazo, chaleco, chándal, chimenea, chorizo, cocodrilo, colador, cómoda, compás, congelador, constructor, coqueta, corcho, cosechar, cuarto trastero, cucharilla, cucharón, culo*. A partir de esta muestra que nos sirve de ejemplo, enseguida se aprecia que hay algunos vocablos que deben ser considerados en un nivel básico, como *baloncesto, bañador* o *cafetería*, y muchas más que claramente se encuadran dentro de un nivel umbral.

Por otra parte, es evidente que, debido al mayor número de unidades que hemos extraído, en general, de Tomé Cornejo, si las comparamos con el léxico disponible de López Meirama también obtenemos más palabras que no encuentran equivalencia. Cuando analizamos cuáles de ellas no están tampoco en el léxico frecuente, la cantidad continúa siendo considerable. Volvamos a los listados de ejemplo y centrémonos en la letra A. En este caso, nada menos que 76 vocablos, de un total de 147 de Tomé Cornejo (2015), no están en López Meirama (2008). Son los siguientes: *acantilado/s, acelerador, acelerar,*

acomodador, acondicionador, acusación/es, acusado/a, adelgazar, ADN, adolescencia, afluente/s, agacharse, agencia inmobiliaria/s, agenda/s, agnóstico/s, agobio, aguja/s, agujero/s negro/s, ahorro/s, Alá, albañil, albornoz, alcachofa (de la ducha), alcalde, alfombra, alquilar, alquiler, altavoz/es, altruismo, amable, amapola/s, ambulancia/s, ambulatorio, amigo/a/os/as, análisis, anarquía, anciano/a/os, andar, Android, anemia, anestésista/s, aniversario, anorexia, antebrazo/s, antibiótico/s, anticiclón, antipático, año/s luz, apartamento, aperitivo/s, aplicaciones, Apple, archipiélago/s, arcoíris, ardilla, articulación/es, ascenso/s, asesinato, asesino/s, asignatura/s, asteroide/s, astro/s, astronauta/s, ateísmo, ateo/s, átomo/s, atraco/s, atractivo/a, ATS, aula/s, auscultar, autoescuela, autovía, auxiliar/es, avispa, azafato/a.

Además, algunas de las unidades que obtenemos en los léxicos disponibles que nos han servido de fuente parecen claramente corresponder a un nivel inicial, sin embargo no tienen equivalencia ni en Davies (2006) ni en el otro recuento de disponibilidad. Para comprobar en qué medida se pueden considerar básicas esas unidades léxicas hemos partido de todas las palabras del léxico disponible que aparecen solo en uno de los recuentos —de las letras A, B y C— y las hemos buscado en los niveles A1 y A2 del *PCIC*. Las que sí se registran en este documento son: *ajedrez, araña, baloncesto, bañador, braga, cafetería*, en el caso del léxico disponible de Galicia, y *agenda, Alá, alquilar, amable, ambulancia, amigo, andar, antipático, apartamento, aperitivo, aula, auxiliar, azafato/a*, en el listado de Tomé Cornejo. Con esto se confirma una vez más la complementariedad de los datos obtenidos de López Meirama (2008) y de Tomé Cornejo (2015), incluso en cuanto a unidades de nivel básico. A pesar de que Tomé Cornejo (2015) recoge un número mucho mayor de vocablos que López Meirama (2008), no contempla algunos de los que sin duda deben considerarse básicos y sí aparecen el recuento de López Meirama (2008). Además, es patente el hecho de que bastantes palabras que hemos obtenido de los léxicos disponibles no se encuentran en el *PCIC* (*vid.* Anexo 4).

Por último, la diferencia es mayor si comparamos todo el léxico disponible con el léxico frecuente. En este caso, en la letra A, por ejemplo, las palabras frecuentes que no encuentran equivalencia en los recuentos de disponibilidad contemplados son 238, de un total de 295 que tiene todo el conjunto de las frecuentes. Son estas: *a, abajo, abandonado, abandonar, abandono, abarcar, abierto, abordar, abril, abrir, absolutamente, absoluto, absorber, absurdo, abuso, acá, acabar, académico, acaso, acceder, acceso, acento, aceptación, aceptar, acomodar, acompañar, aconsejar, acontecimiento, acordar, acostumbrado, acostumar, actitud, actividad, activo, acto, actuación, actual, actualidad, actualmente, acudir, acuerdo, acumular, acusar, adaptar, adecuado, adelantar, adelante, además, adentro, adiós, administración, administrador, administrativo, admirable, admiración, admirar, admitir, adoptar, adquirir, adulto, advertir, aéreo, afán, afectar, afecto, afirmación, afirmar, afortunadamente, afuera, agarrar, agencia, agente, agosto, agotado, agotar, agradecer, agregar, agresivo, aguantar, agudo, ah, ahí, ahora, ahorro, ajeno, ajustar, ala, alcance, alcanzar, alejado, alejar, alemán, algo, alguien, alguno, alimentación, alimentar, allá, allí, alma, almacén, almuerzo, alrededor, área, alterar, alternativa, altura, amable, amante, amargo, amarillo, ambición, ambiente, ámbito, ambos, amenaza, amenazar, americano, amiga, amo, ampliar, amplio, análisis, analizar, ancho, anciano, andar, anécdota, ángel, ángulo, animar, ánimo, anónimo, ante, antecedente, anterior, anteriormente, antes, anticipar, antiguo, anunciar, añadir, año, apagar, aparato, aparecer, aparente, aparentemente, aparición, apariencia, apartar, aparte, apenas, aplicación, aplicado, aplicar, aportar, apoyar, apoyo, apreciar, aprendizaje, apretar, aprobar, apropiado, aprovechar, aproximadamente, aproximar, apuntar, aquel, aquello, aquí, árabe, arco, área, arena, argentino, argumento, armado, armonía, arrancar, arrastrar, arreglar, arreglo, arriba, arrojar, arroyo, artificial, artístico, ascender, asegurar, asesor, así, asistencia, asistir, asociación, asociar, asomar, aspecto, aspiración, aspirar, asumir, asunto, atacar, ataque, atar, atento, atmósfera, atracción, atraer, atrás, atravesar, atrever, atribuir, aumentar, aumento, aun, aún, aunque, ausencia, ausente, auténtico, auto, automático, automóvil, autor, autoridad, avanzado, avanzar, aventura, averiguar, avisar, aviso, ay, ayer, ayudante, ayudar, azar, azul. Nótese que, por una parte, muchas de estas unidades corresponden a categorías diferentes del nombre, es*

decir verbos, adjetivos o adverbios, y, por otra, encontramos los vocablos de carácter gramatical, como *a*, *aún*, *aunque* y léxico propio de la interacción, como *adiós*, *ah*, *ay*.

A simple vista, la gran mayoría de las unidades más frecuentes de Davies parecen adecuadas para un nivel inicial y/o intermedio bajo, sin embargo, como vemos, ni conseguimos extraerlas de los listados de léxico disponible, ni, por cierto, están recogidas en el *PCIC* en sus diferentes apartados —algo que ya habíamos comprobado en un cotejo anterior (*vid.* Anexo 4). No se encuentran entre las nociones específicas, es de suponer, porque no están ligadas directamente a ninguno de sus temas, tampoco en las nociones generales ni en el inventario de funciones. Las coincidencias, en este caso, son muy escasas.

Otra cuestión interesante que confirman los datos a los que hemos llegado es que del léxico básico obtenemos familias de palabras usuales, como por ejemplo: *acceder-acceso*, *aceptar-aceptación*, *actividad-activo*, *administración-administrador-administrativo*, *admirable-admiración-admirar*, *afecto-afectar*, *agotado-agotar*, *alcance-alcanza*, etc. Hemos explicado anteriormente en qué consisten estas familias de palabras según Nation (*vid. supra* § 4.1.2.) y qué aportan de positivo a nuestra práctica docente, ya que ayudan a solidificar contenidos y a establecer uniones entre las unidades de nuestro lexicón mental. Pensemos en la *idiomaticidad* (*idiom principle*), de la que hablamos en los capítulos iniciales (*vid. supra* § 2.3.), para sustentar esto. Según el principio de la idiomaticidad, los hablantes operamos con bloques semiconstruidos que combinamos al expresarnos por escrito o en la oralidad. De este modo, entendemos el lenguaje, en gran medida, como un conjunto de segmentos léxicos prefabricados constituidos por diferentes unidades que vamos ensamblando. De este concepción se desprende la imagen de que el vocabulario es como una gran pared, y que para construirla debemos unir y ensamblar sus componentes: cemento, masa y bloques. En conclusión, lo que conseguimos mediante el vocabulario fundamental es alcanzar las diferentes piezas que conforman esta pared, como hemos ido viendo. En general, hemos comprobado que nuestras hipótesis eran ciertas y que con cada tipo de léxico (disponible y frecuente) obtenemos diferentes unidades. La base de la lengua, el cemento y la masa de

nuestra pared, la obtenemos gracias al léxico básico, mientras que los bloques provienen del léxico disponible. Las palabras gramaticales sirven para unir los diferentes componentes y provienen de los corpus de frecuencias. También aquellos elementos más generales y con menor carga semántica.

Otra idea que se desprende de estos análisis es que nuestro vocabulario fundamental constituye un buen complemento del repertorio propuesto en las nociones específicas del *PCIC*, desde el momento en que el *Plan Curricular* ofrece solo muestras de las palabras que se deben considerar en cada noción específica de los diferentes niveles, pero no contempla su vocabulario como un conjunto totalizador. Las diferencias entre nuestro vocabulario fundamental y los listados del *PCIC* deben interpretarse desde esta óptica.

9.6. Futuros trabajos e investigaciones

Llegados a este punto, consideramos que hemos logrado nuestro objetivo principal: determinar una metodología para alcanzar un vocabulario fundamental del español que satisfaga nuestras hipótesis. También hemos ido detectando algunas debilidades que dificultan la obtención de una base más sólida. Por una parte, la línea de investigación de la disponibilidad léxica aún debe perfeccionar sus modos de proceder en cuanto a los centros de interés, como demuestra que alcancemos vocablos tan diversos a partir de los léxicos seleccionados (López Meirama 2008 y Tomé Cornejo 2015). Tampoco está resuelta del todo la sobrerrepresentación de sustantivos en detrimento de otras categorías. Por otra parte, ya hemos comentado también las dificultades que entraña trabajar con léxicos básicos que no tienen en cuenta las colocaciones, por ejemplo, o que presentan algunas imperfecciones metodológicas. Estas cuestiones influyen en nuestro trabajo desde el punto de partida, pero en cierto modo no pertenecen a su núcleo. Dado que hemos utilizado los datos que la lingüística del corpus y la disponibilidad ponen a nuestro alcance, nuestro trabajo será más perfecto cuando dichos datos también lo sean, es decir, cuando consigan, por ejemplo, la lingüística del corpus dar acceso de forma sistemática a

todo tipo de unidades léxicas y la disponibilidad léxica refinar más su concepción de los centros de interés.

Por otra parte, nuestro trabajo no se presenta como un producto final, pues para comprobar si nuestros resultados son realmente útiles, deberían ser trasladados a la práctica docente con el objeto de valorar en qué medida son aplicables en el contexto de la didáctica de ELE y hasta qué punto se observa una mejora en el aprendizaje de nuestro alumnado, cuando se parte de ellos.

Algunos trabajos que sí deben ser encarados en un siguiente paso para progresar en nuestra línea de investigación tienen que ver con la presentación de un léxico más completo, fruto de la fusión del léxico básico y disponible en un vocabulario que consigamos organizar temáticamente y que presente una estructura lo más satisfactoria posible para quien elabora materiales docentes o quien enseña español. En la presente tesis hemos comenzado a transitar por este camino y hemos conseguido ofrecer un vocabulario que distingue series, clases de palabras y temas, y también hemos apuntado modos de proceder para una fusión total y completa de los términos frecuentes que agrupamos en clases de palabras (los sustantivos, verbos, adjetivos y adverbios del Anexo 2) y el resto de unidades organizadas en temas (Anexo 3). Sin embargo, todavía debemos completar esta labor y comprobar si la logramos con éxito.

Asimismo, un reto que nos gustaría afrontar en el futuro es la difusión en línea de nuestro vocabulario fundamental, no solo porque con ello conseguiríamos llegar a un mayor número de usuarios, sino también por las posibilidades que el soporte digital ofrece de cara a su consulta. De esta forma, contar con una herramienta en red de este tipo nos permitiría manejar los resultados con mayor versatilidad y ofrecer datos que no siempre es posible exponer de manera sencilla en papel. Otra información que creemos que es imprescindible incorporar es la clasificación por niveles de cada palabra, de tal modo que podamos saber a qué nivel corresponde cada vocablo al buscarlo o bien explorar sobre las unidades correspondientes a un nivel concreto.

Conclusões

- *A importância do vocabulário no ensino-aprendizagem de línguas*

Nos dois capítulos que dizem respeito ao enquadramento teórico deste trabalho (os dois iniciais), estudámos as questões relativas ao desenvolvimento da competência lexical, os dados que conhecemos sobre o funcionamento do léxico mental e as características do vocabulário. Interligamos todos estes aspetos com o ensino-aprendizagem de uma segunda língua (L2) ou uma língua estrangeira (LE) para que nos servissem de guia nas nossas opções e, sobretudo, porque têm vindo a ganhar importância na didática dos idiomas. Assim, uma das primeiras conclusões à que chegámos foi a necessidade de ressaltar a importância que adquire o vocabulário no ensino de um idioma. Seguimos diferentes estudos e autores que assim o defendem, tanto em espanhol como noutras línguas e, desta forma, fomos apontando algumas ideias ao respeito como as seguintes:

- a) outorgar uma maior centralidade ao vocabulário que à gramática ou à sintaxe,
- b) organizar os programas didáticos em torno dos núcleos temáticos ou situações comunicativas e não baseando-os na gramática,
- c) trabalhar o léxico de forma implícita e explícita,
- d) ensinar todo tipo de palavras (incluídas as expressões multipalavras) e fazer com que se gerem associações entre elas, para ajudar a processar, organizar, compreender e consolidar a informação nova,

- e) procurar estratégias para desenvolver a competência lexical dentro e fora da sala de aula, também tendo em conta a aprendizagem autónoma.

A teoria que propõe a existência de um limiar lexical é um trabalho que confirma a importância imprescindível do vocabulário na progressão da aprendizagem de um idioma para que se produza compreensão leitora e auditiva adequada ou ótima. Esta hipótese sustenta-se em estudos que analisaram o grau de sucesso na compreensão de alunado de inglês como língua estrangeira em função da percentagem de cobertura lexical textual do seu conhecimento lexical. A cobertura lexical é a percentagem de palavras que compõem um texto (oral ou escrito) que faz parte do conhecimento do aprendiz e o limiar lexical é o nível mínimo de conhecimento lexical que os aprendizes de uma L2 devem de ter para atingir uma compreensão efetiva. Neste sentido, as percentagens que se equacionam são 95% de cobertura para que o nível de compreensão seja aceitável e 98% para que seja ótimo. Traduzida esta percentagem em número de palavras, na oralidade precisamos de aproximadamente 3.000 para uma cobertura lexical de 95% e 5.000 palavras para uma de 98%. Na escrita, estes números aumentam para 5.000 palavras, no que se refere a uma cobertura lexical de 95%, e entre 8.000 e 9.000 para uma cobertura de 98%.

A compreensão lexical efetiva tem-se associado à habilidade do aluno para inferir ou adivinhar o significado de termos desconhecidos num texto, de tal modo que se não conhece um número de palavras suficiente, pode que esta inferência não se produza.

- *É necessário levar a cabo uma seleção léxica*

Outra das conclusões às quais chegamos desde uma fase inicial, foi a necessidade de levar a cabo um processo de seleção lexical, a qual já foi apontada pelo *Quadro Europeu Comum de Referência para as Línguas (QEER)*. É um imperativo que surge das características do vocabulário e da constatação de que existem diferentes tipos de unidades lexicais, às quais devemos recorrer em função dos níveis e contextos educativos. Quanto às características do vocabulário, podemos afirmar que a sua alta complexidade deriva do seu

carácter aberto, variável e ilimitado, isto é, ao falar do vocabulário de uma língua estamos a tratar de um número de unidades quase infinito e nas quais se manifesta a variação (diatópica, diastrática e diafásica) de forma mais evidente, além de estas estarem sujeitas a um grande dinamismo ao longo dos anos (diacronia). Por outro lado, há estudos suficientes que criticam os materiais de espanhol para estrangeiros precisamente por ter sido feita uma escolha dos termos pouco adequada ao nível educativo ou sem critérios explícitos. Os resultados destas análises demonstram que devemos começar a tratar esta questão com mais rigor e recorrer a métodos objetivos.

Para levar a cabo a seleção lexical, devemos responder duas questões básicas: quanto vocabulário ensinar em cada nível e que unidades lexicais escolher.

- *Quantas palavras devemos seleccionar?*

Para determinar quantas unidades lexicais é recomendável introduzir em cada nível do *QECR*, optamos por recorrer aos estudos sobre compreensão lexical e às percentagens de cobertura lexical textual, que se realizaram, sobretudo, com estudantes de inglês como língua estrangeira e que tomaram o léxico frequente como ferramenta de trabalho. Graças a estas contribuições, elaborámos uma tabela na qual definimos o número de palavras frequentes que consideramos oportuno oferecer em cada nível. Isto serviu-nos de referência para estabelecer um critério objetivo, com o qual delimitamos quantos termos extraímos do léxico frequente, para somar ao disponível e formar o vocabulário fundamental. Como referimos no Capítulo 8, no qual expomos uma definição deste tipo de vocabulário e estabelecemos o respetivo nível do *QECR*, o intervalo de palavras que incluir no nível B1, que pretende ser a nossa meta, deve-se situar entre as 3.000 e as 5.000, pois são as que possibilitam uma percentagem de cobertura lexical textual entre 95% e 98% e, portanto, um nível de compreensão textual aceitável.

Assim sendo, consideramos 3.103 palavras frequentes, paralelamente a 2.449 disponíveis, que, ao serem fundidas, derivam em pouco menos de 4.500 unidades diferentes, uma vez descartadas as repetições.

- *Que palavras selecionar e com que critérios?*

Porém, definir o número de unidades que deve conformar o nosso vocabulário fundamental não é suficiente, pois também é preciso estabelecer quais vamos selecionar e com que critérios. Para atingir soluções estudamos os possíveis critérios de seleção, os diferentes métodos, os diversos tipos de unidades léxicas, que unidades oferece cada um destes métodos e quais são os seus pontos fortes e débeis.

Sobre os critérios, procurámos as indicações dos documentos oficiais como o *QEER* ou o *Plan Curricular del Instituto Cervantes (PCIC)*, embora sejam demasiado gerais e incompletos para satisfazer as nossas necessidades. Por este motivo, recorreremos a analisar as propostas de diferentes autores, com o que chegamos a que todos os critérios esgrimidos (a utilidade, a eficácia, a rentabilidade, a produtividade e a facilidade na aprendizagem) cabem dentro de um mesmo “guarda-chuva”: o da utilidade. Segundo os estudos, as palavras mais úteis são as frequentes e usuais, as disponíveis, as que se consideram no vocabulário instrumental e de sala de aulas, as mais familiares e regulares, as mais necessárias e fáceis de aprender, etc. Relacionado com a frequência encontra-se a cobertura textual e o índice de uso, e no caso das disponíveis devemos atender ao seu índice de disponibilidade. Todos os critérios que encontramos se podem classificar em: a) critérios dependentes da situação de ensino (ou critérios pedagógicos), b) baseados na dificuldade das palavras e c) relacionados com a descrição da língua e com o seu uso.

Por outro lado, observamos que métodos de seleção lexical são possíveis. Neste caso, pudemos comprovar que, apesar de existirem três modos de operar (1. limitação subjetiva, 2. limitação estritamente objetiva e 3. limitação objetiva corrigida), só os procedimentos objetivos ou objetivos corrigidos são os que convém pôr em prática, já que os subjetivos levam a resultados insatisfactorios e aleatórios.

Ao mesmo tempo sabemos que existe uma grande variedade de unidades léxicas, que podemos classificar em palavras (simples e compostas) e unidades pluriverbais lexicalizadas e habitualizadas. Também podemos distinguir entre

palavras conteúdo (“léxicas” ou “plenas”) e palavras função (“gramaticais” ou “vazias”) ou inclusive podemos elaborar uma lista de unidades que deveríamos incluir num vocabulário essencial como: a) modalizadores discursivos, b) verbos de regime, c) palavras interactivas, d) marcadores do discurso, e) deícticos ou termos relativos à dêixis pessoal, espacial e temporária, f) adjetivos básicos, g) advérbios básicos e h) verbos básicos. A isto ainda devemos somar outras catalogações, que são as que se ocupam de discernir os diferentes tipos de vocabulário existentes, como o técnico, o subtécnico (ou semitécnico), o temático (ou disponível) e o básico (ou frequente).

Também as nossas investigações e comparações entre o léxico frequente, o léxico disponível e as noções do *PCIC* apontaram que existem notáveis divergências entre eles.

Finalmente, tentamos pôr em relação estes três tipos de léxico para observar que unidades se alcançavam com cada método de seleção lexical, seguindo os diferentes critérios que tínhamos observado. A conclusão, neste caso, é que devemos aspirar a métodos de seleção lexical objetivos e combiná-los, com a finalidade de atingir um vocabulário plural e inclusivo quanto aos seus componentes. Com isso combinamos as suas vantagens e conseguimos colmatar as suas lacunas. Também podemos usar estes métodos objetivos em separado e ajustá-los às nossas necessidades, se estas são concretas e não gerais.

Depois de avaliar todas as possibilidades, decidimos tomar como objeto o desenvolvimento de um *vocabulário fundamental* derivado de métodos léxico-métricos (objetivos, portanto), que combine diferentes tipos de vocabulário adequados aos níveis iniciais (nomeadamente, o frequente e o disponível), que represente melhor o nosso léxico mental e que promova a nossa competência lexical de um modo mais completo. Com o referido vocabulário fundamental obteríamos o núcleo do idioma, o pilar básico que sustenta a língua que devemos transferir para a sala de aula de espanhol como língua estrangeira, e que posteriormente podemos completar seguindo outras vias. Falamos de implementar este vocabulário com o instrumental, com o visual, incluir também mais multipalabras (colocações), marcadores do discurso, recomendações que

vêm dadas pelo contexto educativo, outras indicações que derivam da proximidade ou contraste entre a língua objeto de estudo e a língua materna dos formandos, etc.

- *O vocabulário fundamental como método de seleção léxica objetiva e eficaz*

Pelo dito anteriormente, podemos concluir que o vocabulário fundamental que pretendemos elaborar é a via mais adequada que encontramos por cumprir com vários requisitos: trata-se de um método de seleção objetiva, baseado em procedimentos léxico-métricos e empíricos, que oferece os diferentes tipos de unidades que se complementam e, além disso, corrige as deficiências da disponibilidade lexical e da frequência, ao mesmo tempo que tira partido das suas vantagens.

Assim sendo, apostámos em ir por este caminho e em encontrar os melhores procedimentos para desenvolver este método. Para isso tivemos que procurar respostas a três questões: que critérios usamos para escolher as unidades lexicais que formam o vocabulário fundamental, como as unimos e como as estruturamos para a sua apresentação final. Conforme foi avançando o trabalho, percebemos que o objetivo era demasiado ambicioso e que esta tese ia ter que se centrar eminentemente na primeira das anteriores questões, isto é, a de encontrar uma forma objetiva e satisfatória de selecionar os termos disponíveis e as palavras mais frequentes.

Por último, optamos por aplicar estes critérios de seleção a três obras concretas; de uma obtivemos o léxico básico (Davies 2006) e das outras duas extraímos o léxico disponível (López Meirama 2008 e Tomé Cornejo 2015). Relativamente ao léxico básico optamos pelas mais frequentes: as primeiras 3.000 e algumas pertencentes a conjuntos temáticos elaborados pelo seu autor dentro das primeiras 5.000. Para o léxico disponível estudamos todos os centros de interesse de ambas obras de forma individual, embora tenhamos aplicado os mesmos critérios. Neste último caso, o resultado foi um conjunto de unidades algo menor que no primeiro.

- *Que é o vocabulário fundamental, como se chega a ele e como se organiza*

Uma vez centrados no vocabulário fundamental, analisamos o conceito para oferecer uma definição, estudamos a sua história, as suas origens, as diferentes contribuições que se tinham feito tanto em espanhol como noutros idiomas de léxicos fundamentais e, por último, expusemos as nossas propostas sobre o tema.

No que diz respeito à sua definição, observámos o vocabulário fundamental a partir de duas perspetivas, que denominamos analítica e funcional. Assim, se relativamente à primeira entendemos a noção como um método de seleção lexical, no que diz respeito à segunda concentramo-nos no seu produto e no que este nos permite fazer.

Baseámos a seleção em critérios matemáticos, fundamentalmente estatísticos, que garantissem a sua objetividade. No caso do léxico básico, estes sustentam-se na cobertura textual das palavras frequentes e no limiar léxico de compreensão textual. Para o léxico disponível, propomos estudar os resultados da cada centro de interesse como um conjunto de unidades. Deste modo observamos o comportamento e evolução do índice de disponibilidade das palavras, mas, sobretudo, o número de termos que se obtêm se dividimos o centro de interesse em grupos. Medimos a sua linha de tendência com gráficos, mais precisamente com histogramas de frequência, que nos deixaram ver os dados de cada centro de interesse como se de uma radiografia se tratasse. Assim, pudemos distinguir em que parte do centro de interesse, vendo-o como um conjunto, se agrupa o maior número de termos, em que ponto se obtêm menos, qual é a sua evolução e onde se estabelecem cortes na sua progressão. Estes “cortes naturais” serviram-nos de limites que equiparámos aos diferentes níveis do *QECR*. Pela nossa parte, optámos por definir o corte definitivo no nível B1, pois, como explicámos na definição do vocabulário fundamental, achamos que este deve estar formado por aquele léxico necessário para que um falante possa desenvolver-se com eficácia em situações comunicativas quotidianas e tratando temas habituais. O nível B1 é o ponto de inflexão a partir do qual os utentes duma língua têm esta capacidade, pelo que o nosso horizonte deve ser este.

Contudo, interessa também estabelecer outras balizas na progressão do ensino-aprendizagem de uma L2, para as quais podemos usar igualmente o processo de seleção que nos oferece o vocabulário fundamental, só que com outros objetivos linguísticos mais elementares.

Embora não o tenhamos posto em prática, sim oferecemos outra solução para sequenciar as palavras disponíveis de modo mais fluído naqueles casos nos quais se observa, mais que um corte, uma progressão. Para isso, propomos o uso de modelos matemáticos da lógica difusa como alternativa. Com ela conseguimos fugir dos limites rígidos e adaptar-nos a cada centro de interesse e às nossas necessidades específicas com procedimentos também objetivos e empíricos.

Por último, reflectimos sobre a fusão das unidades procedentes de cada fonte e a estruturação do vocabulário fundamental. Por uma parte optamos por uma disposição em ordem alfabética, com a qual podemos comparar as unidades que se repetem, numa e noutra fonte, e quais são as divergências entre elas. Isto permitiu-nos observar que os dois dicionários de disponibilidade lexical que utilizamos oferecem, não só uma quantidade muito díspar de unidades devido ao diferente número de centros de interesse com os que trabalharam, mas também um grau tão considerável de semelhanças como de dissimilitudes ao nível qualitativo. Noutro documento tentamos organizar o vocabulário fundamental pensando numa disposição que fosse útil e prática para quem elabora materiais docentes e para quem dá aulas de espanhol como língua estrangeira, de tal modo que optamos por uma estruturação em séries, classes de palavras e núcleos temáticos.

As séries de palavras costumam estar formadas pelas palavras gramaticais, que fazem parte de grupos mais ou menos limitados, como por exemplo os artigos, as conjunções, as preposições, etc., mas também outros conjuntos de palavras que se costumam estudar juntas, como as cores.

Agrupámos os nomes, verbos, adjetivos e advérbios mais frequentes tendo em conta a classe de palavra à que pertencem. Além disso, alargamos esta lista com as locuções, expressões e colocações das quais pudemos obter informação sobre a sua frequência de uso.

Em separado, no Anexo 3, oferecemos os termos que organizámos em núcleos temáticos. Neles concentramos o léxico disponível e as palavras mais frequentes que na obra de Davies (2006) aparecem organizadas em conjuntos. Ao estruturar este vocabulário temático tomámos como referente o índice de noções específicas do *PCIC*, para facilitar uma posterior comparação com este documento.

- *Que unidades compõem o vocabulário fundamental*

Como sabemos, a disponibilidade lexical nasceu como medida corretora das listas de palavras frequentes. De facto, há numerosos estudos que confirmam que o tipo de unidades às que chegamos por um método e por outro são diferentes, e que o vocabulário disponível nasceu para corrigir as deficiências da frequência. Em consequência, sabemos que os *corpora* linguísticos são capazes de nos oferecer palavras gramaticais e vazias ou outras que, sem serem gramaticais, são pouco específicas (nomes e adjetivos de significado muito geral como *coisa, casa, tempo, local, pessoa, sítio*) e com débil valor informativo. Não obstante, também reconhecemos que não estabelecem uma sequenciação precisa do vocabulário para a docência e, sobretudo, que existem palavras pouco frequentes, mas muito úteis a nível didático. Pelo contrário, estas palavras pouco frequentes, mas muito úteis em determinadas situações comunicativas (sobretudo substantivos temáticos e palavras com alto valor informativo), são justamente aquelas que encontramos nos dicionários de disponibilidade.

Com a nossa investigação também confirmamos estes dados. No entanto, fomos além deles e dedicámos parte dos nossos esforços a estudar mais pormenorizadamente as palavras que encontramos nos léxicos básicos e nos disponíveis, e quais são as similitudes entre os dicionários de disponibilidade que seguem o *Proyecto Panhispánico de Disponibilidad Léxica* e a nova proposta de Tomé Cornejo (2015). Para isso, realizamos diferentes comparações, que mostrámos no Anexo 4, e analisamos os resultados do nosso vocabulário fundamental no Capítulo 9. Além das dissimilitudes entre as duas listagens de léxico disponível e entre estes dois e o léxico básico, observámos também as

divergências entre todos estes repertórios e as unidades que regista o *PCIC* nas suas noções específicas. De tudo isto concluímos que o *PCIC* serve como guia no processo de seleção léxica, mas as suas contribuições são insuficientes e não nascem de critérios e processos objetivos, senão da acumulação de saberes, da experiência docente e do conteúdo dos manuais de espanhol com os que se veio trabalhando. Por este motivo, o vocabulário fundamental que pretendemos elaborar nasce não só como uma obra complementar desse projeto, mas também como um convite a refletir sobre a adequação do inventário de palavras propostas pelo *PCIC* e sobre a sua sequenciação, já que o processo seguido para chegar ao nosso vocabulário procurou a maior objetividade e empirismo possíveis. Da observação das diferenças entre o léxico disponível de López Meirama e o de Tomé Cornejo (2015) surgiu a decisão de considerar ambos para obter o nosso vocabulário fundamental. Por outra lado, comprovámos que do léxico básico é possível extrair famílias de palavras, que podem ser aproveitadas no plano educativo, enquanto que a metodologia da disponibilidade lexical é suscetível de se converter em fonte de vocabulários técnicos, subtécnicos e académicos, que se podem transladar a cursos de espanhol para fins específicos. Em suma, não temos dúvidas ao afirmar que, de modo geral, ambas metodologias usadas em conjunto são bem mais satisfatórias que por separado.

Referencias bibliográficas*

- Aguirre, E. & Edmonds, P. (eds.) (2006). *Word Sense Disambiguation: Algorithms and Applications*. Dordrecht: Springer.
- Ahumada Lara, I. (2006). *El léxico disponible de los estudiantes de Jaén*. Jaén: Servicio de Publicaciones de la Universidad de Jaén.
- Ainciburu, M. C. (2008). La enseñanza del léxico en las lenguas extranjeras. ¿Qué palabras hay que enseñar? En *III Jornadas de Español como Lengua Extranjera, I Congreso Internacional de Enseñanza e Investigación en ELSE, 21-23 de mayo de 2008*. Argentina: Universidad Nacional de Córdoba.
- Ainciburu, M. C. (2014). El léxico en L2: la adquisición del vocabulario y sus mitos ¿Existe el vocabulario rentable? *SIGNOS ELE*, 8, abril.
- Aitchison, J. (1987). *Words in the Mind: An Introduction to the Mental Lexicon*. Oxford: Basil Blackwell.
- Alameda, J. R. & Cuetos, F. (1995). *Diccionario de Frecuencias de las Unidades Lingüísticas del Castellano*. Oviedo: Universidad de Oviedo.
- Alba Ovalle, O. (1995a). *Léxico disponible de la República Dominicana*. Santiago de los caballeros: Pontificia Universidad Católica Madre y Maestra.
- Alba Ovalle, O. (1995b). Disponibilidad léxica en el español dominicano: análisis cuantitativo. En L. Rodríguez-Fonseca & I. Vázquez (Eds.), *Actas del III Seminario Internacional sobre 'Aportes de la lingüística a la enseñanza de la lengua materna (Revista de Adquisición de la Lengua Española)* (pp. 353-365).
- Alba Ovalle, O. (1996). Disponibilidad léxica en el español dominicano: aspectos sociolingüísticos. En Arjona Iglesias *et al.* (Eds.), *Actas del X Congreso Internacional de ALFAL* (pp. 742-749). México: UNAM.
- Alba Ovalle, O. (1998). Variable léxica y dialectología hispánica. *La Torre*, 7-8, 299-316. Puerto Rico: Universidad de Puerto Rico.
- Alba Ovalle, O. (1999). Densidad de anglicismos en el léxico disponible de la República Dominicana. En J. A. Samper Padilla *et al.* (Eds.), *Actas del XI Congreso Internacional de ALFAL, II*, (pp. 853-865). Las Palmas de Gran Canaria, Universidad de las Palmas de Gran Canaria.
- Alcaraz Varó, E. (2002). La tercera didáctica de las lenguas modernas. Las lenguas de especialidad. En Ministerio de Educación Ciencia y Deporte (Ed.), *La lengua, vehículo cultural multidisciplinar* (pp. 95-120). Madrid: Secretaría General Técnica, MECD.

* Todas las direcciones de internet referenciadas en este apartado, así como las incluidas a lo largo de los capítulos de la tesis han sido visitadas en la primera semana del mes de junio de 2017.

- Alderson, J. C. (1984). Reading in a foreign language: A reading problem or a language problem? En J. C. Alderson & A. H. Urquhart (Eds.), *Reading in a foreign language* (pp. 1-27). Londres: Longman.
- Alexander, L. G. (1967). *First Things First: An Integrated Course for Beginners*. Harlow, England: Longman.
- Allen, F. A. (1983). *Techniques in teaching vocabulary*. New York: Oxford University Press.
- Almela, R. et al. (2005). *Frecuencias del español. Diccionario y estudios léxicos y morfológicos*. Madrid: Universitas.
- Alonso Ramos, M. (1994). Hacia una definición del concepto de colocación: De J. R. Firth & I. A. Mel'čuk". *Revista de Lexicografía*, 1, 9-28.
- Alonso Ramos, M. (2004). *Diccionario de colocaciones del español*. Disponible en: <http://www.dicesp.com>.
- Alonso Ramos, M. (2008). Papel de los diccionarios de colocaciones en la enseñanza de español como L2. En J. A. DeCesaris & E. Bernal (Eds.), *Proceedings of the XIII EURALEX International Congress (Barcelona, 15-19 July 2008)*, (pp. 1215-1230). Barcelona: IULA / Universitat Pompeu Fabra.
- Alonso Ramos, M. (2009). Hacia un nuevo recurso léxico: ¿fusión entre corpus y diccionario? En P. Cantos Gómez, (Ed. lit.) & A. Sánchez Pérez (Ed. lit.), *A survey of corpus-based research [Recurso electrónico]* (pp. 1191-1207). Murcia: Asociación Española de Lingüística del Corpus.
- Alonso Ramos, M. (2012). Explorando la frecuencia léxica para el Diccionario de colocaciones del español. En T. Jiménez Juliá, B. López Meirama, V. Vázquez Rozas & A. Veiga Rodríguez (Eds.), *Cum corde et in nova grammatica. Estudios ofrecidos a Guillermo Rojo* (pp. 19-40). Santiago de Compostela: Universidade de Santiago de Compostela, Servizo de Publicacións e Intercambio Científico.
- Alonso Ramos, M. (2013). Colocaciones, diccionario y corpus de aprendices. En M. Casas Gómez (Dir.), R. Vela Sánchez (Ed.), *Eugenio Coseriu, in memoriam. XIV Jornadas de Lingüística* (pp. 57-71). Cádiz: Servicio de publicaciones, Universidad de Cádiz.
- Alvar Ezquerro, M. (2003). *La enseñanza del léxico y el uso del diccionario*. Madrid: Arco Libros.
- Alvar Ezquerro, M. (2005). La frecuencia léxica y su utilidad en la enseñanza del español como lengua extranjera. En M. A. Castillo Carballo (Coord.), *Las gramáticas y los diccionarios en la enseñanza del español como segunda lengua, deseo y realidad. Actas del XV Congreso Internacional de ASELE, Sevilla, 2004*. (pp. 19-39). Sevilla: ASELE.
- Alvar Ezquerro, M. & Villena Ponsoda, J. A. (Coords.) (1994). *Estudios para un corpus del español*. Málaga: Universidad de Málaga.
- Álvarez de la Granja, M. (2011). La presencia del gallego en el léxico disponible del español de Galicia. Análisis formal y funcional. En B. López Meirama (Ed.) *Estudios sobre disponibilidad léxica en el español de Galicia* (pp. 17-102). Santiago de Compostela: Servizo de publicacións e intercambio científico U. de Santiago de Compostela.
- Álvarez de la Granja, M. & López Meirama, B (2013). A presenza do galego no léxico dispoñible do español de Galicia. Análise distribucional. En E. Gugenberger, H. Monteagudo & G. Rei-Doval (Eds.), *Contacto de linguas*,

- hibrididade, cambio: contextos, procesos e consecuencias* (pp. 49-96). Santiago de Compostela: Consello da Cultura Galega.
- American Council on the Teaching of Foreign Languages (ACTFL), (2012). *Performance Descriptors for Language Learners*. Alexandria: The American Council on the Teaching of Foreign Languages.
- Anderson, R. & Freebody, P. (1981). Vocabulary knowledge. En J. T. Gurthie (Ed.), *Comprehension and teaching: Research reviews* (pp. 77-117). Newark, DE: International Reading Association.
- Anderson, R. C. Y Freebody, P. (1981). Vocabulary knowledge. En J. T. Guthrie (Ed.), *Comprehension and Teaching Research Reviews* (pp. 77-117). Newark, DE: International Reading Association.
- Anglin, T. (1970). *The Growth of Word Meaning*. Cambridge Mass: MIT Press.
- Areta Lara, M. (2009). *El léxico disponible de los estudiantes preuniversitarios navarros. Estudio sociolingüístico*. Tesis doctoral. Navarra: Universidad de Navarra.
- Aristizabal, M. (1938). *Détermination expérimentale du vocabulaire écrit pour servir à l'enseignement de l'orthographe à l'école primaire*. Louvain.
- Arnal, M. L. (Coord.), Castañer, R., Enguita, J. M^a., Lagüéns, V. & Moliné, A. B. (2004). *Léxico disponible de Aragón*. Zaragoza: Libros Pórtico.
- Arnal, M. L. (Ed.) (2008). *Estudios sobre disponibilidad léxica en los jóvenes aragoneses. Actas de las Jornadas sobre disponibilidad léxica en los jóvenes aragoneses (Zaragoza, 2005)*. Zaragoza: Institución Fernando El Católico.
- Ávila Martín, M^a C. (2010). Estadística y lingüística de corpus: implicaciones pedagógicas en la enseñanza y el aprendizaje del léxico. *Cauce. Revista internacional de Filología, Comunicación y sus Didácticas*, 33, 163-175.
- Ávila Muñoz, A. M. (1999). *Léxico de frecuencia del español hablado en la ciudad de Málaga*. Málaga: Servicio de Publicaciones de la Universidad de Málaga.
- Ávila Muñoz, A. M. (2006). *Léxico disponible de los estudiantes preuniversitarios de Málaga*. Málaga: Servicio de Publicaciones de la Universidad de Málaga.
- Ávila Muñoz, A. M. (2016). El léxico disponible y la enseñanza del español. Propuesta de selección léxica basada en la teoría de conjuntos difusos. *Journal of Spanish Language Teaching*, Volume 3, 2016, Issue 1, 31-43.
- Ávila Muñoz, A. M. & Sánchez Sáez, J. M. (2010). La disponibilidad léxica. Antecedentes y fundamentos. En A. M. Ávila Muñoz & J. A. Villena Ponsoda (Eds.), *Variación social del léxico disponible en la ciudad de Málaga. Diccionario y análisis* (pp. 38-81). Málaga: Sarriá.
- Ávila Muñoz, A. M. & Sánchez Sáez, J. M. (2011). La posición de los vocablos en el cálculo del índice de disponibilidad léxica: procesos de reentrada en las listas del léxico disponible de la ciudad de Málaga. *ELUA. Estudios de Lingüística*, 25, 45-74.
- Ávila Muñoz, A. M. & Sánchez Sáez, J. M. (2014). Fuzzy sets and prototype theory. Representational model of cognitive community structures based on lexical availability trials. *Review of Cognitive Linguistics*, 12, 133-159.
- Ávila Muñoz, A. M. & Villena Ponsoda, J. A. (Eds.) (2010). *Variación social del léxico disponible en la ciudad de Málaga. Diccionario y análisis*. Málaga: Sarriá.
- Ayora Esteban, M^a C. (2006). *Disponibilidad léxica en Ceuta: aspectos sociolingüísticos*. Cádiz: Universidad de Cádiz, Servicio de Publicaciones.

- Azurmendi, M^a J. (1983). *Elaboración de un modelo para la descripción sociolingüística del bilingüismo y su aplicación parcial a la Comarca de San Sebastián*. San Sebastián: Caja de Ahorros Provincial de Guipúzcoa.
- Bacelar do Nascimento, M. F.; Garcia Marques, M. L. & Segura da Cruz, M. L. (1987). *Português Fundamental, Métodos e Documentos, tomo 1, Inquérito de Frequência*. Lisboa: INIC, CLUL.
- Bacelar do Nascimento, M. F.; Rivenc, P. & Segura da Cruz, M. L. (1987). *Português Fundamental, Métodos e Documentos, tomo 2, Inquérito de Disponibilidade*. Lisboa: INIC, CLUL.
- Bahns, J. (1993). Lexical collocations: A contrastive view. *ELT Journal*, 47(1), 56-63. Oxford: Oxford University Press.
- Bailey Victory, J. (1971). *A study of lexical availability among monolingual-bilingual speakers of Spanish and English*. Tesina inédita. Houston: Rice University.
- Bailly, D. (1997). *Didactique de l'anglais (Volume 1). Objectifs et contenus de l'enseignement*. Paris: Nathan (coll. Perspectives didactiques).
- Baker, N. (1988). Subtechnical Vocabulary and the ESP Teacher: an Analysis of some Rhetorical Items. En *Medical Journal Articles. Reading in a Foreign Language*, 4 (2), 91-105.
- Baker, B., Hill, K. & Devylder, R. (2000). Core Vocabulary is the Same Across Environments. Comunicación presentada en *California State University at Northridge (CSUN) Conference on Disabilities*. Los Angeles, California.
- Baralo, M. (2001). La construcción del lexicón en español/LE: transferencia y construcción creativa. En M^a A. Martín Zorraquino (Ed.), *¿Qué español enseñar? Normas y variación lingüísticas en la enseñanza del español a extranjeros. Actas del XI Congreso Internacional de ASELE* (pp. 165-174). Zaragoza: Universidad de Zaragoza.
- Bartol Hernández, J. A. (2001). Reflexiones sobre la disponibilidad léxica. En J. A. Bartol Hernández (Coord.), *Nuevas aportaciones al estudio de la lengua española: investigaciones filológicas* (pp. 221-235). Salamanca: Luso-española Ediciones.
- Bartol Hernández, J. A. (2002). El léxico que enseñamos (aportaciones desde la disponibilidad léxica. En V. González Martín (coord.), *Hacia la unidad en la diversidad: difusión de las lenguas europeas* (pp. 62-74). Salamanca: Universidad de Salamanca y Diputación de Salamanca.
- Bartol Hernández, J. A. (2003). Léxico disponible y norma lingüística. En F. Moreno Fernández, J. A. Samper Padilla, M. Vaquero, M. L. Gutiérrez Araus, C. Hernández Alonso & F. Gimeno Menéndez (Coord.), *Lengua, variación y contexto: estudios dedicados a Humberto López Morales* (pp.127-143). Vol. 1. Madrid: Arco Libros.
- Bartol Hernández, J. A. (2004). *Léxico disponible de Soria*. Burgos: Instituto Castellano y Leonés de la Lengua.
- Bartol Hernández, J. A. (2006). La disponibilidad léxica. *Revista española de lingüística*, Año n^o 36, Fasc. 1, 379-384.
- Bartol Hernández, J. A. (2007). El proyecto panhispánico de léxico disponible. En C. Hernández Alonso (Aut.), L. Castañeda San Cirilo (Aut.), *El español de América. Actas del VI Congreso Internacional de "El español de América"*

- (Tordesillas, Valladolid, 25-29 de octubre 2005) (pp. 305-318). Valladolid: Diputación de Valladolid.
- Bartol Hernández, J. A. (2010). Disponibilidad léxica y selección del vocabulario. En R. Castañer Martín & V. Lagüéns Gracia (Eds.) *De moneda nunca usada. Estudios filológicos dedicados a José M. Enguita Utrilla* (pp. 85-107). Zaragoza: Institución Fernando el Católico, CSIC.
- Bartol Hernández, J. A. (2011). Léxicos disponibles de zonas bilingües: interferencias sobre el español". En B. López Meirama (Ed.), *Estudios sobre disponibilidad léxica en el español de Galicia* (pp. 157-88). Santiago de Compostela: Servicio de publicaciones e intercambio científico U. de Santiago de Compostela.
- Bartol Hernández, A. (2014). Idoneidad de los centros de interés clásicos en los estudios de disponibilidad léxica aplicados al español como lengua extranjera: acerca del artículo de González Fernández. *Revista Nebrija de Lingüística Aplicada*, 8(5).
- Bauer, L & Nation, P. (1993). Word Families. *International Journal of Lexicography*, (6) 4. Oxford: Oxford University Press.
- Bauman, J. & Culligan, B. (1995). [Versión de *A General Service List of English Words* (West 1953) ordenada según frecuencia]. Obtenida el 4 de noviembre de 2007 de: <http://jbauman.com/gsl.html>
- Beaumont, P. de (1963). *Le Cheval fou*. Paris: Didier, collection Lire et Savoir.
- Becerra Hiraldo, J. M. (1994). En la enseñanza del español como segunda lengua, semántica y vocabulario. En S. Montesa Peydró & A. Garrido Moraga (Eds.), *Actas del II Congreso Nacional de ASELE, Español para extranjeros: Didáctica e investigación, Málaga* (pp. 313-323). Málaga: ASELE.
- Bellón Fernández, J. J. (2011). *Léxico disponible de la provincia de Córdoba*. Córdoba: Universidad de Córdoba.
- Benítez Pérez P. (1991). Reseña a Azurmendi (1983). *Revista Española de Lingüística*, 21, 355-357.
- Benítez Pérez, P. (1992). Disponibilidad léxica en la zona metropolitana de Madrid. *Boletín de la Academia Puertorriqueña de la Lengua Española*, Segunda Época, 71-102. San Juan.
- Benítez Pérez, P. (1994a). Léxico real/léxico irreal en los manuales de español para extranjeros. En S. Montesa Peydró & A. M. Garrido Moraga (Coords.). *Actas del II Congreso Nacional de ASELE: español para extranjeros: didáctica e investigación* (pp. 325-334). Málaga: ASELE.
- Benítez Pérez, P. (1994b). ¿Qué vocabulario hay que enseñar en las clases de español como lengua extranjera? En L. Miquel & N. Sans (Coord.), *Didáctica del Español como lengua extranjera* (pp. 9-12). Fundación Actilibre: Madrid.
- Benítez Pérez, P. (1995). El vocabulario enseñado en los manuales de ELE (Español como Lengua Extranjera) ¿Es el adecuado? En *Actas de la Jornada Didáctica para profesores de español* (pp. 5-21). Utrech: UFSIA.
- Benítez Pérez, P. (2009). ¿Qué vocabulario hay que enseñar en las clases de español como lengua extranjera? *Monográficos MarcoELE*, vol. 8, 9-12.
- Benítez Pérez, P. & Zebrowski, J. (1993). El léxico español en los manuales polacos. En S. Montesa Peydró & A. Garrido Moraga (Eds.), *Actas III Congreso Nacional de ASELE, El español como lengua extranjera: De la teoría*

- al aula. Málaga, del 10 al 12 de octubre de 1991* (pp. 223-230). Málaga: ASELE.
- Benítez Pérez, P.; Hernández Cabrera, C. E. & Samper Padilla, J. A. (1995). Léxicos básicos de España (LEBAES) y de Canarias (LEBAICan). *Proyectos de Investigación, REALE*, 3, 9-18.
- Berry-Bravo, J. (1993). Teaching the Art of Circumlocution. *Hispania* 76 (2), 371-77.
- Bezdek, J. C. (1981). *Pattern Recognition With Fuzzy Objective Function Algorithms*. New York: Plenum.
- Biber, D., Johansson, S., Leech, G., Conrad, S. & Finegan, E. (1999). *Longman Grammar of Spoken and Written English*. London: Longman.
- Blas Arroyo, J. L. & Casanova Ávalos, M. (2003). La influencia de la lengua materna en algunos valores de la disponibilidad léxica en el español de una comunidad bilingüe". En I. Doval Reixa & M^a R. Pérez Rodríguez (Eds.), *Adquisición, enseñanza y contraste de lenguas, bilingüismo y traducción* (pp. 49-61). Vigo: Universidad de Vigo.
- Bogaards, P. (1994). *Le vocabulaire dans l'apprentissage des langues étrangères*. Paris: Hatier-Didier, Colección Langues et Apprentissage des Langues.
- Bogaards, P. (2001). Lexical units and the learning of foreign language vocabulary. *Applied Linguistics*, 23, 321-343.
- Bogaards, P. (2008). Frequency in Learners' Dictionaries. En E. Bernal & J. DeCesaris, (Eds.), *Proceedings of the XIII Euralex International Congress* (pp. 1231-1236). Barcelona: Universitat Pompeu Fabra.
- Bombarelli, Á. (2005). *La disponibilidad léxica como herramienta didáctica: una propuesta de selección del vocabulario para un nivel umbral de ELE*. Memoria Máster, Universidad de Salamanca.
- Borrego Nieto, J. (2004). Algunas preguntas en relación con el concepto de «léxico disponible». En J. Prado Aragonés & M. V. Galloso Camacho (Eds.), *Diccionario, léxico y cultura* (pp. 59-69-). Huelva: Universidad de Huelva.
- Borrego Nieto, J. & Fernández Juncal, C. (2002). Léxico disponible: aplicaciones a los estudios dialectales. En M. D. Muñoz Núñez *et al.* (Eds.), *IV Congreso de Lingüística General* (pp. 297-306). Cádiz: Universidad de Cádiz / Universidad de Alcalá, vol. 2.
- Bosque Muñoz, I. (Dir.) (2004). *Redes. Diccionario combinatorio del español contemporáneo*. Madrid: SM.
- Bosque Muñoz, I. (Dir.) (2006). *Diccionario combinatorio práctico del español contemporáneo: las palabras en su contexto*. Madrid: SM.
- Briz, A. (2005). Los corpus del español hablado. Presentación. *Oralia*, 8, 7-12.
- Buchanan, M. A. (1927). *A Graded Spanish Word Book*. Toronto: American and Canadian Committees on Modern Languages.
- Butrón, G. (1987). *El léxico disponible: índices de disponibilidad*. Tesis doctoral. Río Piedras: Universidad de Puerto Rico.
- Butrón, G. (1989). Aspectos sociolingüísticos de la disponibilidad léxica. *Asomante* 112, 29-37.
- Butrón, G. (1991). Nuevos índices de disponibilidad léxica. En H. López Morales, (Ed.), *La enseñanza de la lengua materna. Actas del II Seminario Internacional sobre "Aportes de la Lingüística a la enseñanza de la lengua materna"*. Río Piedras: Editorial de la Universidad de Puerto Rico.

- Callealta Barroso, F. J. & Gallego Gallego, D. J. (2016). Medidas de disponibilidad léxica: comparabilidad y normalización. *Boletín de Filología*, Vol. 51, nº. 1, junio 2016.
- Canale, M. (1995). De la competencia comunicativa a la pedagogía comunicativa del lenguaje. En M. Llobera *et al.*, *Competencia comunicativa. Documentos básicos en la enseñanza de lenguas extranjeras* (pp. 63-81). Madrid: Edelsa, Colección Investigación Didáctica.
- Candela, P. & Benegas, G. (2000). *Una rayuela*. Madrid: SGEL.
- Cañizal Arévalo, A. V. (1987). *Disponibilidad léxica en escolares de primaria terminada. Análisis de seis centros de interés*. Memoria de licenciatura inédita. México, D.F.: Universidad Autónoma de México.
- Caramazza, A. (1997). How many levels of processing are there in lexical access? *Cognitive Neuropsychology*, 14, 177-208.
- Carcedo González, A. (1997-1998). Desarrollo de la competencia léxica en el español LE: análisis de cuatro fases de disponibilidad. *Pragmalingüística*, 5-6, 75-94.
- Carcedo González, A. (1998a). Sobre las pruebas de disponibilidad léxica para estudiantes de español/LE, *RILCE: Revista de filología hispánica*, (monográfico *Español como lengua extranjera: investigación y docencia*), 14, 2, 205-224.
- Carcedo González, A. (1998b). Tradición y novedad en las aportaciones hispánicas a los estudios de disponibilidad léxica, *Lingüística*, 10, *ALFAL*, 5-68.
- Carcedo González, A. (1999a). Estudio comparativo del vocabulario español (LE) disponible de estudiantes finlandeses y el de la sintopía madrileña: propuestas didácticas. *Documentos de Español Actual*, 1, 73-87.
- Carcedo González, A. (1999b). Vocabularios básicos, fundamentales y disponibles. Renovación metodológica de la investigación hispánica. En Y. Gambier & E. Suomela-Salmi (Eds.), *Publications du Département d'Etudes Françaises 2* (pp. 191-207). Jalons, Turku: Universidad de Turku.
- Carcedo González, A. (1999c). Análisis de errores léxicos del español en la interlengua de los finlandeses. En T. Jiménez, Losada & Márquez (Eds.), *Español como lengua extranjera: enfoque comunicativo y gramática* (pp. 465-472). Santiago de Compostela: Universidad de Santiago de Compostela,
- Carcedo González, A. (2000). *Disponibilidad léxica en español como lengua extranjera: el caso finlandés (estudio del nivel preuniversitario y cotejo con tres fases de adquisición)*. Turku: Universidad de Turku.
- Carcedo González, A. (2001). *Léxico Disponible de Asturias*. Turku: Universidad de Turku, Publicaciones del Departamento de Español.
- Carcedo González, A. (2003). Unidad y variedad diatópica de la disponibilidad léxica del español: comparación de los inventarios de Puerto Rico, Cádiz y Asturias. En F. Moreno Fernández *et al.* (Coords.). *Lengua, variación y contexto: estudios dedicados a Humberto López Morales* (pp. 199-225), Vol. 1. Madrid: Arco Libros.
- Carcedo González, A. (2004). La variable léxica disponible en la comparación interdialectal: compatibilidad de la norma asturiana con otras sintopías hispanohablantes. En V. Sánchez Corrales (Ed.), *Actas del XIII Congreso*

- Internacional de la Asociación de Lingüística y Filología de la América Latina (ALFAL)* (pp. 661-679). San José: Universidad de Costa Rica.
- Carroll, J. B.; Davies, P. & Richman, B. (1971). *The American Heritage Word Frequency Book*. Boston: Houghton Mifflin Company.
- Cartaya Febres, N. (2011). La hipótesis del umbral léxico en la comprensión de lectura en L2. *Lingua Americana* 104 / Año XV N° 28 (enero-junio), 103-118.
- Carter, R. (1987). *Vocabulary: Applied Linguistic Perspectives*, London: Allen and Unwin.
- Carter, R. (1996, 1998). *Vocabulary. Applied Linguistic Perspectives*. London: Routledge.
- Carter, R. & McCarthy, M. (1996). *Vocabulary and Language Teaching*. New York: Longman (1ª ed.: 1988).
- Carter, R. & McCarthy, M. (2005). *Cambridge Grammar of English*. Cambridge: Cambridge University Press.
- Casanova Avalos, M. (2006). La disponibilidad léxica en la Comunidad Valenciana. J. L. Blas Arroyo, M. Velando Casanova & M. Casanova Avalos (Coord.), *Discurso y sociedad: contribuciones al estudio de la lengua en contexto social* (pp. 737-752). Valencia: Universidad Jaume I, Servicio de Publicaciones.
- Cassany, D., Luna, M. & Sanz, G. (1998). *Enseñar lengua*. Barcelona: Graó, 4ª ed.
- Casares, J. (1950). *Introducción a la lexicografía moderna*. Madrid: CSIC.
- Casso, J. (2010). *Análisis y revisión crítica de los materiales de evaluación de la competencia léxica. Elaboración de un test de vocabulario de nivel umbral*. Memoria de máster. Universidad Antonio de Nebrija.
- Castañer, R. M. (2008). Los extranjerismos en el léxico disponible de los estudiantes aragoneses. En M. L. Arnal (Ed.), *Estudios sobre disponibilidad léxica en los jóvenes aragoneses* (pp. 51-74). Zaragoza: Institución Fernando el Católico.
- Cervero, Mª J. & Pichardo, F. (2000). *Aprender y enseñar vocabulario*. Madrid: Edelsa.
- Chamot, A. U., Barnhardt, S., El-Dinary, P. B., & Robbins, J. (1999). *The learning strategies handbook*. White Plains, NY: Addison Wesley Longman.
- Chung, T. & Nation, I. S. P. (2003). Technical vocabulary in specialised texts. *Reading in a Foreign Language* 15, 2, 103-116.
- Closset, F. (1958). *La enseñanza de los idiomas modernos: objetivos, métodos, procedimientos y material*. Madrid: Ministerio de Educación Nacional.
- Coady, J. (1997). L2 vocabulary acquisition through extensive reading. En J. Coady & T. Huckin (Eds.), *Second language vocabulary acquisition: A rationale for pedagogy* (pp. 225-237). Cambridge: Cambridge University Press.
- Cobas, L. [Autor]; Llinares, G. [Tutor]. (2003). *La instrucción directa de vocabulario de vista de alta frecuencia y la comprensión de lectura en L2* [Tesis de maestría no publicada]. Caracas: Universidad Simón Bolívar.
- Cobb, T. (sin fecha). Why & How to Use Frequency Lists to Learn Words. Online en <http://www.lex Tutor.ca/research/>. [consultado a 28 de abril de 2017].
- Cook, V. (2001). *Second Language Learning and Language Teaching* (3rd ed.). London: Arnold.

- Corda, A. & Marelló, C. (1999). *Insegnare e imparare il lessico*. Torino, Paravia: Colección Italiano Lingua Straniera. Formazione degli Insegnanti.
- Corpas Pastor, G. (1996). *Manual de fraseología española*. Madrid: Gredos.
- Corpas, J. Garmendia, A. & Soriano, C. (2005). *Aula internacional 1*. Barcelona: Editorial Difusión.
- Coxhead, A. (2000). A New Academic Word List. TESOL. *Quarterly* 34(2), 213-238. [Enlace a la lista: <http://www.victoria.ac.nz/lals/resources/academicwordlist/>]
- C.R.E.D.I.F (1958). *Méthode de lecture à l'intention des travailleurs originaires d'Afrique du Nord*. (1^o et 2^o degrés). Paris: l'Institut Pédagogique National.
- C.R.E.D.I.F (1962). *Voix et Images de France*. Paris: Didier.
- C.R.E.D.I.F. (1963). *Bonjour Line*. Paris: Didier.
- C.R.E.D.I.F. *Méthode de lecture destinée aux femmes dont la langue maternelle est l'arabe, le kabyle ou le berbère*, réalisée.
- Criado, R. & Sánchez, A. (2012). Lexical frequency, textbooks and learning from a cognitive perspective. A corpus-based sample analysis of ELT materials. *Revista Española de Lingüística Aplicada, Volumen Extra 1*, 77-94.
- Crowther, J., Dignen, S. & Lea, D. (Ed.) (2002). *Oxford Collocations Dictionary for Students of English*: Oxford: Oxford University Press.
- Cruse, D. A. (1986). *Lexical semantics*. Nueva York: C.U.P.
- Cruz Alonso, R. (2016). *El léxico disponible de Castilla y León*. Salamanca: Ediciones Universidad de Salamanca.
- Crystal, D. (1994). *Enciclopedia del Lenguaje de la Universidad de Cambridge*. Madrid: Taurus. [Edición española dirigida por J. C. Moreno Cabrera].
- Dalurzo, M. J. & González, L. A. (Colaborador) (2010). Disponibilidad Léxica en Córdoba, Argentina. *Salud y aprendizajes lingüísticos. Complejidades en la enseñanza de la lengua*, Tomo II, 279-344. Córdoba, Argentina: Editorial Comunicarte.
- Davesne, A. (1955). *Le français élémentaire*. Paris: Istra.
- Davies, M. (2006). *A Frequency Dictionary of Spanish: Core Vocabulary for Learners*. New York: Routledge.
- Davies, M. & Face, T. L. (2006). Vocabulary Coverage in Spanish Textbooks: How Representative is It? In N. Sagarra, & A. J. Toribio (Eds.), *Selected Proceedings of the 9th Hispanic Linguistics Symposium* (pp. 132-143). Somerville, MA: Cascadilla.
- De Grève, M. & Van Passel, F. (1971). *Lingüística y enseñanza de lenguas extranjeras*. Madrid: Fragua.
- De Kock, J. (1983). De la fréquence relative des phonèmes en français et de la relativité de ces fréquences. *ITL: International Journal of Applied Linguistics*, n^o.59, 1-54.
- De Prada, M., Salazar, D., & Molero, C. M. (2012). *Uso interactivo del vocabulario y sus combinaciones más frecuentes*. Madrid: Edelsa.
- Del Barrio, F. (2016) Algunas observaciones sobre la disponibilidad léxica en estudiantes itálofonos de español. En E. Sainz, I. Solís, F. del Barrio & I. Arroyo (Eds), *Geométrica Explosión Estudios de lengua y literatura en homenaje a René Lenarduzzi* (pp. 127-144).
- Deutscher Volkshochschul-Verband & Goethe-Institut (1977). *Das Zertifikat Deutsch als Fremdsprache*. Munich: Germany.

- Díaz Castañón, C. (1977). *Vocabulario básico del español y sus aplicaciones a la enseñanza*. Oviedo: ICE.
- DISPOGAL = Disponibilidad léxica en Galicia.
<http://gramatica.usc.es/proyectos/disposal/>
- DISPOLEX = dispoLex/investigación léxica. <http://www.disporex.com/>
- Dos Santos Fernández, M. (2013). *Disponibilidad léxica en alumnos de ELE del distrito de Oporto (Portugal)*. Trabajo Fin de Máster defendido en el curso 2012-2013 en la UNED.
- Dottrens R. & Massarenti, D. (1948). *Vocabulaire fondamental du français. Contribution à un enseignement rationnel de l'orthographe d'usage*. Genève: Delachaux & Niestlé.
- Dudley-Evans, A. & St. John, M. J. (1998). *Developments in ESP. A MultiDisciplinary Approach*. Cambridge: Cambridge University Press.
- Eaton, H. (1940). *An English-French-German-Spanish Word Frequency Dictionary*. New York: Dover Publications, Inc.
- Echeverría, M. S. (1991). Crecimiento de la disponibilidad léxica en estudiantes chilenos de nivel básico y medio. En H. López Morales (Ed.). *La enseñanza del español como lengua materna. Actas del II Seminario Internacional sobre Aportes de la lingüística a la enseñanza del español como lengua materna* (pp. 61-78), Río Piedras: Universidad de Puerto Rico.
- Echeverría, M. S. (2001). *Vocabulario disponible (en CD)*. Concepción, Chile: Universidad de Concepción.
- Echeverría, M. S. (2002). Programas computacionales para el español como lengua materna. *Revista Signos*, 35(51-52), 163-193.
- Echeverría, M. S., Herrera, M. O., Moreno, P. & Pradenas, F. (1987). Disponibilidad léxica en Educación Media. *Revista de Lingüística Teórica y Aplicada* 25, 55-115.
- Echeverría, M. S., Urzúa, P. & Figueroa, I. (2005). *Dispogen II. Programa computacional para el análisis de la disponibilidad léxica*. Concepción, Chile: Universidad de Concepción.
- Echeverría, M. S., Vargas, R., Urzúa, P. & Ferreira, R. (2008). *DispoGrafo: una nueva herramienta computacional para el análisis de relaciones semánticas en el léxico disponible*. *Revista de Lingüística Teórica y Aplicada*, 46, 1, 81-91.
- Echeverría, M. S.; Urzúa, P. & Sáez, K. (2006). Disponibilidad léxica matemática. Análisis cuantitativo y cualitativo. *RLA. Revista de Lingüística Teórica y Aplicada*, 44 (2), II Sem., 59-76.
- Eckersley, Ch. E. (1938). *Essential English for Foreign Students*. London: Longman, Green & Co.
- Eckersley, Ch. E. (1943). *Essential English: a progressive course for foreign students. Book I*. London: Longmans, Green and Co.
- Escoriza Morera, L. (2003). *Lenguas en contacto y disponibilidad léxica: la situación lingüística e intercultural de Ceuta y Gibraltar*, *Linred: Lingüística en red*, nº. 1.
- Escoriza Morera, L. (2007). *Disponibilidad léxica y multilingüismo. El contacto entre inglés y español en Gibraltar*. En P. Cano López (Coord.), *Actas del VI Congreso de Lingüística General (Santiago de Compostela, del 3 al 7 de mayo*

- de 2004), Vol. 2, *Las lenguas y su estructura* (pp. 2477-84). Madrid: Arco/Libros.
- Esquer Torres, R. (1968). *Didáctica de la lengua española*. Madrid: Ediciones Alcalá.
- Estoup, J. B. (1905-1907). *Gammes sténographiques*. París: Institut sténographique.
- Etxebarria Arostegui, M. (1996). Disponibilidad léxica en escolares del País Vasco". *Revista española de lingüística*, nº 26, fasc. 2, 301-26.
- Etxebarria Arostegui, M. (1999). Disponibilidad léxica y enseñanza de la lengua materna en el área metropolitana del Gran Bilbao. En J. A. Samper Padilla M. & Troya Déniz, M. (Coords.), *Actas del XI Congreso Internacional de la Asociación de Lingüística y Filología de la América Latina en Las Palmas de Gran Canarias, del 22 al 27 de julio de 1996*. Vol. 2, 1479-94.
- Ezquerro, R. (1974). Los diccionario de frecuencias en español. *Boletín de la Asociación Europea de Profesores de Español, Año VI. N.º. 10 Marzo*, 3-27.
- Fajardo Aguirre, A. (1991). Los vocabularios básicos. En S. Montesa Peydró & A. Garrido Moraga (Coord.), *Actas del III Congreso Nacional de ASELE. El español como lengua extranjera: de la teoría al aula. Málaga, del 10 al 12 de octubre de 1991* (pp. 215-222). Málaga: ASELE.
- Farrar, F. W. (1865). *Chapters of language*. London: Longmans, Green and Co.
- Fernández Juncal, C. (2013). *Léxico disponible en Cantabria*. Santander: Ediciones Universidad de Cantabria
- Fernández Leyva, H. (2015). Posible aplicación del índice de disponibilidad léxica a la selección del vocabulario de manuales de ELE. *MarcoELE: revista de didáctica ELE*, nº 20.
- Fernández-Merino Gutiérrez, P. V. (2011). Presencia del léxico disponible de inmigrantes en glosarios específicos de vocabulario. *Cuadernos Comillas*, 2, 1-18.
- Fernández-Merino Gutiérrez, P. V. (2013). *Disponibilidad léxica de inmigrantes en Castilla y León*. Tesis doctoral. Valladolid: Universidad de Valladolid.
- Fernández-Merino Gutiérrez, P. V. (2015). *Léxico disponible de inmigrantes en Castilla y León. Estudio y diccionarios*. Editorial Sapere Aude.
- Ferreira, R. & Echeverría, M. S. (2010). Redes semánticas en el léxico disponible de inglés L1 e inglés LE. *Onomázein*, nº. 21, 133-153.
- Firth, J.R. (1957). *Papers in Linguistics 1934-1951*. Londres: Oxford University Press.
- Fourré, P. (1963). *Premier Dictionnaire en images*. Paris: Didier.
- Francis, W. N. & Kučera, H. (1964, 1971, 1979). *The Brown University Standard Corpus of Present-Day American English*. Providence, Rhode Island: Brown University.
- Francis, W. N. & Kučera, H. (1982). *Frequency Analysis of English Usage: Lexicon and Grammar*. Boston: Houghton Mifflin.
- Frey Pereira, M. L. H. (2007). Disponibilidad léxica y escritura del español como lengua extranjera: propuestas de comparación de dos corpus. *Interlingüística*, 17, 366-373.
- Frisby, A. W. (1957). *Teaching English; notes and comments on teaching English Overseas*. London: Longmans.

- Fuentes González, A. D. (1997). Disponibilidad léxica y fines específicos. En J. García Marcos (Ed.), *Estudios de disponibilidad léxica* (pp. 127-146). Almería, Barcelona: GRUSTA.
- Gairns, R. & Redman, S. (1986). *Working with words*. Cambridge: Cambridge University Press.
- Galisson, R. (1979). *Lexicologie et enseignement des langues*. Paris: Hachette.
- Galisson, R. (1991). *De la langue à la culture par les mots*. Paris: Clé International.
- Gallego Gallego, D. J. (2014). *Léxico Disponible de Estudiantes de Español Como Lengua Extranjera en la Comunidad de Madrid*. Tesis doctoral. Alcalá de Henares: Universidad de Alcalá.
- Galoso Camacho, M^a V. (1998). Hacia una definición de léxico disponible: posibles aplicaciones. En J. Fernández González *et al.* (Eds.), *Actas III Congreso de Lingüística General. Lingüística para el siglo XXI* (pp. 711-717). Colección *Aquilafuente*, 9. Salamanca: Universidad de Salamanca.
- Galoso Camacho, M^a V. (2001). *El léxico de los estudiantes preuniversitarios en el distrito universitario de Salamanca*. Tesis doctoral. Salamanca: Universidad de Salamanca.
- Galoso Camacho, M^a V. (2002a). *El léxico disponible de los estudiantes preuniversitarios en el distrito universitario de Salamanca*. Salamanca: Ediciones de la Universidad de Salamanca.
- Galoso Camacho, M^a V. (2002b). *El léxico disponible en el nivel preuniversitario: provincia de Zamora*. Trabajo de Grado. Huelva: Universidad de Huelva.
- Galoso Camacho, M^a V. (2003). *El léxico disponible de Ávila, Salamanca y Zamora*. Valladolid: Fundación Instituto Castellano y Leonés de la Lengua.
- Galoso Camacho, M^a V. & Prado Aragonés, J. (2005). La estructura estadística del léxico disponible de informantes nativos e informantes de ELE. En M. A. Castillo Carballo (coord.), *Las gramáticas y los diccionarios en la enseñanza del español como segunda lengua, deseo y realidad. Actas del XV Congreso Internacional de ASELE, Sevilla 22-25 de septiembre de 2004* (pp. 370-375). Sevilla: Universidad de Sevilla, Secretariado de Publicaciones.
- Gamazo Carretero, E. (2014). Estereotipos en el léxico disponible de universitarios portugueses. En E. Tobar & M. E. Mañas (Coord.), *El español como lengua extranjera en Portugal: retos de la enseñanza de lenguas cercanas* (pp. 28-41). Ministerio de Educación, Cultura y Deporte.
- García Hoz, V. (1953). *Vocabulario usual, común y fundamental: determinación y análisis de sus factores*. Madrid: C.S.I.C.
- García Marcos, F. (1985). *El segmento fónico vocal +s en ocho poblaciones de la Costa Granadina. Aportación sociolingüística, estadística e informática al reexamen de la cuestión*. (Memoria de Licenciatura inédita). Granada: Universidad de Granada.
- García Marcos, F. & Mateo García, M^a V. (1997). Resultados de las encuestas sobre disponibilidad léxica realizadas en Almería. *REALE: revista de estudios de adquisición de la lengua española*, n^o 7, 57-68.
- García Marcos, F. & Mateo García, M^a V. (2003). La selección de materiales léxicos en la enseñanza de lenguas extranjeras. Evaluación y propuestas desde la disponibilidad léxica. En L. Negró (Ed.). *Mélanges offerts à Charles Leselbaum* (pp. 297-312). Collecton Études hispaniques et ibéro-américaines appliquées. París: Editions hispaniques.

- García Megía, A. (2003). *La disponibilidad léxica en la ciudad de Almería entre los nueve y los doce años*. Tesis doctoral. Almería: Universidad de Almería.
- García Surrallés, C. (2012). *Léxico básico de la modalidad andaluza del español*. Cádiz: Universidad de Cádiz.
- Gómez Devís, M^a B. (2004). *La disponibilidad léxica de los estudiantes preuniversitarios valencianos: reflexión metodológica, análisis sociolingüístico y aplicaciones*. Valencia: Servicio de Publicaciones de la Universidad de Valencia.
- Gómez Devís, M^a B. (2005). Aprovechamiento del léxico disponible. *Lenguaje y Textos* 23, 69-75.
- Gómez Devís, M^a B. (2010). Aportes de la disponibilidad léxica a la investigación sobre fenómenos de contacto lingüístico. En M. García (Ed.), *Interacció Comunicativa i Ensenyament de Llengües* (pp. 233-239). Valencia: Universitat de València.
- Gómez Devís, M^a B. & Llopis Rodrigo, F. (2016). *Xarxes semàntiques en el lèxic disponible de valència. Una aproximació al léxico bilingüe*. *Studia Romanica Posnaniensia*, Vol. 43, n^o. 2, 65-83.
- Gómez Molina, J. R. (1997). El léxico y su didáctica: una propuesta metodológica. *REALE, Revista de Estudios de Adquisición de la Lengua Española*, 7, 69-93.
- Gómez Molina, J. R. (2000a). Enseñanza-aprendizaje de las locuciones idiomáticas del español en el aula de lengua extranjera. En *Studi di Lingua e Culturologia. In onore di Aleksandr Lobodanov* (pp. 145-168). Moscú: Università Statale di Mosca.
- Gómez Molina, J. R. (2000b). Las unidades fraseológicas del español: una propuesta metodológica para la enseñanza de las locuciones en la clase de ELE. *Quaderns de Filologia. Estudis Lingüistics*, V, 111-134.
- Gómez Molina, J. R. (2003). Las unidades léxicas: tipología y tratamiento en el aula de ELE. *Mosaico*, 11, 4-11.
- Gómez Molina, J. R. (2004a). Las unidades léxicas en español. *Carabela*, 56, 27-50.
- Gómez Molina, J. R. (2004b). Los contenidos léxico-semánticos. En J. Sánchez Lobato & I. Santos (Dirs.), *Enseñar español como segunda lengua y lengua extranjera. Vademécum para la formación de profesores* (pp. 789-810). Madrid: SGEL.
- Gómez Molina, J. R. (2004c). La subcompetencia léxico-semántica. En J. Sánchez Lobato & I. Santos (Dirs.), *Enseñar español como segunda lengua y lengua extranjera. Vademécum para la formación de profesores* (pp. 491-510). Madrid: SGEL.
- Gómez Molina, J. R. (2005a). La competencia léxica en el currículo de Español para fines específicos (EpFE). En V. de Antonio, R. Cuesta, A. Van Hooft, B. de Jonge, J. Robisco & M. Ruiz (Eds.), *Español para Fines Específicos. Actas del II Congreso Internacional de Español para Fines Específicos, Amsterdam, noviembre de 2003* (pp. 82-104). España: Ministerio de Educación y Ciencia.
- Gómez Molina, J. R. (2005b). La competencia léxica en la enseñanza del español como L2 y LE. *Mosaico*, 5, 23-29.
- Gómez Molina, C. & De Kock, J. (1982). Concordancias e índices automáticos a disposición de la enseñanza del español, lengua extranjera. *LEA: Lingüística española actual*, Vol. 4, n^o 1, 47-82.

- Gómez Molina, J. R. & Gómez Devís, M. B. (2004). *La disponibilidad léxica de los estudiantes preuniversitarios valencianos. Estudio de estratificación sociolingüística*. València: Universitat de València.
- González Fernández, J. (2013) La disponibilidad léxica de los estudiantes turcos de español como lengua extranjera. *Marcoele Revista de Didáctica español como lengua extranjera*, nº 16.
- González Fernández, J. (2014). Idoneidad de los centros de interés clásicos en los estudios de disponibilidad léxica aplicados al español como lengua extranjera. *Revista Nebrija de Lingüística Aplicada*, 8(3).
- González Fernández, J. (2015). *Diccionario básico de aprendizaje. La planta de un diccionario temático monolingüe del nivel acceso (A1) de español como lengua extranjera*. Tesis doctoral. Universidad de Jaén.
- González Martínez, A. (2002). *Disponibilidad léxica de Cádiz*. Cádiz: Universidad de Cádiz.
- González, A., y Orellana, P. (2001). Compatibilidad y discrepancia entre los léxicos disponibles de Puerto Rico y Cádiz. *Pragmalingüística*, 8-9, 179-196.
- Gougenheim, G. (1958). *Dictionnaire fondamental de la langue française*. Paris: Didier
- Gougenheim, G., R. Michéa, P. Rivenc, & A. Sauvageot, (1956): *L'élaboration du Français élémentaire. étude sur l'établissement d'un vocabulaire et d'une grammaire de base*. París: Didier.
- Gougenheim, G.; Michéa, R.; Rivenc, P. & Sauvageot, A. (1956). *L'élaboration du français élémentaire : étude sur l'établissement d'un vocabulaire et d'une grammaire de base*, Didier, Paris. Con una nueva edición revisada: (1964). *L'élaboration du français fondamental : étude sur l'établissement d'un vocabulaire et d'une grammaire de base*. Paris: Didier.
- Gougenheim, G.; Michéa, R.; Rivenc, P. & Sauvageot, A. (1958). *Dictionnaire fondamental de la langue française*. Paris: Didier. (Contiene 3.500 palabras, nueva edición revisada en 1977).
- Goulden, R., Nation, I.S.P. & Read, J. (1990). How can large can a receptive vocabulary be? *Applied Linguistics* 11, 4, 341-363.
- Guerra Salas, L. & Gómez Sánchez, M^a E. (2004). Español de los medios de comunicación: aspectos de disponibilidad léxica. En H. Perdiguero Villarreal & A. A. Álvarez (Coord.), *Actas del XIV Congreso Internacional de ASELE. Burgos 2003* (pp 336-343). Burgos: ASELE.
- Guerra Salas, L. & Gómez Sánchez, M^a E. (2005). El español de los profesionales de la radio: una aproximación desde la disponibilidad léxica. En M. L. Carrió Pastor (Coord.), *Perspectivas interdisciplinarias de la lingüística aplicada, Vol. 1*, (pp. 351-360). AESLA: Valencia.
- Gugenberger, E. (2013). O cambio de paradigma nos estudos sobre contacto lingüístico: pode ser útil o concepto de hibrididade para a lingüística y a política de linguas en España". En E. Gugenberger, H. Monteagudo & G. Rei-Doval (eds.), *Contacto de linguas, hibrididade, cambio: contextos, procesos e consecuencias* (pp. 17-48). Santiago de Compostela, Consello da Cultura Galega.
- Guiraud, P. (1959). *Problèmes et méthodes de la statistique linguistique*. Dordrecht: Reidel

- Haensch, G. & Omeñaca, C. (2004). *Los diccionarios del español en el siglo XXI*. Salamanca: Universidad.
- Halliday, M.A.K. (1966). Lexis as a linguistic level. En C. E. Bazell *et. al.* (Ed.), *In Memory of John Firth*. Londres: Longmans.
- Harmer, J. (1991). *The Practice of English Language Teaching*. Londres: Longman ELT.
- Henmon, V. A. C. (1924). *A French Word Book Based on a Count of 400.000 Running Words*. Wisconsin: Madison Bureau of Educational Research, University of Wisconsin.
- Henríquez Guarín, M.C., Mahecha, V., & Mateus Ferro, G.E. (2016). Análisis de los mecanismos del léxico disponible de cuerpo humano a través de grafos. *Lingüística y Literatura* 69, 229-251.
- Hernández Cabrera, C. E. & Samper Hernández, M. (2004). Léxico disponible, norma culta y norma popular. *Archivo de Filología Aragonesa*, Vol. 59-60, 2, 2002-2004, 1341-1358.
- Hernández Cabrera, C. E. & Samper Padilla, J. A. (2007). Léxico regional en los materiales de disponibilidad léxica. *Revista de filología de la Universidad de La Laguna*, nº 25 (Ejemplar dedicado a Homenaje a Antonio Lorenzo), 287-290.
- Hernández Muñoz, N. (2004). *El léxico disponible de los estudiantes conquenses*. Salamanca: Ediciones Universidad de Salamanca.
- Hernández Muñoz, N. (2005). La disponibilidad léxica: una herramienta fronteriza para el estudio del léxico en Lingüística y Psicología. En E. Díez Villoria, B. Zubiauz de Pedro & M. Á. Mayor Cinca (Coord.), *Estudio sobre la adquisición del lenguaje* (pp. 942-953). Salamanca: Universidad de Salamanca.
- Hernández Muñoz, N. (2006). *Hacia una teoría cognitiva integrada de la disponibilidad léxica: el léxico disponible de los estudiantes castellano-manchegos*. Salamanca: Ediciones Universidad de Salamanca.
- Hernández Muñoz, N. (2007). Relaciones entre palabras y categorías semánticas. Ponencia presentada en el *Curso Superior de Filología*, Salamanca, 9-13 de julio de 2007.
- Hernández Muñoz, N. (2010). El cambio de canal en la producción léxica experimental. *LEA*, XXXII/2, 141-171.
- Hernández Muñoz, N. (2014). Categorías en el léxico bilingüe: perspectivas desde el *priming* semántico interlenguas y la disponibilidad léxica. *Revista Electrónica de Lingüística Aplicada*, 1, 19-38.
- Hernández Muñoz, N. (2015). La evaluación de la competencia léxica adulta: una aproximación a través de la disponibilidad léxica y la especialización académica en preuniversitarios. *Revista de filología de la Universidad de La Laguna*, 33, 79-99.
- Hernández Muñoz, N. & Borrego Nieto, J. (2004). Cuestiones metodológicas sobre los estudios de disponibilidad léxica. En M. Villayandre Llamazares (Coord.), *Actas del V Congreso de Lingüística General: León 5-8 de marzo de 2002*, Vol. 2 (pp. 1519-1527). Madrid: Arco Libros.
- Hernández Muñoz, N. & Bartol Hernández, J. A. (2006). Proyecto del léxico disponible en España. En J. L. Blas Arroyo, M. Velando Casanova, M. Casanova Avalos (Eds.), *Discurso y sociedad: contribuciones al estudio de la*

- lengua en contexto social* (pp. 725-736). Castellón de la Plana: Publicacions de la Universitat Jaume I.
- Hernández Muñoz, N. & Izura, C. (2010). ¿Perro o musaraña? La recuperación léxica en las categorías semánticas, *Ciencia Cognitiva*, (4)1, 1-3.
- Hernández Muñoz, N., Izura, C. & Ellis, A. (2006). Cognitive Aspects of lexical availability. *European Journal of Cognitive Psychology*, 18, 734-755.
- Herreros Marcilla, M. (2014). La disponibilidad léxica como propuesta de investigación docente. Análisis en alumnos inmigrantes no hispanohablantes. Comunicación presentada en el *XXV Congreso Internacional de ASELE* celebrado del 17 al 20 de septiembre de 2014 en la Universidad Carlos III de Madrid.
- Herreros Marcilla, M. (2016). Disponibilidad léxica oral en inmigrantes de origen marroquí. Análisis sociolingüístico en torno a las variables sexo, edad y años de residencia en España. En N. Domínguez García, C. Fernández Juncal & J. L. García Alonso (Dir.), *Innovación y desarrollo en español como lengua extranjera* (pp. 129-145.). Salamanca: Universidad de Salamanca.
- Higueras García, M. (1996). Aprender y enseñar léxico. En L. Miquel & N. Sans (Coors.), *Didáctica del español como lengua extranjera*, 3, (pp. 111-126). Madrid: Colección Expolingua, Fundación Actilibre.
- Higueras García, M. (1997a). La importancia del componente idiomático en la enseñanza del léxico a extranjeros. *Frecuencia L*, 15-19.
- Higueras García, M. (1997b). Las unidades léxicas y la enseñanza del léxico a extranjeros. *REALE: revista de estudios de adquisición de la lengua española*, 8, 35-50
- Higueras García, M. (2004a). Claves prácticas para la enseñanza del léxico. En *La enseñanza de léxico en español como segunda lengua/lengua extranjera. Carabela 56* (pp. 5-25). Alcobendas: SGEL. (El mismo artículo exactamente está también publicado en: 2012, *Revista Nebrija de Lingüística aplicada a la enseñanza de Lenguas*, 11.)
- Higueras García, M. (2004b). *La enseñanza-aprendizaje de las colocaciones en el desarrollo de la competencia léxica en el español como lengua extranjera*. Tesis Doctoral. Universidad Complutense de Madrid.
- Higueras García, M. (2005). Necesidad de un diccionario de colocaciones para aprendientes de ELE. En M. A. Castillo Carballo; O. Cruz Moya; J. M. García Platero; J. P. Mora Gutiérrez (coord.), *Las gramáticas y los diccionarios en la enseñanza del español como segunda lengua: deseo y realidad. Asociación para la Enseñanza del Español como Lengua Extranjera. XV Congreso Internacional de la ASELE: Sevilla, 2004* (pp. 480-490). Sevilla: Universidad de Sevilla, ASELE.
- Higueras García, M. (2006). *Las colocaciones y su enseñanza en la clase de ELE*. Madrid: Arco/libros.
- Higueras García, M. (2007). *Estudio de las colocaciones léxicas y su enseñanza en Español como Lengua Extranjera*. Colección Monografías ASELE, nº 9, 2006. Málaga: Ministerio de Educación, Cultura y Deporte, Centro de Investigación y Documentación Educativa.
- Higueras García, M. (2009). *Aprender y enseñar léxico*. Monográficos marcoELE, nº 9, Didáctica del español como lengua extranjera. Expolingua 1996.

- Hornby, A. S. & Cowie, A. P. (1995). *Oxford Advanced Learner's Dictionary of Current English* (5ª Edición). Oxford: Oxford University Press.
- Horst M., Cobb, T. & Meara, P. (1998). Beyond A Clockwork Orange: Acquiring Second Language Vocabulary through Reading. *Reading in a foreign Language* 11(2), 207-223.
- Howatt, A. P. R. (1984). *A History of English Language Teaching*. Oxford: Oxford University Press.
- Hoz Fernández, C. de la (2005). Las palabras comodín. Sobre cosas, cacharros, chismes y cachivaches. En A. Álvarez et al. (Eds.) *Actas del XVI Congreso Internacional de ASELE* (pp. 396-401). Oviedo: Universidad de Oviedo.
- Hu, M. & Nation, I. S. P. (2000). Unknown vocabulary density and reading comprehension. *Reading in a Foreign Language* 13, 403-430.
- Hugo, N. H. (2003). *Disponibilidad léxica en español como lengua extranjera*. Tesis de licenciatura inédita. Universidad de Saarbrücken.
- Hwang, K. & Nation, I. S. P. (1995). Where would general service vocabulary stop and special purposes vocabulary begin?. *System* 23, 1, 35-41.
- Hymes, D. (1971). *On Communicative Competence*. Philadelphia, PA: University of Pennsylvania Press. Trad. esp. en M. Llobera (Coord.) (1995). *Competencia comunicativa. Documentos básicos en la enseñanza de lenguas extranjeras*. Madrid: Edelsa.
- InfoStat (2004). *InfoStat, versión 2004. Manual del usuario*, (1ª ed.) Córdoba, Argentina: Grupo InfoStat, FCA, Universidad Nacional de Córdoba, Editorial Brujas.
- Consejo de Europa (2001). *Marco común europeo de referencia para las lenguas: aprendizaje, enseñanza, evaluación*. Madrid: Ministerio de Educación, Cultura Deporte y Editorial Anaya, 2002. Disponible en línea <http://cvc.cervantes.es/obref/marco/>
- Instituto Cervantes (2006). *Plan Curricular del Instituto Cervantes. Niveles de referencia para el español*. Madrid: Instituto Cervantes / Edelsa, Biblioteca Nueva (3 vols.). También disponible en red: http://cvc.cervantes.es/ensenanza/biblioteca_ele/plan_curricular/default.htm
- Izquierdo Gil, M. C. (2002). Los corpóra electrónicos en la enseñanza del vocabulario español LE: la utilidad pedagógica de las concordancias. En A. Giméno Sanz (Coord.), *Actas del XII Congreso Internacional de ASELE: tecnologías de la información y de las comunicaciones en la enseñanza de la E/LE. 2001, Valencia* (pp. 201-212). Valencia: ASELE.
- Izquierdo Gil, M. C. (2004). *La selección del léxico en la enseñanza del español como lengua extranjera: su aplicación en el nivel elemental en estudiantes francófonos*. València: Universitat de Valencia, Servei de Publicacions.
- Izquierdo Gil, M. C. (2005). *La selección del léxico en la enseñanza del español como lengua extranjera*. Málaga: ASELE.
- Izquierdo Gil, M. C. (2016). *La selección del léxico en la enseñanza del español como lengua extranjera*. Biblioteca Virtual Miguel de Cervantes.
- Jiménez Berrio, F. (2012). Aprendizaje integrado de contenidos y español para inmigrantes desde la perspectiva de la disponibilidad léxica: enseñanza del vocabulario. En R. Breeze, F. Jiménez Berrio, C. Llamas Saíz, C. Martínez Pasamar & C. Taberero Sala (Ed.), *Teaching Approaches to*

- CLIL/Propuestas docentes en AICLE* (pp. 149-163). Pamplona: Servicio de Publicaciones de la Universidad de Navarra.
- Jiménez Berrio, F. (2013). *Léxico disponible de inmigrantes escolares no hispanohablantes*. Pamplona: Servicio de Publicaciones de la Universidad de Navarra.
- Jiménez Catalán, R. M. (2002). El concepto de competencia léxica en los estudios de aprendizaje y enseñanza de segundas lenguas, *Atlantis*, Vol. XXIV, Núm. 2 (Junio), 149-62.
- Jiménez Catalán, R. M. (2014). *Lexical Availability in English and Spanish as a Second Language*. New York: Springer.
- Jiménez Catalán, R. M. (2017). Estudios de disponibilidad léxica en español y en inglés: revisión de sus fundamentos empíricos y metodológicos. *Revista Nebrija de Lingüística Aplicada* 22.
- Jimeno Panés, M. (2006). El español en Rumanía y Moldavia. *Enciclopedia de español en el mundo: anuario del Instituto Cervantes: 2006-2007* (pp. 304-305). Madrid: Instituto Cervantes.
- Jing, L. (2006). *Estudio de disponibilidad léxica de los estudiantes chinos de español como lengua extranjera*. Memoria de máster inédita. Madrid: Universidad de Alcalá.
- Jing, L. (2012). El estudio de disponibilidad léxica de los estudiantes chinos de español como lengua extranjera. *marcoELE*, 14.
- Jones, R. L. & Tschirner, E. (2005). *A Frequency Dictionary of German: Core Vocabulary for Learners*. Florence: Taylor and Francis
- Juilland, A. & Chang-Rodríguez, E. (1964). *Frequency Dictionary of Spanish Words*. La Haya: Mouton.
- Juilland, A. & Traversa, V. (1973). *Frequency Dictionary of Italian Words*. La Haya: Mouton.
- Juilland, A.; Brodin, D. & Davidocitch, C. (1990). *Frequency Dictionary of French Words*. La Haya/Paris: Mouton.
- Justicia, F. (1984). *Determinación y análisis de vocabulario del niño de 6 a 10 años en Andalucía Oriental*. Resumen de Tesis Doctoral. Granada: Universidad de Granada.
- Justicia, F. (1995). *El desarrollo del vocabulario. Diccionario de frecuencias. (niños de 6 a 13 años de Andalucía oriental)*. Granada: Universidad de Granada.
- Justo Hernández, H. (1987). *Disponibilidad léxica en colores*. Tesina inédita. México: Universidad Nacional Autónoma de México.
- Kaeding, F. W. (1898). *Haufigkeitwörterbuch der deutschen Sprache*. Berlín.
- Keniston, H. (1920). Common words in Spanish. *Hispania* 3, 85-118.
- Kilgarriff, A. (2009). *Putting the Corpus into the Dictionary*. En Ooi, V.B., Pakir, A., Talib, I.S. & Tan, P.K. (Eds.), *Perspectives in Lexicography* (pp. 239-248). Tel Aviv: Asia and Beyond.
- Kilgarriff, A. et al. (2014). The Sketch Engine: ten years on. *Lexicography*, 1-30.
- Kilgarriff, A., Rychlý, P., Smrž, P. & Tugwell, D. (2004a). The Sketch Engine. *Information Technology*, 105, 116- 127.
- Kilgarriff, A., Rychlý, P., Smrž, P. & Tugwell, D. (2004b). The Sketch Engine. En G. Williams & S. Vessier (Eds.), *Proceedings of the 11th EURALEX (European Association for Lexicography) International Congress (Euralex 2004), Lorient: France, 6-10 July 2004. Université de Bretagne Sud. Volume I* (pp.

- 105–116) (reeditado en P. W. Hanks (Ed.), *Lexicology: Critical concepts in Linguistics*, Routledge, 2007).
- Koike, K. (2001). *Colocaciones léxicas en el español actual: Estudio formal y léxico-semántico*. Madrid: Universidad de Alcalá y Universidad de Takoshoku.
- Kubarth, H. & Jódar Alamo, M. A. (2005). Las tics aplicadas a la elaboración de un vocabulario básico hispanoamericano. En M. A. Castillo Carballo (Coord.), *Las gramáticas y los diccionarios en la enseñanza del español como segunda lengua, deseo y realidad. Actas del XV Congreso Internacional de ASELE, Sevilla, 2004*. (pp. 491-498). Sevilla: ASELE.
- Lado, R. (1964). *Language Teaching. A Scientific Approach*. New York: McGraw-Hill, Inc.
- Lagrange, R. (1961). *Le français fondamental par l'action, langage et vocabulaire*. Paris: Didier.
- Lahuerta, J. & Pujol, M. (1996). El lexicón mental y la enseñanza del vocabulario. En C. Segoviano (Ed.), *La enseñanza del léxico español como lengua extranjera* (pp. 117-129). Madrid: Iberoamericana.
- Lara, L. F. (2006). *Curso de Lexicología*. México: El Colegio de México.
- Laufer, B. (1989). What percentage of text-lexis is essential for comprehension? En C. Lauren & M. Nordman (Eds.), *Special language: From humans thinking to thinking machines* (pp. 316-323). Clevedon: Multilingual Matters.
- Laufer, B. (1992). How much lexis is necessary for reading comprehension? En H. Bejoint & P. Arnaud (Eds.), *Vocabulary and applied linguistics* (pp. 126-132). London: Macmillan.
- Laufer, B. (1997a). The lexical plight in second language reading: Words you don't know, words you think you know and words you can't guess. En J. Coady & T. Huckin (Eds.), *Second language vocabulary acquisition: A rationale for pedagogy* (pp. 20-34). Cambridge: Cambridge University Press.
- Laufer, B. (1997b). What's in a word that makes it hard or easy: Some intralexical factors that affect the learning of words. En N. Schmitt & M. McCarthy (Eds.), *Vocabulary: Description, acquisition and pedagogy* (pp. 140-155). Cambridge: Cambridge University Press.
- Laufer, B. & Nation, I. S. P. (1995). Vocabulary Size and Use: Lexical Richness in L2 Written Production. *Applied Linguistics*, (16)3, 307-322.
- Laufer, B. & Nation, I. S. P. (2001). Passive vocabulary size and speed of meaning recognition: Are they related? En S. Foster-Cohen & A. Nizgorodcew (Eds.), *EUROSLA Yearbook 1* (pp.7-28). Amsterdam: Benjamins.
- Laufer, B. & Ravenhorst-Kalovsky, G. (2010). Lexical threshold revisited: Lexical text coverage, learners' vocabulary size and reading comprehension. *Reading in a Foreign Language* 22(1), 15-30.
- Laufer, B. & Sim, D. D. (1985). Measuring and explaining the threshold needed for English for Academic Purposes texts. *Foreign Language Annals* 18, 405-413.
- Leech, G. (1991). The state of the art in corpus linguistics. En K. Aijmer & B. Altenberg (Eds.), *English Corpus Linguistics* (pp. 8-29). London, Longman.
- Leech, G. (2011). Frequency, corpora and language learning. En F. Meunier, S. De Kock, G. Gilquin & M. Paquot (Eds.), *A taste for Corpora: In Honour of Sylviane Granger* (pp. 7-31). Amsterdam & Philadelphia: John Benjamins.
- Leech, G., Rayson, P. & Wilson, A. (2001). *Word frequencies in written and spoken English*. London: Longman.

- Levelt, W. J. M. (1999). Models of word production. *Trends in Cognitive Sciences*, 3, 223-232.
- Levelt, W. J. M. (2001). Spoken word production: A theory of lexical access. En *Proceedings of the National Academy of Sciences* 98 (23), (pp. 13464--13471).
- Levelt, W. J. M., A. Roelofs & A. S. Meyer (1999). A theory of lexical access in speech production. *Behavioraland Brain Sciences*, 22, 1-38.
- Lewis, M. (1993). *The lexical approach: The state of ELT and the way forward*. Hove, England: Language Teaching Publications.
- Lewis, M. (1997a). *Implementing the Lexical Approach*. Londres: Language Teaching Publications.
- Lewis, M. (1997b). Pedagogical implications of the lexical approach. En J. Coady & T. Huckling (Eds.), *Second language vocabulary acquisition: A rationale for pedagogy* (pp. 255-270). Cambridge: Cambridge University Press.
- Lewis, M. (2000). *Teaching collocations. Further developments in the Lexical Approach*. Londres: Language Teaching Publications.
- Lin, J. (2012) El estudio de la disponibilidad léxica de los estudiantes chinos de español como lengua extranjera. *Marcoele: Revista de Didáctica como lengua extranjera* 14.
- Littlewood, W. & Liu, N. F. (1996). *Hong Kong Students and their English*. Hong Kong: Macmillan.
- Llinares, G. (1990). Estudio del "umbral lingüístico" necesario para la comprensión de textos en inglés. En *Actas del II Congreso Nacional de Profesores de Lenguas Extranjeras con Fines Específicos* (pp. 139-143). Caracas: Universidad Simón Bolívar.
- Llinares, G., Leiva, B., Cartaya, N. & St. Louis, R. (2008). Acquisition of L2 vocabulary for effective reading: Testing teachers' classroom practice. *The Reading Matrix* 8(2), 55-69. Obtenido de: <http://www.readingmatrix.com/articles/llinares-leiva-cartayastlouis/article.pdf>
- Llopis Rodrigo, F. (2008). Disponibilitat léxica del valencià: classificació dels fenòmens de contacte lingüístic. En J. L. Blas Arroyo, M. Casanova Avalos, M. Velando Casanova & J. Vellón Lahoz (Coord.) *Discurso y sociedad II: Nuevas contribuciones al estudio de la lengua esn contexto social* (pp. 571-582). Castellón de la Plana: Universitat Jaume I.
- López Casado, M. (2013). *Disponibilidad léxica de los estudiantes de la Escuela Francesa de Ibiza (curso 2012-2013)*. Trabajo de grado. Ibiza: UNED
- López Chávez, J. (1992). Alcances panhispánicos del léxico disponible. *Lingüística*, 4, 26-124.
- López Chávez, J. (1993). *El léxico disponible de escolares mexicanos*. México: Editorial Alambra Mexicana.
- López Chávez, J. (1994). Comportamiento sintáctico de algunos verbos ordenados según su grado de disponibilidad léxica. *REALE* 1, 67-84.
- López Chávez, J. (1995). Léxico fundamental panhispánico: realidad o utopía. En *Actas del III Congreso Internacional sobre el Español de América, vol. II* (pp.1006-1014). Santiago de Chile: Universidad Católica de Chile.
- López Chávez, J. & Strassburger Frías, C. (1987) *Otro cálculo del índice de disponibilidad léxica*. México: Universidad Nacional Autónoma de México.

- López Chávez, J. & Strassburger Frías, C. (1991a). El diseño de una fórmula matemática para obtener un índice de disponibilidad léxica confiable. *Anuario de letras XXXVIII*, 227-251.
- López Chávez, J. & Strassburger Frías, C. (1991b). Un modelo más para el cálculo de disponibilidad léxica individual. En H. López Morales (Ed.), *La enseñanza del español como lengua materna* (pp. 91-112). Río Piedras: Universidad de Puerto Rico.
- López Chávez et al. (1993). Colección de cuadernos de *El léxico disponible de escolares mexicanos*: 1. *Léxico de preescolares mexicanos* (con Rosa M^a Mesa Canales). 2. *Léxico disponible de primer grado de primaria* (con M^a Trinidad Madrid Guillén). 3. *Léxico disponible de segundo grado de primaria* (con Martha Julián Peña). 4. *Léxico disponible de tercer grado de primaria* (con Rosala Bolfeta Montes de Oca). 5. *Léxico disponible de cuarto grado de primaria* (con Marcela Flores Cervantes). 6. *Léxico disponible de quinto grado de primaria* (con Lilia Castellanos Medina). 7. *Léxico disponible de sexto grado de primaria* (con Eva Núñez Alonso). México: Editorial Alhambra Mexicana.
- López Chávez, J. & Pérez Durán, M. A. (2014). Los estudios de la disponibilidad léxica en México desde el año 2000. *Tonos digital. Revista electrónica de Estudios Filológicos*, 26, Enero 2014, 14 pp.
- López García, Á. (2008). Disponibilidad léxica, gramatical y textual. *Lingüística española actual*, XXX/1, 65-86.
- López González, A. M. (2009). La disponibilidad léxica en las secciones bilingües de Lódz (Poznan, Polonia). En A. Barrientos Clavero et al. (Eds.), *El profesor de español LE / L2. Actas del XIX Congreso Internacional ASELE*. (pp. 567-582). Vol. 2. Cáceres: Universidad de Extremadura – ASELE.
- López González, A. M. (2010). La evaluación del desarrollo de la competencia léxica en L2 por medio de la disponibilidad léxica. *redELE: red electrónica de Didáctica de ELE*, 18.
- López González, A. M. (2014). Desarrollo de los estudios de disponibilidad léxica en español lengua extranjera (ELE). En N. M. Contreras Izquierdo (Ed. Lit.), *Actas del XIV Congreso Internacional de ASELE (Jaen 2013). La enseñanza del Español como LE/L2 en el siglo XXI* (pp. 397-360). Jaen: ASELE.
- López González, A. M. (2016). Las asociaciones léxicas en el léxico disponible en lengua materna y en lengua extranjera. *Tonos digital: Revista electrónica de estudios filológicos*, 31.
- López Jiménez, M. D. (2013). Colocaciones léxicas: estudio comparativo de libros de texto de inglés y de español como L2. En *Odisea*, nº 14, 101-120.
- López Meirama, B. (2008). *Léxico disponible en el español de Galicia*. Santiago de Compostela: Universidade de Santiago de Compostela, Colección Lalia (Series Maior, nº 22).
- López Meirama, B. (Ed.) (2011). *Estudios sobre disponibilidad léxica en el español de Galicia*. Santiago de Compostela: Servizo de publicacións e intercambio científico Universidade de Santiago de Compostela.
- López Meirama, B. & Álvarez de la Granja, M. (2014). *Léxico dispoñible en galego*. Anexo 71 de *Verba*. Santiago de Compostela: Servizo de Publicacións e Intercambio Científico da USC.

- López Morales, H. (1973). *Disponibilidad léxica en escolares de San Juan*. Puerto Rico: MS.
- López Morales, H. (1978). Frecuencia léxica, disponibilidad y programación curricular. En H. López Morales (ed.), *Actas del I Seminario Internacional sobre 'Aportes de la lingüística a la enseñanza del español como lengua materna, BAPLE'*, 6 (1), (pp. 73-86).
- López Morales, H. (1979). Disponibilidad léxica y estratificación socioeconómica. En *Dialectología y sociolingüística. Temas puertorriqueños* (pp. 173- 181). Madrid: Hispanova de Ediciones.
- López Morales, H. (1990). Los estudios de disponibilidad léxica: pasado y presente, I. En V. M. Sánchez Corrales (Ed.), *Memoria del IV Congreso costarricense de Filología, Lingüística y Literatura* (pp. 29-40). San José: Universidad de Costa Rica.
- López Morales, H. (1993). En torno al aprendizaje del léxico. Bases psicolingüísticas de la planificación curricular. En S. Montesa & A. Garrido (Eds.), *Actas del tercer Congreso Nacional de ASELE: el español como lengua extranjera: de la teoría al aula: Málaga, del 10 al 12 de octubre de 1991* (pp. 9-22). Málaga: ASELE.
- López Morales, H. (1995-6). Los estudios de disponibilidad léxica: pasado y presente. *Boletín de Filología de la Universidad de Chile*, 35, 245-259.
- López Morales, H. (1996). Los estudios de disponibilidad léxica: pasado y presente. En *Homenaje a Kodolfo Oroz Boletín de Filología de la Universidad de Chile XXXV* (pp. 245-259).
- López Morales, H. (1999a). Anglicismos en el léxico disponible de Puerto Rico. En L. A. Ortiz López, *El caribe hispánico: perspectivas lingüísticas actuales. Homenaje a Manuel Álvarez Nazario* (pp. 147-70). Vervuert Iberoamericana,
- López Morales, H. (1999b). *Léxico disponible de Puerto Rico*. Madrid: Arco Libros.
- López Morales, H. (2001). Tendencias del léxico disponible. Hispanoamérica. *Revista de Occidente*, 240, 5-24.
- López Morales, H. (2014). Lexical Availability Studies. En Jiménez Catalán, R. M^a (Ed.). *Lexical Availability in English and Spanish as a Second Language*. New York: Springer.
- López Morales, H. & García Marcos, F. (1995). Disponibilidad léxica en Andalucía. Proyecto de investigación. *Revista de Estudios de Adquisición de la lengua española (REALE)*, números especiales, 67-76.
- López Pérez, M. V. (2005). Cuestiones relativas al léxico de la lengua de instrucción en el marco de la enseñanza del Español/L2 a inmigrantes. En M. A. Castillo Carballo et al. (Eds.), *Actas del XV Congreso Internacional de ASELE. Las gramáticas y los diccionarios en la enseñanza del español como segunda lengua: deseo y realidad* (pp. 534-541). Sevilla: Universidad de Sevilla, Secretariado de Publicaciones.
- López Pérez, M. V. (2013). Criterios de selección de voces para la enseñanza del español como L2 en contextos escolares. *Revista de Educación, Extraordinario*, 154-176.
- López-Mezquita Molina, M^a T. (2007). *La evaluación de la competencia léxica: test de vocabulario. Su fiabilidad y validez. Colección de Investigación*, 177. Madrid: Ministerio de Educación, Cultura y Deporte, Centro de

- Investigación y Documentación Educativa. (Tesis Univ. Granada. Departamento de Filologías Inglesa y Alemana, leída el 11 de febrero de 2005).
- Lorán, R. (1987). *La tasa de sustitución y el índice de disponibilidad léxica*. Mayagüez: Universidad de Puerto Rico, Departamento de Matemáticas, MS.
- Lorán, R. & López Morales, H. (1983). *Nouveau calcul de l'indice de disponibilité*. Ms.
- Luque Durán, J. de D. (1998). Introducción a la tipología léxica. En B. Gallardo Paúls (Ed.), *Temas de Lingüística y Gramática* (pp. 122-145). Valencia: Universitat de València.
- Luque Durán, J. de D. & Manjón Pozas, F. J. (1998). Colocaciones léxicas: cuestión lingüística o estilística. En *Teoría y práctica de la lexicología: V Jornadas internacionales sobre estudio y enseñanza del léxico, Granada 1997* (pp. 15-24). Granada: Médoto Ediciones.
- Mackey, W. F. (1971). *Le vocabulaire disponible du Français*, 2 vols. Paris: Didier.
- Madison, M. (2013). *La selección del vocabulario: las palabras de alta frecuencia en español. Análisis del vocabulario de dos manuales de ELE*. Tesis de máster inédita defendida en la Universidad de Tartu (Estonia).
- Magnúsdóttir, S. (2012). *Disponibilidad léxica en alumnos de español como lengua extranjera. Estudio sobre el léxico disponible en alumnos de ELE en la secundaria en Islandia*. Memoria de máster inédita. Háskóli Íslands: Sigillum Universitatis Islandiae.
- Malaca Casteiro, J. (1984). *Português Fundamental, Vocabulário e Gramática, tomo 1, Vocabulário*. Lisboa: INIC, CLUL.
- Marconi, D. (2000). *La competencia léxica*. Madrid: Antonio Machado Libros.
- Márquez Villegas, L. (1975). *Vocabulario del español hablado*. Madrid: Sociedad General Española de Librería (SGEL).
- Martín Martín, S. (1999). La revisión del concepto de vocabulario en la gramática de E/LE. En L. Miquel & N. Sans (Coord.), *Didáctica del español como lengua extranjera (4)* (pp. 157-163). Madrid: Fundación Actilibre-Colección Expolingua.
- Martín Peris, E. et al. (2008). *Diccionario de términos clave de ELE*. Madrid: SGEL.
- Martínez J. A. & García, E. (2004). *Frecuencias del castellano escrito en niños de 6 a 12 años*. Salamanca: Servicio de Publicaciones de la Universidad Pontificia de Salamanca.
- Martínez Olmos, E. (2007). *Disponibilidad léxica en las comunidades de habla alicantinas*. Alicante: Universidad de Alicante. Tesis doctoral.
- Martínez Olmos, E. (2015). *El léxico disponible de las comunidades de habla alicantinas. Estudio sociolingüístico y diccionarios*. Alicante: Publicacions de la Universitat d'Alacant.
- Martínez Segura, G. E. (2012). *Disponibilidad léxica de los/las estudiantes de Bachillerato Académico, Sección "A" del Instituto 1º de mayo de 1954. Una propuesta para la enseñanza del vocabulario*. Tesis de máster. Tegucigalpa: Universidad Pedagógica Nacional Francisco Morazán
- Mas Álvarez, I. & Santos Palmou, X. (2014a). Avances y retos en la aplicación de la metodología de los estudios sobre disponibilidad léxica a la enseñanza de ELE. En *Actas del II Congreso Internacional Nebrija en Lingüística aplicada a*

- la Enseñanza de Lenguas, Madrid, 26-28 de junio de 2014*, (pp. 391-403). Madrid: Universidad Nebrija.
- Mas Álvarez, I. & Santos Palmou, X. (2014b). Disponibilidad léxica y contacto del español con otras lenguas. En J. M. Santos Rovira (Ed.), *Fronteras y diálogos. El español y otras lenguas* (pp. 25-38). Lugo: Axac.
- Mateo García, M. V. (1998). *Disponibilidad léxica en el C.O.U. almeriense: estudio de estratificación social*. Almería: Universidad de Almería.
- Mateus Ferro, G. E. & William Santiago, Á. (2006). Disponibilidad léxica en estudiantas bogotanos. *Folios*, 3-26. Bogotá: Universidad Pedagógica Nacional.
- Mauger, G. (1963). *Contes et récits en français facile*. Paris: Hachette.
- Mauger, G. & Gougenheim, G. (1955). *Le Français Élémentaire*. Paris: Hachette.
- McCarthy, M. J. (1984). A new look at vocabulary. *EFL, Applied Linguistics*, 5, 12-22.
- McCarthy, M. J. (1990). *Vocabulary*. Oxford: Oxford University Press.
- McCarthy, M. J. (1999). What constitutes a basic vocabulary for spoken communication? *Studies in English Language and Linguistics*, 1, 233-249.
- McCarthy, M. J. (2001). *Issues in Applied Linguistics*. Cambridge: Cambridge University Press.
- McCarthy, M. J. & Carter, R. (1997). Written and spoken vocabulary. En N. Schmitt & M. McCarthy (Eds.), *Description, Acquisition and Pedagogy* (pp. 20-39). Cambridge: Cambridge University Press.
- McEnery, T. & Wilson, A. (1972). *Corpus Linguistics*. Edinburgh: Edinburgh University Press.
- Medina Arejita, E. T. (2009). *Las nociones específicas del PCIC y la disponibilidad léxica como instrumento de selección del vocabulario. El caso de 43 estudiantes alemanes de español en Berlín*. Memoria de Máster. Instituto Cervantes - Universidad Menéndez Pelayo.
- Mel'čuk, I., A. Clas & Polguère, A. (1995). *Introduction à la lexicologie explicative et combinatoire*. Louvain-la-Neuve: Duculot.
- Mena Osorio, M. (1986). *Disponibilidad léxica infantil en tres niveles de enseñanza básica*. Tesis de Máster. Concepción, Chile: Universidad de Concepción.
- Mesa Betancor, E. (2011). *La enseñanza del léxico de la banca española: el léxico disponible y los manuales de ENE*. Memoria de Máster. Alicante: Universitat d'Alacant.
- Mesa Betancor, E. (2012). La enseñanza del léxico de la banca española: el léxico disponible y los manuales ENE. *redELE: red electrónica de Didáctica de ELE, Biblioteca 2012*, 13. Madrid: Ministerio de Educación, Cultura y Deporte.
- Mesa Canales, R. M^a (1989). *Disponibilidad léxica en preescolares*. México, D.F.: Universidad Autónoma de México.
- Mezynski, K. (1983). Issues concerning the acquisition of knowledge: Effects of vocabulary training on reading comprehension. *Review of Educational Research*, 53, 253-279.
- Michéa, R. (1953). Mots fréquents et mots disponibles: un aspect nouveau de la statistique du langage. *Les Langues Modernes*, 47, 338-344.
- Miller, G. A. (1986). Dictionaries in the Mind. *Language and Cognitive Processes*, (1)3, 171-185.

- Miyazaki H. (Director) (1988). *Mi vecino Totoro*. [Cinta cinematográfica]. País: Japón. Studio Ghibli.
- Molero Huertas, J. (1999). *El vocabulario usual de nuestros alumnos*. Granada: Consejería de Educación y Ciencia de la Junta de Andalucía.
- Molero, C. (2013). La adquisición del léxico en el aula de E/LE en contextos multilingües: la teoría de las colocaciones léxicas como solución a las interferencias léxicas. En B. Blecua, S. Borrell, B. Crous & F. Sierra (Eds.), *Actas del XXIII Congreso Internacional de ASELE: Plurilingüismo y enseñanza de ELE en contextos multiculturales. Girona, 2012* (pp. 605-612). Girona: ASELE.
- Molero, C. M., & Salazar, D. (2012). Ejercicios prácticos para trabajar las colocaciones léxicas en el aula de E/LE. En *Actas del I Encuentro Internacional de profesores de ELE: El aprendizaje y la enseñanza de español en contextos multiculturales y multilingües*. Bruselas: Instituto Cervantes de Bruselas y Universidad de Oxford, p. 74-87.
- Molero, C. M. & Salazar, D. (2013). Análisis contrastivo, criterios de selección y didáctica de las colocaciones léxicas en el aula de español. En B. Pregelj (Ed.), *Estudios Hispánicos en el siglo XXI*. Belgrado: Universidad de Belgrado.
- Morales, A. (1986). *Léxico básico del español de Puerto Rico*. San Juan de Puerto Rico: Academia Puertorriqueña de la Lengua Española.
- Moreno Fernández, F. (2007). Anglicismos en el léxico disponible de los adolescentes hispanos de Chicago. En K. Potowski & R. Cameron (Eds.), *Spanish in Contact* (pp. 41-58). Amsterdam: John Benjamins.
- Moreno Fernández, F. (2009). *Principios de sociolingüística y sociología del lenguaje*. Barcelona: Ariel.
- Moreno Fernández, F. (2012). Disponibilidad léxica: cuestiones metodológicas. A propósito de disponibilidad léxica de los estudiantes hispanos de Redwood City, CA. *Revista Nebrija de Lingüística aplicada a la enseñanza de Lenguas*, 6(2), 51-59.
- Moreno Fernández, F.; Moreno Fernández J. E. & García de las Heras, A. J. (1995). Cálculo de disponibilidad léxica. El programa LexiDisp. *Lingüística*, nº 7, 243-250.
- Moreno Ortiz, A. (2000). Diseño e implementación de un lexicón computacional para lexicografía y traducción automática. *Estudios de Lingüística del Español*, Volumen 9.
- Morgan, B. Q. (1928). *German Frequency Word Book*. New York: Macmillan.
- Morris, I. (1964). *The art of teaching English as a living language*. New York: Macmillan.
- Müller, Ch. (1964). *Essai de statistique lexicale*. Paris: Bibliothèque française et romaine.
- Müller, Ch. (1973). *Estadística Lingüística*. Trad. Arturo Amites. Madrid: Editorial Gredos.
- Müller, F. M. (1862). *Lectures on the Science of Language: Delivered at the Royal Institution of Great Britain in April, May & June 1861*. London: Longmans, Green & Roberts.
- Muñoz Hernández, I. O. (2005). *Disponibilidad léxica en alumnos de enseñanza media de Almería*. Tesis doctoral. Almería: Universidad de Almería.
- Muñoz, B. (2005). *Mis primeros días en Secundaria*. Madrid: SGEL

- Murillo Rojas, M. (1993). Disponibilidad léxica en los preescolares: estudio de cinco campos semánticos. *Káñina* 17(2), 117-127. Universidad de Costa Rica.
- Murillo Rojas, M. (1994). Comidas y bebidas: estudio de disponibilidad léxica en preescolares. *Káñina* 18(2), 117-133. Universidad de Costa Rica.
- Nassaji, H. (2003a). Higher-level and lower-level text processing skills in advanced ESL reading comprehension. *The Modern Language Journal* 87(II), 261-276.
- Nassaji, H. (2003b). L2 Vocabulary learning from context: Strategies, knowledge, sources and their relationship with success in L2 lexical inferencing. *TESOL Quarterly*, 37(4), 645-670.
- Nation, I. S. P. (1983). Testing and teaching vocabulary. *Guidelines* 5, 1, 12-25.
- Nation, I. S. P. (1990). *Teaching and learning vocabulary*. Rowley, MA: Newbury House.
- Nation, I. S. P. (2001). *Learning Vocabulary in Another Language*. Cambridge: Cambridge University Press.
- Nation, I. S. P. (2006). How large a vocabulary is needed for reading and listening? *The Canadian Modern Language Review* 63, 59-82.
- Nation, I. S. P. & Waring, R. (1997). Vocabulary size, text coverage, and word lists. En N. Schmitt & M. McCarthy (Eds.), *Vocabulary: Description, acquisition and pedagogy* (pp. 6-19). Cambridge: Cambridge University Press.
- Nattinger, J. R. & De Carrico, J. S. (1992). *Lexical Phrases and Language Teaching*. Oxford University Press
- Navarro Marrero, Y. (2009). *Terminología especializada en el área de Fisioterapia: acercamiento desde la metodología de la disponibilidad léxica específica*. Comunicación presentada en la Universidad Autónoma de Barcelona. [Disponible en <http://filcat.uab.cat/clt/XXIVAJL/Interlinguistica/Encuentro%20XXIV/NavarroMarreroREVF.pdf>].
- Navarro Marrero, Y. (2011). *Léxico disponible específico de los estudiantes de Fisioterapia de la Universidad de Salamanca: primeros resultados*. Trabajo de grado. Salamanca: Universidad de Salamanca.
- Navarro, C. (2007). Fraseología contrastiva del español y el italiano: análisis de un corpus bilingüe. *Tonos: Revista Electrónica de Estudios Filológicos*, 13.
- Njock, P. E. (1979). *L'univers familier de l'enfant africain*. Quebec: CIRB, Université Laval.
- Núñez Romero, L. (2008a). Presencia de elementos culturales en el léxico disponible en E/LE de una muestra de estudiantes arabófonos. En J. Villatoro (Ed.), *Actas del Segundo Congreso Virtual sobre E/LE. La red como espacio de colaboración*, vol. I, (pp. 267-279).
- Núñez Romero, L. (2008b). Análisis de los contenidos léxicos en el aula de E/LE: propuestas para la selección de vocabulario. En J. Villatoro (Ed.), *Actas del Segundo Congreso Virtual sobre E/LE. La red como espacio de colaboración*, vol. I, (pp. 280-289).
- O'Malley, J. & Chamot, A. U. (1990). *Learning strategies in second language acquisition*. Cambridge: Cambridge University Press.
- Oehler, H. & Khoury, R. G. (1970). *Grundwortschatz Deutsch*. Stuttgart: Klett.

- Ogden, C. K. (1930). *Basic English: a general introduction with rules and grammar*. London: Kegan Paul. (1ª edición).
- Ortega Olivares, J. (1985). Algunas observaciones sobre los diccionarios de frecuencias. En R. Fente & F. Serrano (Eds.), *Actas del II Congreso Nacional de Lingüística Aplicada. Las lenguas de España. Adquisición, aprendizaje y uso* (pp. 429-442). Madrid: AESLA.
- Ostyn, P. y Godin, P. (1985). RALEX: An alternative approach to language teaching. *Modern Language Journal* 69, 346-353.
- Ostyn, P., Vandecasteele, M., Deville, G. & Kelly, P. (1987). Towards an optimal programme of FL vocabulary acquisition. En A. M. Cornu, J. Vanparijs, M. Delaheye & L. Baten (Eds.), *Beads or bracelet? How do we approach LSP?* (pp. 292-305). Oxford: Oxford University Press.
- Oxford, R. L. (1990). *Language Learning Strategies: What Every Teacher Should Know*. Boston: Heinle & Heinle.
- Palapanidi, K. (2012). La aplicación de la disponibilidad léxica a la didáctica del léxico de LE. *Revista Nebrija de Lingüística Aplicada a la Enseñanza de Lenguas*, 11 (6), 69-77.
- Palmer, H. E. & Hornby, A. S. (1937). *Thousand-word English: what it is and what can be done with it*. London: G.G. Harrap & Co, Ltd.
- Paredes García, F. (2012). Desarrollos teóricos y metodológicos recientes de los estudios de disponibilidad léxica. *Revista Nebrija de Lingüística Aplicada*, 11 (6), 78-100.
- Paredes García, F. (2014). A vueltas con la selección de 'centros de interés' en los estudios de disponibilidad léxica: para una propuesta renovadora a propósito de la disponibilidad léxica en ELE. *Revista Nebrija de Lingüística Aplicada a la Enseñanza de Lenguas*, 8(4).
- Paredes García (2015). Disponibilidad léxica y enseñanza de ELE: el léxico disponible como fuente curricular y como recurso en el aula. *Linred: lingüística en la Red*, nº 13.
- Pastor Millán, M. A. & Sánchez García, F. J. (2008). *El léxico disponible de Granada y su provincia*. Granada: Universidad de Granada.
- Patterson, W. & Urrutibéheity, H. (1975). *The lexical structure of Spanish*. The Hague: Mouton.
- Pedroni Torres, T. (2015). *El léxico disponible de los profesores de E/LE en formación en la ciudad de São Paulo (Brasil)*. Tesis doctoral defendida en el curso 2014-2015 en la Universidad de Salamanca.
- Pérez Basanta, C. (1999). La enseñanza del vocabulario desde una perspectiva lingüística y pedagógica. En S. Salaberri (Ed.), *Lingüística Aplicada a las Lenguas Extranjeras* (pp. 262-307). Almería: Universidad de Almería.
- Pérez Durán, M. A. (2015). Análisis contrastivo de los centros de interés con mayor producción verbal en profesores de Tlaxcala y en alumnos de Ayamonte (Huelva), España. *Revista Electrónica del Lenguaje*, nº 2.
- Pérez Serrano, M. (2009). *Estudio de disponibilidad léxica en estudiantes de ELE en los centros de interés "Medios de transporte" y "Profesiones y oficios"*. Memoria de máster. Instituto Cervantes, UIMP.
- Peters, P. (2004). *The Cambridge Guide to English Usage*. Cambridge: Cambridge University Press.

- Pfeffer, A. (1964). *Grunddeutsch: Basic (Spoken) German Word List*. Englewood Cliffs, New Jersey: Prentice Hall, INC.
- Picoche, J. (1992). *Précis de lexicologie française. L'étude et l'enseignement du vocabulaire*. Paris: Nathan Université, (1^a ed.: 1977).
- Picoche, J. (1993). *Didactique du vocabulaire français*. Paris: Nathan Université.
- Picoche, J. (1994). *Précis de lexicologie française. L'étude et l'enseignement du vocabulaire*. Nueva edición revisada y actualizada. Nathan: París.
- Picoche, J. (1999). Dialogue autour de l'enseignement du vocabulaire. *Etudes de Linguistique Appliquée (ELA)*, n^o 116, 421-433.
- Prado Aragonés, J. & Galloso Camacho, M^a V. (2005). *Léxico disponible de Huelva: nivel preuniversitario*. Huelva: Servicio de Publicaciones de la Universidad de Huelva.
- Prado Aragonés, J. & Galloso Camacho, M^a V. (2006a). *Diccionario del Léxico Disponible de alumnos de 2^o curso de Bachillerato de Huelva*. Huelva: Universidad de Huelva.
- Prado Aragonés, J. & Galloso Camacho, M^a V. (2006b). *Diccionario del Léxico Disponible de alumnos de 4^o curso de Enseñanza Secundaria Obligatoria de Huelva*. Huelva: Universidad de Huelva.
- Prado Aragonés, J. & Galloso Camacho, M^a V. (2006c). *Diccionario del Léxico Disponible de alumnos de 6^o curso de Educación Primaria de Huelva*. Huelva: Universidad de Huelva.
- Prado Aragonés, J. & Galloso Camacho, M^a V. (2014). *El léxico disponible de Extremadura y comparación con el de Andalucía*. Huelva: Servicio de Publicaciones de la Universidad de Huelva.
- Prado Aragonés, J., Galloso Camacho, M^a V. & Célio Conceição, M. (2010). *La disponibilidad léxica en situaciones de contacto de lenguas en las zonas limítrofes de Andalucía y Extremadura (España) y Algarve y Alentejo (Portugal)*. Huelva: Universidad de Huelva.
- Prieto, S.; Mosqueira, E. & Vázquez, N. (2009). Córpora y enseñanza de lenguas: se buscan colocaciones. En P. Cantos Gómez & A. Sánchez Pérez (Eds.), *A Survey of Corpus-based Research. Panorama de investigaciones basadas en corpus* (pp. 366-373). Murcia: AELINCO.
- Pulido, D. (2007). The relationship between text comprehension and second language incidental vocabulary acquisition: A matter of topic familiarity? *Language Learning*, 57 (Supplement 1), 155-199.
- Pulido, D. (2009). Vocabulary processing and acquisition through reading: Evidence of the rich getting richer. En Z. Hang & N. Anderson (Eds.), *Second language reading research and instruction: Crossing the boundaries* (pp. 65-81). Michigan: The University of Michigan Press.
- Pulido, D. & Hambrick, D. (2008). The virtuous circle: Modeling individual differences in L2 reading and vocabulary development. *Reading in a Foreign Language*, Volume 20, Number 2, October. [Publicación en línea]
- Quillian, M. R. (1968). Semantic Memory. En M. Minsky (Ed.), *Semantic Information Processing* (pp. 27-70). Cambridge, Mass: The MIT Press.
- Real Academia Española. *Diccionario de la lengua española (versión en línea)*.
- Redston, C. & Cunningham, G. (2005). *Face2Face Elementary Student's Book (F2F henceforth)*. A2 level. Cambridge: Cambridge University Press

- Regla Cruz Ventura, S. de (2016). *La disponibilidad léxica de hablantes canarios adultos y comparación con los corpus juveniles*. Tesis doctoral defendida en el curso 2015-2016 en la Universidad de Las Palmas de Gran Canaria.
- Riazi, A.M. & Alvari, A. (2004). Strategy activation in learning English words. *Academic Exchange Quarterly*, 6 (2), 199-203.
- Richards, J. & Rogers, T.S. (2011). *Approaches and Methods in Language Teaching*. New York: Cambridge University Press
- Richards, J. C., Platt, J., & Platt, H. (1992). *Language teaching and applied linguistics*. London: Longman.
- Rivenc, P. (1967). État actuel des enquêtes sur les langues parlées et les langues de spécialité. *Actes du 1^o congrès international de Linguistique Appliquée (Nancy 1964)* (pp. 153-176). Conseil de l'Europe, AIDELA, (Les Langues vivantes en Europe).
- Robinson, P. (1989). Procesural vocabulary and language learning. *Journal of Pragmatics* 13, 523-546.
- Rodríguez Bou, I. (1952). *Recuento de vocabulario español*, I. Río Piedras (Puerto Rico): Universidad de Puerto Rico.
- Rodríguez Muñoz, F. J. & Muñoz Hernández, I. O. (2009). De la disponibilidad a la didáctica léxica. *Tejuelo*, Vol. 4, 8-18.
- Rojo Sastre, A. J. (1979). *Bases lexicológicas en una metodología audiovisual de la lengua española para extranjeros*. Madrid: Universidad Complutense. [Tesis Doctoral dirigida por Víctor García Hoz. Texto manuscrito e inédito.]
- Rojo Sastre, J. A., Rivenc P. & Ferrer, A. (1968-1969). *Vida y diálogos de España. 1er grado. Metodo audiovisual*, Paris: Didier. Hay 4 volúmenes: V.1 Libro del profesor, V. 2. Textos de lectura. Ortografía del español, preparación para el dictado, V. 3. Lista alfabética de formas empleadas, V. 4. Libro de imágenes.
- Rojo, G. (2001). La explotación de la Base de datos sintácticos del español actual (BDS). En J. De Kock (Ed.), *Gramática Española, Enseñanza e Investigación. I. 7. Lingüística con corpus. Catorce aplicaciones sobre el español* (pp. 255-286). Salamanca: Universidad de Salamanca.
- Rojo, G. (2005-2006). Informática y Lingüística: Las lenguas en la sociedad del conocimiento. *RedIRIS: boletín de la Red Nacional de I+D RedIRIS*, N^o 74-75, 1-8.
- Rojo, G. (2005). Lingüística basada en el análisis de corpus y enseñanzas de lenguas. En J. M. Oro Cabanas, J. D. Anderson & J. Varela Zapata (Coord.), *Lingüística aplicada al aprendizaje de lenguas* (pp. 13-34). Santiago de Compostela: USC, Servicio de publicaciones.
- Rojo, G. (2008). Lingüística de corpus y lingüística del español. Ponencia plenaria en el *XV Congreso de la Asociación de Lingüística y Filología de América Latina (Montevideo, 18-21 de agosto de 2008)*. Publicación en red en <http://www.narr.de/download/bachelorwissen/spanischesprachwissenschaft/rojo.pdf>
- Rojo, G. (2009). Sobre la construcción de diccionarios basados en corpus. *Tradumática* 7. Disponible en:
- Rojo, G. (2010). Sobre codificación y explotación de corpus textuales: Otra comparación del Corpus del español con el CORDE y el CREA. *Lingüística*, N^o 24, 11-50.

- Rojo, G. (2015). Lingüística de corpus e investigación lingüística: El Corpus del español del siglo XXI. Presentación en power point disponible en: http://gramatica.usc.es/~grojo/Presentaciones/lgca_corpus_e_investigacion_lgca.pdf
- Rojo, G. (2016). Los corpus textuales del español. En Gutiérrez-Rexach, J. (Ed.), *Enciclopedia lingüística hispánica* (pp. 285-296). Oxon: Routledge.
- Rojo, G. & Sánchez, M. (2010). *El español en la red*. Madrid/Barcelona: Fundación Telefónica/Ariel.
- Rojo, G. & Palacios, I. (Dir.). (2014). *Corpus de Aprendices de Español (CAES)*. Instituto Cervantes. (Versión:1.0–octubre 2014). Disponible en: <http://galvan.usc.es/caes>.
- Román, B. (1985). *Disponibilidad léxica en estudiantes de Dorado*. Tesis inédita. Río Piedras: Universidad de Puerto Rico.
- Rubio Sánchez, R. (2016). Estudio de disponibilidad léxica en 60 aprendices portugueses. En N. Domínguez García, C. Fernández Juncal & J. L. García Alonso (Dir.), *Innovación y desarrollo en español como lengua extranjera* (pp. 107-128). Salamanca: Universidad de Salamanca.
- Rubio Sánchez, R. (2015). Estudio de disponibilidad léxica en aprendices italianos de español: análisis cualitativo. *e-AESLA 2015*.
- Ruiz Basto, A. (1987). *Disponibilidad léxica de los alumnos de primer ingreso en el Colegio de Ciencias y Humanidades Plantel Naucalpan*. Memoria de Licenciatura. México, D.F.: Universidad Autónoma de México.
- Salazar García, V. (1994). Aprendizaje del léxico en un currículum centrado en el alumno. En L. Miquel & N. Sans (Coors.), *Didáctica del español como lengua extranjera 2* (pp. 165-204). Madrid: Colección Expolingua, Fundación Actilibre.
- Salazar García, V. (1996). Una propuesta para el análisis de necesidades léxicas en el aula de español. En C. Segoviano (Coord.), *La enseñanza del léxico como lengua extranjera: homenaje a Anton e Inge Bermmerlein* (pp. 13-21). Iberoamericana: Vervuert.
- Salazar García, V. (2004). Acercamiento crítico a la selección objetiva de contenidos léxicos en la enseñanza de E/LE1. *Estudios de Lingüística*. 18. Alicante: Universidad, 243-273.
- Salcedo, P., Ferreira, A. & del Valle, M. (2014). *La disponibilidad léxica matemática en estudiantes de enseñanza media de la ciudad de Concepción de Chile: Investigación Lexicométrica*. Chile: Universidad de Concepción.
- Samper Hernández, M. (2001). Dificultades de los estudios de disponibilidad léxica en ELE: los criterios de selección. En J. A. Bartol, S. Crespo, C. Fernández Juncal, C. Pensado, E. Prieto & N. Sánchez (Eds.), *Nuevas aportaciones al estudio de la lengua española. Investigaciones filológicas* (pp. 277-286). Salamanca: Luso-Española de Ediciones.
- Samper Hernández, M. (2002). *Disponibilidad léxica en alumnos de español como lengua extranjera*. Colección Monografías nº 4. Málaga: ASELE.
- Samper Padilla, J. A. (1998a). Criterios de edición del léxico disponible: sugerencias. *Lingüística*, nº 10, 311-333.
- Samper Padilla, J. A. (Dir.) (1998b). *Léxico del habla culta de Las Palmas de Gran Canaria*. Las Palmas de Gran Canaria: Cabildo Insular de Gran Canaria.

- Samper Padilla, J. A. (1999). Léxico disponible y variación dialectal: datos de Puerto Rico y Gran Canaria. En E. Forastieri Braschi, J. Cardona, H. López Morales & A. Morales de Walters (Coord.), *Estudios de lingüística hispánica: homenaje a María Vaquero* (pp. 550-573).
- Samper Padilla, J. A. (2000). El léxico de la norma popular de Las Palmas de Gran Canaria. Comparación con el de la norma culta. En C. Corrales Y D. Corbella (Coords.), *Estudios de dialectología dedicados a Manuel Alvar* (pp. 39-74). La Laguna: Instituto de Estudios Canarios.
- Samper Padilla, J. A. (2003). El proyecto panhispánico de disponibilidad léxica: logros y estado actual. En C. Varo Varo & M. Casas Gómez (Coord.), *VII Jornadas de Lingüística* (pp. 193-226).
- Samper Padilla, J. A. (2006). Disponibilidad léxica y sociolingüística. En J. L. Blas Arroyo, M. Velando Casanova & M. Casanova Avalos (Coord.), *Discurso y sociedad: contribuciones al estudio de la lengua en contexto social* (pp. 99-120).
- Samper Padilla, J. A. (2009). El proyecto panhispánico de disponibilidad léxica: datos andaluces. En M. V. Camacho Taboada, J. J. Rodríguez Toro & J. J. Santana Marrero (Coord.), *Estudios de lengua española: descripción, variación y uso: homenaje a Humberto López Morales* (pp. 595-618).
- Samper Padilla, J. A. & Hernández Cabrera, C. E. (1997). El estudio de la disponibilidad léxica en Gran Canaria: datos iniciales y variación sociolingüística. En J. Dorta Luis & M. V. Almeida Suárez (Coord.), *Contribuciones al estudio de la lingüística hispánica: homenaje al profesor Ramón Trujillo* (pp. 229-240), Vol. 2.
- Samper Padilla, J. A. & Hernández Cabrera, C. E. (2003). Los dialectalismos en el léxico disponible de Gran Canaria. Análisis de un centro de interés. En F. Moreno Fernández, J.A. Samper Padilla, M. Vaquero, M. L. Gutiérrez Araus, C. Hernández Alonso & F. Gimeno Menéndez (Coord.), *Lengua, variación y contexto: estudios dedicados a Humberto López Morales* (pp. 339-353), Vol. 1.
- Samper Padilla, J. A. & Hernández Cabrera, C. E. (2009). Aportaciones canarias al estudio de la disponibilidad léxica. En D. Corbella Díaz, J. Dorta (eds.), *La investigación dialectológica en la actualidad* (pp. 53-74). Santa Cruz de Tenerife: Agencia Canaria de Investigación, Innovación y Sociedad de la Información del Gobierno de Canarias.
- Samper Padilla, J. A. & Samper Hernández, M. (2006). Aportaciones recientes a los estudios de disponibilidad léxica. *Lynx: Panorámica de estudios lingüísticos*, nº 5, 5-95.
- Samper Padilla, J. A. & Samper Hernández, M. (2007). El proyecto panhispánico de disponibilidad léxica y los cotejos dialectales. *Lingüística*, nº 19, 88-116.
- Samper Padilla, J. A., Bellón Hernández, J. J. & Samper Hernández, M. (2003). El proyecto de estudio de la disponibilidad léxica en español. En G. Wotjak (Coord.), *Pautas y pistas en el estudio del léxico hispano(americano)* (pp. 27-140). Frankfurt /Madrid: Iberoamericana Verbuert, Lingüística Iberoamericana, 19.
- Sánchez Corrales, V. M. & Murillo Rojas, M. (2006). *Disponibilidad léxica de los niños preescolares costarricenses. Seires ELEXIHCÓS*. Ciudad Universitaria Rodrigo Facio: Editorial Universidad de Costa Rica.

- Sánchez Gómez, C. (2003). *Disponibilidad léxica en alumnos de español como lengua extranjera*. Tesina inédita. Huelva: Universidad.
- Sánchez Gómez, C. (2006). Naturaleza gramatical del léxico disponible en informantes de español como lengua extranjera”, *Interlingüística*, 16, 1-10.
- Sánchez-Saus Laserna, M. (2009). La variable “nivel de español” en el léxico disponible de los estudiantes de español como lengua extranjera. *Pragmalingüística*, 17, 140-153.
- Sánchez-Saus Laserna, M. (2010). La variación en el léxico disponible de los estudiantes de español como lengua extranjera. Resultados cuantitativos. En J. L. Bueno Alonso *et al.*, *Analizar datos: Describir variación / Analysing data: Describing variation* (pp. 265-273). Vigo: Universidade de Vigo.
- Sánchez-Saus Laserna, M. (2011). *Bases semánticas para el estudio de los centros de interés del léxico disponible. Disponibilidad léxica de informantes extranjeros en las universidades andaluzas*. Cádiz: Universidad de Cádiz. Tesis doctoral inédita.
- Sánchez-Saus Laserna, M. (2012). *Bases semánticas para el estudio de los centros de interés del léxico disponible. Disponibilidad léxica de informantes extranjeros en las universidades andaluzas*. Cádiz: Servicio de Publicaciones de la Universidad de Cádiz.
- Sánchez-Saus Laserna, M. (2016). *Léxico disponible de estudiantes de español como lengua extranjera en las universidades andaluzas*. Sevilla: Universidad de Sevilla.
- Sánchez, A. (1995). Organización del corpus “Cumbre”. En A. Sánchez, R. Sarmiento, P. Cantos & J. Simón (Ed.), *CUMBRE. Corpus lingüístico del español contemporáneo. Fundamentos, metodología y análisis* (pp. 25-39). Madrid: SGEL.
- Sánchez, A. (2000). Language Teaching before and after ‘Digitalized corpora’. Three main issues. *Cuadernos de Filología Inglesa*, 9(1), 5-37.
- Sánchez, A., Sarmiento, R., Cantos, P. & Simón, J. (1995). *CUMBRE. Corpus lingüístico del español contemporáneo. Fundamentos, metodología y análisis*. Madrid: SGEL.
- Sancho Sánchez, M. (2006). Disponibilidad léxica en los hispanos de los EE.UU. *Cuadernos Cervantes de la lengua española*, 63, 22-31.
- Sandu, B. (2009). *Disponibilidad léxica en español como lengua extranjera de los alumnos rumanos (Estudio exploratorio)*. Diploma de Estudios Avanzados. Las Palmas de Gran Canaria: Universidad de Las Palmas de Gran Canaria.
- Sandu, B. (2010). Cotejo del léxico disponible en español (LE) de dos trabajos realizados en Salamanca y Rumanía. *Interlingüística XXI*, 528-539.
- Sandu, B. (2012). La disponibilidad léxica en alumnos rumanos de ELE: incidencia de la variable ‘sexo/género’ y su correlación con el ‘nivel escolar’. *Lingua Americana Año XVI*, 31, 61-85.
- Sandu, B. (2013). Romanian Students’ Lexical Availability in Spanish as a Foreign Language. *Estudios interlingüísticos*, 1, 121-133.
- Sans, N., Marín, E. & Garmendia, A. (2011). *Bitácora 1*. Barcelona: Editorial Difusión.
- Santiago Guervós, J. (2008). *Léxico disponible de Segovia. Estudio y diccionario*. Burgos: Instituto Castellano y Leonés de la Lengua.

- Santos Palmou, X. (2016). La selección del vocabulario en ELE: estado de la cuestión y nuevas metodologías. En A. M. Sain García (Dir.), *El español como lengua extranjera en Portugal II: retos de la enseñanza de lenguas cercanas* (pp.164-178). Madrid, Ministerio de Educación, Cultura y Deporte.
- Santos Palmou, X. (2017). El Vocabulario Fundamental: historia, definición y nuevas propuestas aplicadas a la enseñanza de ELE. *E-AESLA*, nº. 3, 110-120. Asociación Española de Lingüística Aplicada (AESLA) y el Instituto Cervantes.
- Santos Palmou, X. & Mas Álvarez, I. (2017). Léxico disponible en alumnado de español en la enseñanza superior en Oporto: propuestas para su análisis. *Investigaciones actuales en Lingüística. Vol. I: Aplicaciones de la Lingüística* (pp. 93-111).
- Sanz Álava, I. (2000). El léxico en la enseñanza de ELE: una aplicación práctica para el nivel intermedio. En M. A. Martín Zorraquino & C. Díez Pelegrín (Eds.), *¿Qué español enseñar? norma y variación lingüísticas en la enseñanza del español a extranjeros: Actas del XI Congreso Internacional ASELE, Zaragoza 13-16 de septiembre de 2000* (pp. 635-646). Zaragoza: ASELE.
- Saragi T., Nation, P. & Meister, G. (1978). Vocabulary learning and reading. *System* 6, 70-78.
- Saralegui, C. & Taberner, C. (2008). *Aportación al Proyecto Panhispánico de Léxico Disponible: Navarra*. En I. Olza Moreno, M. Casado Velarde & R. González Ruiz (Eds.), *Actas del XXXVII Simposio Internacional de la Sociedad Española de Lingüística (SEL)* (pp. 745-761). Pamplona: Servicio de publicaciones de la U. de Navarra.
- Schlyter, B. (1951). *Centrala ord förrädet i franskan* (Vocabulaire central du français). Upsal: Almqvist et Wiksell.
- Schmitt, N. (1997). Vocabulary learning strategies. En N. Schmitt, & M. McCarthy (Eds.) *Vocabulary: Description, acquisition and pedagogy* (pp. 199-227). Cambridge: Cambridge University Press.
- Schmitt, N. (2000). *Vocabulary in Language Teaching*. Cambridge, U.K: Cambridge University Press.
- Schmitt, N. & McCarthy, M. (1997). *Vocabulary: Description, acquisition and pedagogy*. Cambridge: Cambridge University Press.
- Schonell, F. Meddleton, I., Shaw, B., Routh, M., Popham, D., Gill, G., Mackrell, G. & Stephens, C. (1956). *A Study of the Oral Vocabulary of Adults. Brisbane and London: University of Queensland Press / University of London Press*.
- Sebastián, N., Cuetos, F., Martí, M. A., & Carreiras, M. F (2000). *LEXESP: Léxico informatizado del español [CD-ROM]*. Barcelona: Edicions de la Universitat de Barcelona.
- Segalowitz, N. (2007). Access fluidity, attention control, and the acquisition of fluency in a second language. *TESOL Quarterly* 41, (1), 181-186.
- Serfati, M. (2010). *Disponibilidad léxica en alumnos marroquíes de español como lengua extranjera (nivel universitario)*. Memoria de máster inédita. Agadir, Marruecos: Universidad de Ibn Zohr.
- Serrano Zapata, M. (2002). *Análisis sociolingüístico del léxico castellano disponible en la ciudad de Lleida*. Lleida: Universitat de Lleida.
- Serrano Zapata, M. (2003). Disponibilidad léxica en la provincia de Lérida: aspectos metodológicos. *Interlingüística*, 14, 929-937.

- Serrano Zapata, M. (2006). Consecuencias del contacto de lenguas en Lérida: interferencias detectadas en las encuestas de disponibilidad léxica. En J. L. Blas, M. Casanova Ávalos & M. Velando Casanova (eds.), *Discurso y sociedad. Contribuciones al estudio de la lengua en un contexto social* (pp. 811-29). Castellón de la Plana: Servicio de Publicaciones de la U. Jaume I.
- Šifrar Kalan, M. (2009). Disponibilidad léxica en español como lengua extranjera: el cotejo de las investigaciones en Eslovenia, Salamanca y Finlandia. *Verba Hispanica*, 17, 165-182.
- Šifrar Kalan, M. (2011). *Leksikalna razpoložljivost v spanscini kot tujem jeziku*. Tesis doctoral inédita. Ljubljana: Universidad.
- Šifrar Kalan, M. (2012) Análisis comparativo de la disponibilidad léxica en español como lengua extranjera (ELE) y lengua materna (ELM). *Marcoele: Revista de Didáctica del español como lengua extranjera*.
- Šifrar Kalan, M. (2014). Disponibilidad léxica en diferentes niveles de español/lengua extranjera. *Studia Romanica Posnaniensia*, 41(1), 63-85.
- Sinclair, J. M. (1966). Beginning the study of lexis. En C. E. Bazell *et al.* (Ed.), *In Memory of John Firth*. Londres: Longmans.
- Sinclair, J. M. (Ed.) (1987). *Collins COBUILD English Language Dictionary*. London: Collins ELT.
- Sinclair, J. M. (1991). *Corpus, concordance, collocation*. New York: Oxford University Press.
- Sinclair, J. M. (2003). *Collins COBUILD advanced learner's English dictionary*. Glasgow: HarperCollins Publishers.
- Sinclair, J. M. & Renouf, A. (1988). A lexical syllabus for language learning. En R. Carter & M. McCarthy (Eds.), *Vocabulary and Language Teaching* (pp. 140-158). London: Longman.
- Sinclair J. M. & Renouf, A. (1996). A lexical syllabus for language learning. En R. Cárter & M. McCarthy (Eds.), *Vocabulary and Language Teaching* (pp. 140-160). London: Longman.
- Slagter, P. J. (1979). *Un nivel umbral*. Estrasburgo: Consejo de Europa.
- Smedberg, A. (1896). *Svenska Landsmålen* XI, 9, 1-28.
- Sökmen, A. J. (1997). Current trends in teaching second language vocabulary. En N. Schmitt & M. McCarthy, *Vocabulary. Description, Acquisition and Pedagogy* (pp. 237-257). Cambridge: Cambridge University Press.
- Soler-Espiauba Conesa, D. (1994). De la España tópica a la postmodernidad a través de nuestros manuales de español como lengua extranjera: ¿Qué imagen de la sociedad presentamos a nuestros alumnos? En Sánchez Lobato, J. & Santos Gargallo, I. (Coord.), *Actas del IV Congreso Internacional de ASELE. Problemas y métodos en la enseñanza del español como lengua extranjera* (pp. 479-490). Madrid: SGEL.
- Stanovich, K. (1986). Matthew effects in reading: Some consequences of individual differences in the acquisition of literacy. *Reading Research Quarterly* 21, 360-400.
- Stern, H. H. (1983). *Fundamental Concepts of Language Teaching*. Oxford: Oxford University Press.
- Sternberg, R. (1987). Most vocabulary is learnt from context. En M. McKeown & M. Curtis (Eds.), *The nature of vocabulary acquisition* (pp. 89-101). Hillsdale: Lawrence Earlbaum.

- Taberero Sala, C. (2008). Disponibilidad léxica y contacto de lenguas. *Oihenart. Cuadernos de lengua y Literatura*, 23, 545-65.
- Terrádez Gurrea, M. (2001). Frecuencias léxicas del español coloquial: análisis cuantitativo y cualitativo. *Cuadernos de Filología, Anejo XLI*. Valencia: Universitat de València.
- Tharp, J. B. (1934). The Basic French Vocabulary. *Modern Language Journal*, January 1934.
- The British National Corpus*, version 2 (BNC World). (2001). Oxford: Oxford University. Disponible en: <http://www.natcorp.ox.ac.uk/>
- The British National Corpus*, version 3 (BNC XML Edition) (2007). Oxford: Oxford University. Disponible en: <http://www.natcorp.ox.ac.uk/>
- Thornbury, S. (2002). *How to Teach Vocabulary*. Paperback: Pearson Education ESL.
- Thorndike, E. L. (1921). *The Teacher's Word Book*. New York: Columbia University, Teachers College.
- Thorndike, E. L. (1932). *A Teacher's Word Book of the 20.000 Words Found Most Frequently and Widely in General Reading for Children and Young People*. New York: Columbia University, Teachers College.
- Thorndike, E. L. & Lorge, I. (1944). *The Teacher's Word Book of 30.000 Words*. New York: Columbia University, Teachers College.
- Tomé Cornejo, C. (2010). *Aspectos metodológicos de los estudios de disponibilidad léxica. Los centros de interés*. Trabajo de Grado dirigido por J. A. Bartol Hernández. Salamanca: Universidad de Salamanca.
- Tomé Cornejo, C. (2011). Reflexiones en torno a los centros de interés. En V. Galloso Camacho (Coord.), *La investigación de la Lengua y la Literatura en la Onubense* (pp. 119-159). Huelva: Universidad de Huelva.
- Tomé Cornejo, C. (2012). La utilización de guiones ('scripts') como centros de interés en los estudios de disponibilidad léxica. En P. Álvarez *et al.* (Eds.), *Interlingüística*, 22, 2, (pp. 459-474). Salamanca: Luso-Española de Ediciones, S.L.
- Tomé Cornejo, C. (2015). *Léxico disponible. Procesamiento y aplicación a la enseñanza de ELE*. Tesis Doctoral inédita defendida en el curso 2014-2015 en la Universidad de Salamanca. Salamanca: Universidad de Salamanca.
- Tony McEnery and Andrew Wilson (2001). *Corpus Linguistics: An Introduction* (2nd. edition). Edinburgh: Edinburgh University Press.
- Torres González, A. N. & García Falero, A. B. (2009). Planificación y primeros resultados de la disponibilidad léxica en Tenerife. En D. Corbella Díaz, J. Dorta (eds.), *La investigación dialectológica en la actualidad* (pp. 75-101). Santa Cruz de Tenerife: Agencia Canaria de Investigación, Innovación y Sociedad de la Información del Gobierno de Canarias.
- Tréville, M.C. & Duquette L. (1996). *Enseigner le vocabulaire en classe de langue*. Paris, Hachette: Colección F.
- Trigo Ibáñez, E. (2011). *Dialectología y cultura: el léxico disponible de los preuniversitarios sevillanos*. Valencia: Aduana Vieja.
- Trigo Ibáñez, E. & González, A. (2011) Estudio del comportamiento de la variable sexo en el léxico disponible de los preuniversitarios sevillanos. *Diálogo de la lengua III*, 28-41.

- Urzúa, P., Sáez, P. & Echeverría M. S. (2006). Disponibilidad léxica matemática. Análisis cuantitativo y cualitativo. *RLA. Revista de Lingüística Teórica y Aplicada*, 44 (2), 59-76.
- Valdés, G., Dearholt, D.W., Barrera, R. & Cárdenas, M. (1986). La derivación de una lista práctica de frecuencias basada en *El recuento de vocabulario* de Rodríguez Bou. *Lectura y Vida*. Año 7. N.º. 1, 9-12.
- Valencia, A. (2005). Dialectalismos en el léxico disponible chileno. En *Actas del XV Congreso de la ALFAL*. Monterrey: Edición en CD Rom.
- Valencia, A. (1994a). *El léxico de los estudiantes de 4º año de Educación Media. Centro de Interés "Procesos mentales"*. Santiago de Chile: Serie Documentos de estudio 26: CPEIP.
- Valencia, A. (1994b). Disponibilidad léxica en Educación Media: III y IV. En *X Encuentro en torno a la admisión Universitaria*. Santiago de Chile: Talleres Gráficos DIVEST.
- Valencia, A. (1997). Disponibilidad léxica, muestreo y estadísticos. *Onomazein*, n.º 2, 197-226.
- Valencia, A. (1998a). Aspectos sociolingüísticos del léxico estudiantil de la Región Metropolitana. En M. Bernales & C. Contreras (Orgs.), *Por los caminos del lenguaje* (pp. 39-45). Temuco: Universidad de la Frontera.
- Valencia, A. (1998b). Léxico estudiantil. Comentario sobre un repertorio. *Boletín de Filología* 37 (Homenaje a Ambrosio Rabanales) (pp. 1211-1221).
- Valencia, A. (2000). Ciencia y tecnología. Análisis del léxico estudiantil. *Nueva Revista del Pacífico*, n.º. 45, 89-98. Valparaíso: Universidad de Playa Ancha.
- Valencia, A. & Echeverría, M. S. (1998). *Disponibilidad léxica en estudiantes de cuarto año de Educación Media*. Santiago de Chile: Universidad de Chile / Universidad de Concepción.
- Valencia, A. & Echeverría, M. S. (1999a). *Disponibilidad léxica en estudiantes chilenos*. Santiago de Chile: Universidad de Chile / Universidad de Concepción.
- Valencia, A. & Echeverría, M. S. (1999b). El factor geográfico en la disponibilidad léxica chilena. En J. A. Samper Padilla *et al.* (Eds.), *Actas del XI Congreso Internacional de la ALFAL, Las Palmas de Gran Canaria* (pp. 1705-1711). Las Palmas de Gran Canaria: Universidad de Las Palmas de Gran Canaria.
- Vander Beke, G. E. (1935). *French Word Book*. New York: The Macmillan Company. Publications of the American and Canadian Committees on Modern Languages, Vol. XV.
- Varela, S. (2003). Léxico, morfología y gramática en la enseñanza de español como lengua extranjera. *Estudios de Lingüística de la Universidad de Alicante*, 17, 571-588.
- Vargas Sandoval, P. (1991). La disponibilidad léxica: un diagnóstico al hablante medio de la V región en seis centros de interés. *Nueva revista del Pacífico*, n.º. 33-36, 115-123.
- Verdeses-Mirabal, R. T. (2012). Disponibilidad léxica de los estudiantes hispanos de Redwood City, California. *Revista Nebrija de Lingüística Aplicada*, 11 (6), 3-50.
- Verlée, L. (1954). *Basis Woordenboek voor de Franse Taal*. Amsterdam: Neulenhoff.

- Villalba Martínez, F. & Hernández, M. T. (2004). *Español segunda lengua. Educación Secundaria*. Madrid: Anaya.
- Villarreal Castillo, M. (2013). *Léxico disponible de Panamá*. Panamá: Meduca.
- Villena Ponsoda, J. A., Ávila Muñoz A. M. & J. M. Sánchez Sáez (2010). Patrones sociolingüísticos del vocabulario disponible. Condicionamiento estratificacional de la capacidad léxica en la ciudad de Málaga. Proyecto CONSOLEX. En E. Ridruejo Alonso; T. Solías Arís; N. Mendizábal de la Cruz & S. Alonso Calvo (Coords.), *Tradición y progreso en la Lingüística General* (pp. 409-432). Valladolid: Área de Lingüística General, Departamento de Lengua Española de la Universidad de Valladolid.
- Vincze, O., Alonso Ramos, M., Mosqueira Suárez, E. & Prieto González, S. (2011). Exploiting a learner corpus for the development of a CALL environment for learning Spanish collocations. En I. Kosem & K. Kosem (Eds.), *Electronic lexicography in the 21st century: New applications for new users. Proceedings of eLex 201* (pp. 280-285). Ljubljana: Trojina, Institute for Applied Slovene Studies.
- Watts, A. F. (1946). *The language and Mental Development of children*. London: George G. Harrap & Co. Ltd.
- Weinreich, U. (1953). *Languages in contact: Findings and problems*. The Hague: Mouton.
- West, M. (1935). *The new method English dictionary*. London: Longman, Green and Co.
- West, M. (1953). *A General Service List of English Words*. London: Longman.
- Widdowson, H. G. (1984). *Explorations in Applied Linguistics 2*. Oxford: Oxford University Press.
- Wilkins, D. A. (1972). *Linguistics in language teaching*. London: Edward Arnold.
- Willis, D. (1999). Syllabus Design and the Pedagogic Corpus. En J. Aitchison, H. Funk, R. Galisson, G. List, M.A. Mochet, C. O'Neil, C. Owen, W. Ulrich, G. Vigner & D. Willis (Eds.), *Vocabulary learning in a foreign language* (pp. 115-148). Fontenay Saint-Cloud: ENS Editions.
- Willis, J. & Willis, D. (1988). *Collins COBUILD English Course*. London: Collins.
- Wingeyer, H. R. (2014). Temas de la disponibilidad léxica de la región NEA de Argentina: mitos populares, la ciudad, el campo, insultos y diversiones nocturnas. Aportes para la elaboración de programaciones interculturales y de materiales en la enseñanza de E/LE. *SIGNOS ELE (Revista de español como lengua extranjera)*, jul. 2014.
- Xue, G. & Nation, I. S. P. (1984). A University Word List. *Language Learning and Communication* 3 (2), 215-229.
- Zimmerman, C. B. (1997). Historical trends in second language vocabulary Acquisition. En J. Coady & T. Huckin (Eds.), *Second language vocabulary Acquisition: A rationale for pedagogy* (pp. 5-19). Cambridge, UK: Cambridge University Press.
- Zimmermann, H. J. (2001). *Fuzzy Set Theory and its Applications*. Boston: Kluwer Academic Publishers.
- Zuluaga Ospina, A. (2002). Los 'enlaces frecuentes' de Maria Moliner. Observaciones sobre las llamadas colocaciones. *LEA: Lingüística española actual, Vol. 24, n° 1, 97-114*.

Índice de cuadros

Pág.	CUADROS
24	1. Aspectos que entran en juego al conocer nuevas palabras (Harmer 1991: 158)
26	2. Matices que se pueden enseñar de la palabra <i>majo-a</i> según Higuera García (2004a:11) 26
49	3. Relación entre tipos de aprendizaje y algunas de sus técnicas (Nation 2001, <i>apud</i> Higuera García 2004a)
74	4. Clasificación del tipo de unidades léxicas
159	5. Ejemplo de una familia de palabras ordenada en niveles, según los criterios de Bauer & Nation (1993: 254)
184	6. Relación entre los grados de comprensión de la hipótesis del nivel umbral y el número de palabras necesarias para lograrlo, distinguiendo oralidad y escritura

Índice de figuras

Pág.	IMÁGENES
18	1. Asociaciones de la palabra <i>coche</i> (Higueras García 2004a: 14)
41	2. Página interior del manual de Eckersley (1938)
42	3. Ejemplo de ejercicio para repetición (Alexander 1967)
43	4. Portada del libro de Rojo Sastre, Rivenc y Ferrer (1968)
341	5. Definición del vocabulario fundamental desde una óptica funcional
342	6. Métodos objetivos de selección léxica
343	7. Definición del vocabulario fundamental desde una óptica analítica
364	8. Relación entre la cobertura léxica textual y los umbrales de comprensión léxica

Pág.	GRÁFICOS
369	1. Evolución del ID de los vocablos del CI <i>El cuerpo humano</i> extraídos de Tomé Cornejo (2015)
372	2. Evolución del ID de los vocablos del CI <i>Calefacción e Iluminación</i> , López Meirama (2008)
375	3. Histograma de 10 conjuntos a partir de la frecuencia de palabras según su ID en el CI <i>El cuerpo humano</i> de Tomé Cornejo (2015)
375	4. Histograma de 10 conjuntos a partir de la frecuencia de palabras según su ID en el CI <i>El cuerpo humano</i> de López Meirama (2008)
376	5. Histograma de 10 conjuntos con su línea de tendencia a partir de la frecuencia de palabras según su ID en el CI <i>El cuerpo humano</i> de Tomé Cornejo (2015)
376	6. Histograma de 10 conjuntos con su línea de tendencia a partir de la frecuencia de palabras según su ID en el CI <i>El cuerpo humano</i> de López Meirama (2008)
377	7. Histograma de 10 conjuntos a partir de la frecuencia de palabras según su ID en el CI <i>Profesiones y oficios</i> de Tomé Cornejo (2015)
377	8. Histograma de 10 conjuntos a partir de la frecuencia de palabras según su ID en el CI <i>Profesiones y oficios</i> de López Meirama (2008)

378	9. Histograma de 10 conjuntos con su línea de tendencia a partir de la frecuencia de palabras según su ID en el CI <i>Profesiones y oficios</i> de Tomé Cornejo (2015)
378	10. Histograma de 10 conjuntos con su línea de tendencia a partir de la frecuencia de palabras según su ID en el CI <i>Profesiones y oficios</i> de López Meirama (2008)
388	11. Histograma de 20 conjuntos con su línea de tendencia a partir de la frecuencia de palabras según su ID en el CI <i>El cuerpo humano</i> de Tomé Cornejo (2015)
389	12. Límites de selección de las palabras que se corresponden con los niveles del MCERL en el CI <i>El cuerpo humano</i> de Tomé Cornejo (2015)
396	13. Histograma de 20 conjuntos a partir de la frecuencia de palabras según su ID en el CI <i>Alimentación y nutrición</i> de Tomé Cornejo (2015)
400	14. Representación de una <i>Fuzzy -C partition</i> y las etiquetas lingüísticas de los <i>clusters</i> (Zimmermann 2001: 304)
401	15. <i>Fuzzy C partition</i> de las palabras disponibles y etiquetas lingüísticas acuñadas para cada grupo

Índice de tablas

Pág.	TABLAS
48	1. Relación entre el tipo de vocabulario, cantidad, frecuencia, cobertura textual e implicaciones didácticas (Nation 2000, <i>apud</i> Higuera García 2004a: 8)
81	2. Clasificación de los factores que pueden dificultar o facilitar el aprendizaje de una palabra (Laufer 1997a y 1997b <i>apud</i> López Pérez 2013)
82	3. Tipos de vocabularios según Bogaards (1994 <i>apud</i> López Pérez 2013)
127-128	4. Comparación entre los centros de interés de la disponibilidad léxica, los temas y subtemas del Consejo de Europa y las nociones específicas del <i>PCIC</i> (Bartol Hernández 2010: 95)
130	5. Comparación entre la noción <i>Ocio</i> (<i>PCIC</i>) y las primeras 50 palabras del CI 15. <i>Juegos y diversiones del Léxico disponible en el español de Galicia</i> (López Meirama 2008)
134-137	6. Los centros de interés propuestos por Tomé Cornejo (2015: 441-4)
142	7. Resumen de los puntos fuertes y débiles de la frecuencia y la disponibilidad léxicas para la selección del vocabulario
155	8. Relación entre el número de palabras frecuentes y el porcentaje de cobertura textual obtenido con ellas según los estudios
163	9. Relación entre <i>tokens</i> , <i>types</i> y <i>word families</i> del manual <i>F2F</i> organizados por Range según Criado & Sánchez (2012: 83)
164	10. Número de <i>tokens</i> , <i>types</i> y <i>word families</i> que contienen las 15 listas elaboradas por Nation (2006: 62-67)
169	11. Número de palabras y relación con su cobertura léxica textual en los diferentes tipos de textos (Nation 2001: 17)
172	12. Relación entre los porcentajes de cobertura en textos escritos y orales y las listas de las palabras más frecuentes de Nation (2006: 79)
175	13. Relación entre el porcentaje de cobertura textual, la densidad de palabras conocidas y el número de líneas del texto por cada palabra desconocida
183	14. Relación entre el número de palabras frecuentes, su cobertura léxica textual, el nivel de comprensión que se alcanza y los niveles del <i>MCERL</i>
197	15. Algunos corpus textuales del inglés y sus características
198	16. Las primeras 50 palabras más frecuentes en algunos corpus del inglés

319-320	17. Centros de interés usados en los proyectos del francés, el español y el portugués fundamental
347-348	18. Equivalencias entre los temas considerados en López Meirama (2008), Tomé Cornejo (2015), Davies (2006) y el <i>PCIC</i>
353	19. Número de palabras contabilizadas en los núcleos temáticos de López Meirama (2008) y Tomé Cornejo (2015) con un índice de disponibilidad $\geq 0,1$, dentro de las 5.000 más frecuentes en Davies (2006) y pertenecientes a los niveles A1-A2 y B1 en el <i>PCIC</i> . Número de unidades coincidentes entre los cuatro repertorios
390	20. Palabras disponibles seleccionadas según su ID para el nivel elemental (A1-A2) dentro del CI <i>El cuerpo humano</i> de Tomé Cornejo (2015)
390-391	21. Palabras disponibles seleccionadas según su ID para el nivel umbral (B1) dentro del CI <i>El cuerpo humano</i> de Tomé Cornejo (2015)
392-393	22. Palabras disponibles seleccionadas según su ID para el nivel avanzado (B2) dentro del CI <i>El cuerpo humano</i> de Tomé Cornejo (2015)
394-395	23. Palabras disponibles seleccionadas según su ID para el nivel competente (C) dentro del CI <i>El cuerpo humano</i> de Tomé Cornejo (2015)
419-428	24. Muestra del vocabulario fundamental organizado alfabéticamente. Letra A
430	25. Preposiciones del vocabulario fundamental según el orden de frecuencia de Davies (2006)
432-435	26. Adverbios que forman parte del vocabulario fundamental según el orden de frecuencia de Davies (2006)
436	27. Los colores más frecuentes según Davies (2006) que forman parte del vocabulario fundamental
437-439	28. Palabras del vocabulario fundamental dentro del subtema <i>El cuerpo humano</i> (1. <i>El individuo</i> , 1.1. <i>Dimensión física, perceptiva y anímica</i>)
440-442	29. Temas considerados para la organización del vocabulario fundamental temático

Índice de abreviaturas y siglas

CI: centro de interés

DiCE: Diccionario de Colocaciones del Español (Alonso Ramos 2004). Disponible en: <http://www.dicesp.com>.

DL: disponibilidad léxica

ELE: español como lengua extranjera

ENE: español de los negocios

EpFE: español para fines específicos

FDSW: Frequency Dictionary of Spanish Words (Juilland & Chang-Rodríguez 1964).

ID: índice de disponibilidad

L1: primera lengua o lengua inicial

L2: segunda lengua

LB: léxico básico

LD: léxico disponible

LE: lengua extranjera

LF: léxico frecuente

LM: lengua materna

loc.: locución

MCER: Marco común europeo de referencia para las lenguas: aprendizaje, enseñanza, evaluación.

PCIC: Plan curricular del Instituto Cervantes (Instituto Cervantes 2006). También disponible en red:

http://cvc.cervantes.es/ensenanza/biblioteca_ele/plan_curricular/default.htm

PPHDL: Proyecto Panhispánico de Estudio de la Disponibilidad Léxica

STTR: *standardised type/token ratio*

TIC: Tecnologías de la información y la comunicación

TTR: *type-token ratio*

VF: vocabulario fundamental