

ACERCAMIENTO CRÍTICO A LA SELECCIÓN OBJETIVA DE CONTENIDOS LÉXICOS EN LA ENSEÑANZA DE E/LE¹

VENTURA SALAZAR GARCÍA
Universidad de Alicante
Ventura.Salazar@ua.es

Resumen

The aim of this paper consists in offering a critical perspective on the main procedures for the objective selection of lexical contents in second language teaching. Three criteria are considered: frequency, use and availability. After analysing the methodological bases on this research field, and its results in the study of Spanish vocabulary, it is found out that none of them is satisfactory enough. The frequency criterion is revealed as the most operative of the three ones; thus, it may be a useful tool for teaching if we complement it with other data. But, into the context of a pedagogy based on a learner-centred curriculum, any objective selection of lexical contents must be subject to the learner's needs, the classroom negotiation, and the design of communicative learning tasks.

1. Introducción

1.1. En un ensayo sobre adquisición del léxico que merecería más atención de la recibida hasta ahora, René Appel (1996: 397-399) formula cinco principios básicos, emanados de la reciente investigación psicolingüística, que han de guiar la enseñanza del vocabulario en una lengua extranjera. Me interesa destacar ahora dos de ellos, situados en segundo y cuarto lugar. Por un lado, Appel establece que aprender el vocabulario de un idioma extranjero requiere por parte del estudiante un esfuerzo cognitivo consciente. Por otro, postula que es necesario repetir en varias ocasiones las nuevas palabras antes de esperar que sean adquiridas por parte del aprendiz.

Ambos principios entroncan directamente con asunciones tradicionales de la intervención pedagógica, y por sí solos no resultan novedosos. Su originalidad estriba más bien en que tales principios se han de concretar en unos planteamientos curriculares que superen ciertas prácticas anquilosadas. Por ejemplo, el esfuerzo cognitivo no se corresponde con una actividad exclusivamente mnemotécnica, sino que implica un conocimiento operacional (un saber hacer) manifestado en tareas concretas desempeñadas en situaciones concretas. Consecuentemente, los meros listados de palabras aisladas de contexto, con o sin correspondencia bilingüe, se revelan como una técnica de aprendizaje profundamente trivial. La necesidad de repetición guarda poca semejanza con la reiteración mecánica (pongamos por caso, re-

1 El presente trabajo se inserta dentro del proyecto de investigación HUM2004-04296/FILO, financiado por la Dirección General de Investigación (Ministerio de Educación y Ciencia).

pasando los listados antes aludidos) y mucha con la reaparición de la misma unidad léxica en diversas situaciones de comunicación y aprendizaje en las que el estudiante participa de modo efectivo. Según Jenkins y Dixon (1983), para contar con garantías de que una palabra ha sido aprendida es necesario un mínimo de seis a doce exposiciones. Esta horquilla está lejos de constituir un patrón universal, pues debe ser tenida en cuenta la incidencia de otros factores como el grado de atención, la marcación afectiva, la aplicación de estrategias asociativas, etc. En cualquier caso, se trata de una referencia importante porque obliga a relativizar los efectos de las muestras de lengua (*input*), y sus posibilidades para pasar a formar parte de los contenidos verdaderamente adquiridos (*intake*). En muchos manuales de idiomas y materiales didácticos al uso, la mayor parte de las unidades léxicas que conforman el sílabus aparece apenas una o dos veces, lo cual, si no va acompañado de una actividad didáctica acorde con lo arriba señalado, resulta a todas luces insuficiente. A la cantidad de *input* debe ir unida, además, la calidad, que entronca con aspectos tales como una cierta tensión metalingüística que explote el potencial de aprendizaje del *input*. Creo que es esta variable la que incide, por ejemplo, en el hecho de que, como constatan Pérez y Picó (1995), los programas de lectura extensiva no garanticen un progreso real de la competencia². Al menos por lo que respecta a la adquisición de una segunda lengua, todo apunta a que no es la lectura extensiva la que promueve el aprendizaje del léxico, sino, más bien al contrario, el nivel de competencia léxica el que condiciona el éxito de la lectura; la intuición propia de un lector autónomo, por la cual es posible extraer del contexto el significado de una palabra desconocida, sólo es viable si dicho lector domina la gran mayoría del resto del vocabulario presente en el texto, en una proporción que Laufer (1997: 24) sitúa entre el 90 y el 95%, y que otros autores elevan nada menos que al 98% (cf. Pérez Basanta, 1999: 275).

De todo lo expuesto se deduce que la adquisición de la competencia léxica es un proceso nada fácil, que exige interés, voluntad y no poca dedicación tanto por parte del docente como, en muy principal medida, del aprendiz. Constituye un problema mayor de la didáctica del vocabulario ofrecer los instrumentos para rentabilizar al máximo ese esfuerzo de obligada inversión.

1.2. El número de unidades léxicas de la lengua es enorme, y en la práctica inabarcable. Resulta un hecho comúnmente aceptado que el aprendizaje del léxico de una segunda lengua constituye un proceso limitado y esencialmente gradual. En este aspecto, hay una notable coincidencia entre la adquisición de la lengua materna y la de una segunda lengua. Los hablantes nativos aprenden el léxico de su lengua de una manera paulatina y, si bien su progreso más espectacular se produce durante la infancia, en principio es un camino abierto que puede prolongarse a lo largo de toda la vida, y que realmente nunca llega a completarse. Puede afirmarse que ningún individuo, por culto que sea, 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 el español, en las que la amplitud de su comunidad vernácula entraña una considerable diversidad dialectal, sociolingüística y

2 Por su parte, Paribakht y Wesche (1997) sí aprecian una correlación positiva entre la lectura extensiva centrada en el significado y el progreso de la competencia léxica de los aprendices. Pero constatan asimismo que la combinación de dicha lectura con actividades de instrucción formal ofrece resultados sensiblemente mejores tanto en términos cuantitativos como cualitativos.

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. Dicho en otras palabras, 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.

Para proceder a la selección de contenidos léxicos se han manejado fundamentalmente tres tipos de criterios (cf. Bartol, 2002: 63): el objetivo, fundamentado en investigaciones lexicoestadísticas previas, el subjetivo, que habitualmente apela a la competencia e intuición del profesor o de los redactores de materiales didácticos, y, por último, el objetivo corregido, que combina eclécticamente una base lexicoestadística con un posterior ajuste de naturaleza subjetiva. El presente trabajo tiene como propósito ofrecer una evaluación crítica de los procedimientos de limitación objetiva. Dicha evaluación atenderá tanto a sus presupuestos de partida como a su desarrollo efectivo en el ámbito concreto de la lengua española. Como apuntaré en más de una ocasión, mis críticas no pretenden desestimar los logros y posibilidades de estas aportaciones, sino simplemente llamar la atención sobre las limitaciones que, por diversos motivos, restringen su alcance como herramienta didáctica, y hacen desaconsejable, al menos desde planteamientos de corte comunicativo, su utilización como criterio único a la hora de determinar los contenidos léxicos.

1.3. Tres son, a mi juicio, los parámetros relevantes de cara a una selección objetiva de los contenidos léxicos: frecuencia, uso y disponibilidad, que nos conducen en buena lógica a las nociones de *léxico frecuente*, *léxico usual* y *léxico disponible*. Estas tres nociones están estrechamente relacionadas entre sí, pero, como intentaré mostrar en estas páginas, no por ello dejan de ser conceptualmente distintas³. Los vínculos se mueven en un doble sentido. Por una parte, se aprecia una relación de inclusión: el léxico frecuente aparece como un subconjunto del léxico usual, que a su vez es un subconjunto del léxico disponible. Por otra, se aprecia una relación de implicación. Quiere esto decir que la frecuencia es el factor más decisivo para el uso (aunque no necesariamente al contrario), y, de modo análogo, el uso es el factor más decisivo de cara a la disponibilidad. Así es como creo que debe entenderse la constatación de Aitchison (1987: 180) según la cual las palabras de uso más frecuente son más fáciles de recordar y, consecuentemente, requieren para su procesamiento una menor cantidad de tiempo y de esfuerzo. Veamos todo esto con más detalle.

3 A los problemas de índole conceptual se añaden también las considerables oscilaciones terminológicas. Por ejemplo, García Hoz (1953) manejaba el término *usual* en un sentido diferente al que le doy yo aquí. Para él, el *vocabulario usual* era el formado por el conjunto de vocablos documentados en su corpus de referencia, y que sumaba un total de doce mil novecientas trece unidades. A partir de ahí, extraía como subconjunto el *vocabulario común* (mil novecientas setenta y una unidades) y, como listado esencial último, el *vocabulario fundamental* (doscientas ocho unidades). Esto último tampoco concuerda con los usos terminológicos de la bibliografía posterior, en donde el vocabulario fundamental abarca un volumen léxico más amplio, que emana de la combinación de los datos de frecuencia y los de disponibilidad (vid. 4.2 y nota 10). Tampoco el sentido que le da Justicia (1995) a los términos *vocabulario usual* y *vocabulario común* resulta coincidente con el que he documentado en otras investigaciones lexicométricas.

2. El léxico frecuente

2.1. La noción de frecuencia es en principio válida para cualquier nivel de análisis lingüístico (desde el fónico hasta el discursivo, pasando por el gramatical), pero lo cierto es que ha sido el nivel léxico el que ha acaparado un mayor interés tanto para la investigación como para su aplicación a la enseñanza de segundas lenguas. Por léxico frecuente hay que entender el conjunto de palabras que muestra un mayor grado de presencia en los textos, orales o escritos, emitidos en un idioma. Obviamente, resultaría descabellada cualquier pretensión de analizar la totalidad de enunciados proferidos en una lengua viva, por lo que el cálculo que se quiera hacer al respecto ha de circunscribirse a 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 (cf. Richards, Platt y Platt, 1992: 177). Dicha incidencia queda posteriormente reflejada 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.

Gracias al desarrollo de la lingüística de corpus a lo largo del siglo XX, buena parte de las lenguas de mayor difusión internacional disponen de diccionarios de frecuencias y otros recursos similares, si bien es cierto que faltan aún para un gran número de idiomas. Para el caso de la lengua española, merecen mencionarse los repertorios confeccionados por Juilland y Chang-Rodríguez (1964) y por Alameda y Cuetos (1995), que serán los que reciban una atención prioritaria en el presente trabajo. Asimismo, cabe aludir a otras obras redactadas con anterioridad y que, con irregular fortuna, aspiraban a ofrecer también una caracterización cuantitativa del vocabulario de la lengua española. Entre ellas figuran las que, con una finalidad eminentemente pedagógica, llevaron a cabo Buchanan (1927), Keniston (1933 y 1941), Eaton (1940) y García Hoz (1953). Ahora bien, ciertas características de su formato y metodología, amén del tiempo transcurrido desde su aparición, lastran gravemente su vigencia en nuestros días como instrumentos de partida para una selección objetiva del léxico. Por último, tampoco conviene pasar por alto otras investigaciones que han apelado al criterio de frecuencia con el fin de analizar el desarrollo evolutivo de la competencia léxica durante la infancia. En este sentido, merecen destacarse, para el caso del español, las aportaciones de Fernando Justicia (1985, 1993 y 1995). Los datos aportados por este autor difícilmente pueden ser proyectados a todo el dominio lingüístico panhispánico, pues opera con un corpus circunscrito a unas condiciones diastráticas y diatópicas muy determinadas: muestras de lengua escrita (mediante dos pruebas, una de redacción y otra de asociación de palabras) aportadas por escolares residentes en Andalucía Oriental. No obstante, constituyen un valioso testimonio tanto por su metodología como por sus resultados descriptivos. De ahí que, con las matizaciones correspondientes, puedan servir de apoyo para consideraciones de carácter general.

Todos los partidarios de la selección objetiva del vocabulario en la enseñanza de segundas lenguas admiten la relevancia del parámetro de frecuencia cuando se trata de fijar una gradación de contenidos léxicos en el marco de la planificación curricular. Ahora bien, a la hora de la puesta en práctica de tales menesteres, el manejo de los diccionarios de frecuencias se halla sujeto a ciertos condicionantes que obligan a mantener amplias dosis de cautela. La razón para ello reside en que, como reconocen incluso los propios especialistas

en lingüística de corpus, el cálculo de la frecuencia léxica se enfrenta a diversas dificultades para las que no siempre se dispone de una solución satisfactoria. El diccionario de frecuencias debe ser valorado como una herramienta útil, pero intrínsecamente limitada e incapaz de dar cuenta de todos los aspectos que conforman el papel del vocabulario dentro de la interacción comunicativa. En definitiva, una selección de contenidos léxicos basada *exclusivamente* en el gradiente de frecuencia se apartaría considerablemente de los criterios actualmente asumidos de modo general por los enfoques comunicativos de la enseñanza de segundas lenguas. Dedicaré el próximo epígrafe a exponer y evaluar esas limitaciones, con especial atención a los diccionarios de frecuencias de Juilland y Chang-Rodríguez (1964) y de Alameda y Cuetos (1995).

2.2. Las insuficiencias de los diccionarios de frecuencias han sido constatadas ya en diversas publicaciones. Entre otras muchas referencias, cabe recordar las acertadas puntualizaciones de Jenaro Ortega (1985) a propósito del diccionario de Juilland y Chang-Rodríguez (1964), así como la síntesis crítica, de carácter más general, ofrecida por Bartol (2002: 63 y ss.). Lo que quizá no se ha advertido hasta ahora con el suficiente detalle es que esas insuficiencias responden a motivaciones de distinta índole, que deben ser convenientemente delimitadas si se quiere hacer una valoración verdaderamente ecuaníme del alcance de este tipo de recursos. A grandes rasgos, creo que los problemas reconocibles en la aplicación de los diccionarios de frecuencias a la enseñanza de lenguas pueden ser agrupados en tres grandes apartados. A saber:

a) Fenómenos derivados de la propia naturaleza del léxico, que el diccionario de frecuencias se limita a constatar. Están siempre presentes, independientemente de la calidad del diccionario. Salvo excepciones, son fáciles de reconocer y de subsanar mediante mecanismos correctores no especialmente complejos.

b) Problemas emanados de las limitaciones técnicas que constriñen la confección del diccionario, en virtud de sus condiciones de viabilidad. Los avances técnicos de la moderna informática, así como su plasmación en el campo de la lingüística de corpus y de las llamadas *industrias de la lengua*, han permitido durante los últimos tiempos mejorar enormemente dichas condiciones de viabilidad, y todo apunta a que en el futuro se seguirá progresando en una proporción exponencial. En cualquier caso, todavía estamos lejos de contar con una situación óptima, lo cual provoca que las dificultades de esta índole sigan estando presentes incluso en los diccionarios de más reciente aparición.

c) El último grupo de problemas guarda relación directa con ciertas decisiones metodológicas, generalmente vinculadas a cuestiones tales como la representatividad del corpus de partida, su horquilla diacrónica y el etiquetado de las unidades discretas con las que se opera. Aunque estas circunstancias se encuentran también mediatizadas en parte por condicionantes externos (que van desde las limitaciones de recursos materiales y humanos hasta las exigencias editoriales), en líneas generales la responsabilidad reside mayormente en el diseño de la investigación. De ahí que, a mi modo de ver, son las que mayor incidencia tienen en la calidad del producto final.

2.2.1. Como ejemplo de fenómeno del primer grupo conviene recordar el hecho de que los vocablos más frecuentes de cualquier idioma son siempre de índole gramatical, no léxica. En terminología de Ullmann (1962: 51), las palabras sinsemánticas gozan de índices de

frecuencia mucho más elevados que las palabras autosemánticas. Para el caso del español, es fácil comprobar que la mayor frecuencia corresponde a unidades pertenecientes a clases cerradas: preposiciones, conjunciones, determinantes y pronombres. A ello cabe añadir algunos verbos muy gramaticalizados y de valor sumamente general, por su condición de auxiliares, cópulas o verbos soporte (*haber, ser, estar, tener, hacer*, etc.). En ninguno de los diccionarios de frecuencias del español aparecen nombres o adjetivos entre las cuarenta palabras más frecuentes. Como ya he señalado antes, aquí el diccionario de frecuencias se limita a constatar una realidad que forma parte de la lengua, y que es fruto de la no identidad entre palabras y lexías. En la planificación curricular, sólo estas últimas han de conformar el repertorio de contenidos léxicos. De ahí la necesidad de introducir un mecanismo corrector que fije esta delimitación dentro de los listados de frecuencias. Las palabras sinsemánticas han de contemplarse en el marco de los contenidos gramaticales, para cuya secuenciación la frecuencia se revela como un factor muy secundario en comparación con aspectos tales como la dificultad de procesamiento o la adecuación a las distintas etapas adquisicionales. En ocasiones, y particularmente en el caso de los verbos soporte, resulta difícil trazar la frontera entre las unidades léxicas y las gramaticales. En cualquier caso, y siquiera sea a efectos prácticos, creo que el criterio lexicográfico ofrece suficientes garantías de operatividad. Según dicho criterio, las unidades léxicas son susceptibles de una paráfrasis en forma de definición, por medio de una metalengua de contenido. Las unidades gramaticales, por su parte, sólo admiten paráfrasis en una metalengua de signo. Es decir, pueden ser explicadas, pero no definidas (cf. Seco, 1978: 225).

2.2.2. Entre los problemas del segundo grupo antes mencionado se encuentra, por ejemplo, el hecho de que la inmensa mayoría de los diccionarios de frecuencias, incluidos los del español, adoptan como punto de partida unos *corpora* extraídos exclusivamente a partir de fuentes escritas. Ello conlleva un sesgo muy acusado desde un punto de vista diafásico y diastrático, hasta el punto de que puede asegurarse que los diccionarios de frecuencias conocidos hasta ahora no tienen como universo toda la realidad idiomática, sino sólo aquella circunscrita a los dominios discursivos propios de los textos que han servido de base para la extracción de muestras. Esta insuficiencia encuentra su justificación en los enormes problemas que acarrea hasta hace poco el tortuoso proceso de obtención, transcripción y procesamiento de *corpora* orales de grandes dimensiones. Ahora bien, la situación se ha ido modificando sensiblemente a lo largo de los últimos años.

Un testimonio muy significativo de las nuevas coordenadas, en lo que respecta al español, es la existencia del *Subcorpus Oral* integrado dentro del *Corpus Referencia del Español Actual* (CREA), disponible en el portal electrónico de la Real Academia Española (<http://www.rae.es/nivel1/crea.htm>). Dicho subcorpus incluye, además de materiales propios procedentes de grabaciones de medios de comunicación, otros muchos recabados por proyectos anteriores: el *Corpus Oral de Referencia del Español Contemporáneo* (COREC), elaborado a principios de los años noventa por la Universidad Autónoma de Madrid (cf. Marcos Marín, 1991), el *Macrocorpus de la Norma Lingüística Culta de las principales ciudades del mundo hispánico* (MC-NLCH) (cf. Samper, 1995) y el *Corpus de Análisis de la Conversación* de la Universidad de Alcalá de Henares. En definitiva, todo apunta a que los diccionarios de frecuencias del futuro ya no podrán justificar la omisión de las fuentes orales.

2.2.3. Por último, y por lo que respecta al tercer grupo de problemas, han de ser tomadas en consideración las cuestiones derivadas del método de trabajo seguido en la confección de cada uno de los diccionarios de frecuencias. Aquí la casuística es más amplia, y varía según el repertorio lexicográfico sometido a evaluación.

El diccionario de Juilland y Chang-Rodríguez (1964) constituyó en su día un trabajo encomiable, sobre todo si tenemos en cuenta la fecha de su redacción y su carácter pionero respecto de otros repertorios análogos (para diversas lenguas romances) en los que intervino Alphonse Juilland. Pero ello no es óbice para constatar que, durante las cuatro décadas que median desde su aparición, las aplicaciones pedagógicas derivadas del mismo han sido realmente muy escasas, y, en nuestros días, su utilidad práctica resulta, desde luego, más que discutible. El corpus de partida era bastante restringido (quinientas mil palabras)⁴, y se estableció, para la inclusión de una unidad dentro del diccionario, un umbral de frecuencias excesivamente alto (cinco apariciones absolutas). Esto determinó que el repertorio final contase con poco más de cinco mil entradas diferentes. Eso significa que se trata de una obra que más bien se encuentra a medio camino entre los diccionarios de frecuencias propiamente dichos y los léxicos básicos. Por último, no debe olvidarse que los textos seleccionados tienen su origen en el período comprendido entre 1920 y 1940, lo que abre un más que justificado margen de duda acerca de su correspondencia con la sincronía actual. De todo ello se deduce que, con vistas a la planificación curricular, este diccionario puede servir, a lo sumo, de referencia aproximada en lo relativo a los contenidos propios de los niveles iniciales. Por lo que respecta al vocabulario de los niveles avanzados o con un cierto grado de especialización diatómica, la obra de Juilland y Chang-Rodríguez apenas si aporta elementos de juicio.

El diccionario de Alameda y Cuetos (1995) contiene también abundantes puntos oscuros, algunos de ellos realmente impropios de un diccionario relativamente reciente. Como elemento a su favor, es de agradecer que aspire a ser un diccionario no restrictivo, pues incorpora todas las palabras a partir de una sola aparición. Parte de un corpus muy actual en el momento de su aparición (textos publicados entre 1980 y 1993), con la siguiente distribución proporcional: novela (50%), prensa (25%), ensayo (15%) y divulgación científico-técnica (10%). En total alcanza ya los dos millones de palabras, algo en principio suficiente, por más que se encuentre muy lejos del volumen manejado en los principales diccionarios de frecuencias de otras lenguas. Ahí están, por ejemplo, los cinco millones utilizados hace sesenta años por Thorndike y Lorge (1944) para el inglés, los veintitrés millones y medio que usó Imbs (1971) para el francés y los nada menos que cincuenta y dos millones, trescientas setenta y cinco mil cuarenta y cuatro palabras del corpus textual informatizado que ha permitido la elaboración del diccionario de frecuencias del catalán, bajo la dirección de Rafel i Fontanals (1996-1998)⁵. Incluso para nuestra lengua hay antecedentes de estudios lexicométricos con un corpus mucho más amplio. Al menos eso es lo que se desprende de lo señalado por López Morales (2003: 81), quien afirma que los materiales de partida para

4 Como comparación, baste señalar que un trabajo muy anterior, como el de Buchanan (1927), manejaba un corpus de un millón doscientas mil palabras, procedentes de un más amplio espectro de dominios discursivos.

5 Conviene precisar que tanto Imbs como Rafel i Fontanals operan con textos que abarcan una amplia horquilla diacrónica. Con vistas a su aplicabilidad didáctica, esto último, lejos de ser un aval, constituye un serio inconveniente, ya que distorsiona la perspectiva acerca del estado de lengua realmente existente en el momento de acometer la enseñanza.

la confección del *Recuento de vocabulario español*, publicado en 1952 bajo la dirección de Ismael Rodríguez Bou, alcanzaban casi los seis millones de palabras.

De cualquier modo, los reparos que deben achacarse al corpus utilizado por Alameda y Cuetos residen no tanto en su volumen total o en la distribución de tipos textuales como en su representatividad desde un punto de vista diatópico. Si se examinan los textos seleccionados, se puede comprobar que la inmensa mayoría de los mismos proceden de autores españoles (cf. Alameda y Cuetos, 1995: 4-11). La presencia iberoamericana, que es claramente minoritaria en el apartado de novela y puramente testimonial en el de ensayo, brilla por su ausencia en las otras dos secciones. En total, los textos iberoamericanos apenas si suponen un 10%, según estimaciones de los propios autores (Alameda y Cuetos, 1995: 3). En resumen, el diccionario aquí comentado puede ser, a lo sumo, representativo del léxico del 'español de España', pero no del idioma común a todo el dominio lingüístico hispánico; el cual, no se olvide, está formado en más de un 90% por hablantes del continente americano. Es cierto que el diccionario de Juilland y Chang-Rodríguez omitía los textos iberoamericanos por completo, pero eso era hasta cierto punto explicable por la época de su redacción, en la que todavía pesaban mucho los prejuicios hacia modalidades lingüísticas ajenas a la norma idiomática centropeninsular. Algo que ya no puede ser esgrimido a propósito de un diccionario nacido a las puertas del siglo XXI.

Aun admitiendo la hipótesis, en principio razonable, de que existe una amplia base común en el vocabulario usado en todos los territorios hispanohablantes (al menos en el registro escrito), no es menos cierto que los resultados de un análisis cuantitativo pueden verse sensiblemente sesgados a causa de la desproporción aquí aludida. Creo que un ejemplo puede ilustrarlo fehacientemente. La palabra *autonomía* aparece registrada en cuarenta y nueve ocasiones, a lo que deben añadirse las seis apariciones de *autonomías*. Por su parte, las formas *federal* y *federales* ofrecen una frecuencia sensiblemente inferior (diecisiete y tres, respectivamente). ¿No resulta lógico pensar que dicha diferencia viene motivada por el régimen político actual del estado español, y a las alusiones que al mismo se hacen en los medios de prensa españoles? Me inclino a pensar que sí, y que si en la selección de textos hubieran primado los procedentes de un estado federal (por ejemplo, México), la proporción habría sido muy diferente.

En otro orden de cosas, también es un muy serio inconveniente, compartido por ambos diccionarios, la falta de un etiquetado real de las unidades contempladas. Eso provoca, en concreto, la falta de reconocimiento de los fenómenos de homonimia, homografía y polisemia, que son fundamentales para la aplicación didáctica. Porque, ¿de qué nos sirve conocer la frecuencia total de una palabra como *partido*, si no se nos informa del peso proporcional que tiene en la misma cada uno de sus valores: 'competición deportiva', 'agrupación política' o 'participio de *partir*' (con las diferentes acepciones de dicho verbo)? De nuevo, la ausencia de tal etiquetado parece disculpable, al menos parcialmente, en el añejo diccionario de Juilland y Chang-Rodríguez, pero no así en el mucho más reciente de Alameda y Cuetos; máxime cuando para lenguas como el francés existen desde hace tiempo tablas de repartición que reparan esa insuficiencia (cf. INALF, 1984). Paralelamente, tampoco resulta comprensible que un diccionario de frecuencias publicado en la década de los noventa asigne entradas distintas a todas las variantes morfológicas de una misma unidad léxica (incluidos los incrementos prepositivos: *asfixiar*, *asfixiarlas* y *asfixiarme* figuran como vocablos diferentes), lo cual es algo que tergiversa gravemente el cálculo. Por esto último, hay que des-

cartar por irreal la cifra de vocablos distintos (ochenta y un mil trescientos veintitrés) que, según los autores, incluye este repertorio (Alameda y Cuetos, 1995: 116). Si se agrupan convenientemente las variantes morfológicas de cada lexía, dicho número se vería mermado en una proporción más que considerable⁶.

Creo que las insuficiencias constatadas en el diccionario de Alameda y Cuetos pueden encontrar una explicación, al menos parcial, en una circunstancia que considero muy relevante. Ambos investigadores proceden del campo de la psicología, no de la lingüística. A mi entender, eso motiva el que, aunque hagan un correcto uso del utillaje estadístico exigido para el cálculo de frecuencias, desatiendan muchos de los más elementales fundamentos de la técnica lexicográfica, la cual es de capital importancia si lo que realmente se pretende elaborar es un *diccionario*, en el pleno sentido de la palabra, y no un mero listado de ocurrencias. Desde esta perspectiva, la no agrupación de variantes morfológicas bajo una misma entrada o la omisión de las referencias a la categoría léxica son sólo la punta del iceberg dentro de todo un cúmulo de carencias.

2.2.4. Como conclusión, es obligado admitir que aún no disponemos de un diccionario de frecuencias que pueda ser usado, con garantías, como base para una selección objetiva de los contenidos léxicos en la enseñanza del español como lengua extranjera. La obra de Juilland y Chang Rodríguez (1964) y sus numerosos precedentes están ya superados, y hoy día apenas si tienen más interés que el meramente historiográfico. En cuanto al repertorio de Alameda y Cuetos (1995), creo haber puesto de relieve que no resiste la más mínima comparación ante diccionarios dedicados a otras lenguas, que, como el de Rafel i Fontanals (1996-1998), se muestran mucho más solventes desde el punto de vista lexicográfico.

3. El léxico usual

3.1. Por lo que se refiere al léxico usual, cabe hablar de él a la hora de referirse a las palabras que son comúnmente manejadas por los hablantes de un idioma. Una apreciación superficial podría hacer creer que léxico frecuente y léxico usual son una misma cosa. De hecho, a menudo se pasa por alto esta precisión terminológica, que yo estimo muy necesaria. La frecuencia se asocia a la lengua, en tanto que instrumento de comunicación; el uso emana de los hablantes, en tanto que responsables de los actos de comunicación. El cálculo de frecuencias fija una medición que en principio se postula como general y unitaria, haciendo abstracción de la variabilidad sociolingüística y la consiguiente diversidad vernacular existente entre los distintos miembros de una comunidad idiomática. Ciertamente, el léxico usual incorpora sin duda una considerable proporción de unidades comunes a todos los hablantes, con escaso margen de variación. Esas unidades se corresponden esencialmente con el léxico frecuente que se documenta en un corpus verdaderamente representativo. Eso es lo que da lugar a la relación de inclusión a la que aludí en el epígrafe 1.3. Ahora bien, existe igualmente otra parte del léxico usual, más o menos amplia según los casos, que ya no mostrará tal coincidencia con el léxico frecuente, y que tenderá a ofrecer unos conside-

6 En cambio, la cifra de lemas diferentes del diccionario catalán dirigido por Rafel i Fontanals (1996-1998), que suma ciento veintidós mil seiscientos cuarenta y siete, sí resulta totalmente creíble, pues se establece tras un correcto etiquetado de las palabras del corpus.

rables márgenes de variación entre los hablantes en virtud de factores sociales tales como el lugar de procedencia, el sexo, la edad, la especialización académica o profesional, etc. Es precisamente este segundo grupo de palabras de aparición más dispersa, habida cuenta de las características de los individuos, el que marca la especificidad del léxico usual respecto del léxico frecuente. En última instancia, cabe admitir que un vocablo será tanto más frecuente en una lengua cuanto mayor sea el número de hablantes que lo tengan como parte de su léxico usual. Por todo ello, no es de extrañar que la investigación lexicométrica actual, a la hora de establecer una medición estadística estándar, haya definido el índice de uso como el producto de multiplicar la frecuencia por la dispersión (López Morales, 2003: 82)⁷. Frecuencia y uso coincidirán únicamente en los *corpora* idiolectales (extraídos de un único informante), normalmente de escaso interés para la lingüística. En cuanto el corpus manejado resulte representativo de todo un colectivo hablante (ya sea éste la comunidad idiomática en su conjunto, ya una parte de la misma, reconocible socialmente como tal), dicha coincidencia deja de existir.

No faltan testimonios que constatan la asimetría entre léxico frecuente y léxico usual. Siquiera sea a título de anécdota, cabe traer a colación una actividad incluida por McCarthy y O'Dell (1994: 2) en su manual de enseñanza de vocabulario. En ella se pregunta a los estudiantes acerca del número de palabras que maneja como promedio un hablante de inglés, así como cuántas son las palabras que constituyen el 45% de lo que se escribe en dicha lengua. La primera pregunta hace referencia a lo que aquí denomino léxico usual, la segunda, al léxico frecuente. La clave de los ejercicios da como correctas (lamentablemente, sin indicar las fuentes) las siguientes respuestas: *cinco mil* para la primera pregunta, y *cincuenta* para la segunda. Ello significaría nada menos que un diferencial de cien a uno entre el léxico usual y el frecuente. En términos más moderados, Pedro Benítez (1994: 10), haciéndose eco de trabajos anteriores, señala, por una parte, que las mil palabras más frecuentes del español abarcan aproximadamente el 85% del contenido léxico de cualquier texto no especializado; por otra, que los hablantes nativos hacen uso de una parte relativamente pequeña del caudal léxico de su lengua, de acuerdo con una horquilla que puede ser situada entre los poco más de dos mil vocablos que maneja lo que Benítez denomina el *hombre de la calle* (cualquier cosa que esto sea) y las cinco mil de los hablantes dedicados a actividades intelectuales. Cifras similares maneja Humberto López Morales (2003: 80). De tales datos se deduce, en primer lugar, que el léxico usual desborda al léxico frecuente, y, en segundo lugar, que mientras el cómputo del léxico frecuente se postula de acuerdo con un patrón único, la medición del léxico usual está sujeta a fluctuaciones (por la incidencia estadística de la dispersión), en función del tipo de hablante que sea tomado como referencia.

3.2. Mientras que la investigación de la frecuencia léxica se halla circunscrita a una línea de trabajo en general bien delimitada (la lingüística de corpus y la elaboración de diccionarios de frecuencias), no cabe decir lo mismo a propósito del léxico usual, para el que cabe apreciar diversas vías de acercamiento según la orientación y objetivos de los distintos investigadores. Quizá ello sea debido a que en esta parcela los datos cualitativos sean, como mínimo, tan relevantes como los estrictamente cuantitativos. El análisis de muestras

7 La fórmula estadística que permite hallar el índice de dispersión aparece especificada en Juilland *et al.* (1970: lii).

extraídas de un único individuo tiene normalmente, como ya he señalado en el epígrafe 3.1, muy escasa relevancia desde el punto de vista lingüístico. Pero para ciertas parcelas de los estudios literarios (particularmente, la estilística) sí puede ser interesante una indagación de este tipo a propósito de la obra de un determinado autor. Ello es lo que da lugar, por ejemplo, a la aparición de estudios lexicométricos como el que Bellón (1973) llevó a cabo a partir de la poesía de Luis Cernuda. Desde otro punto de vista, la llamada *lingüística de campo* (que engloba la dialectología, la geografía lingüística, la lingüística antropológica, etc.) ofrece desde hace tiempo interesantes aportaciones en lo relativo al conocimiento cualitativo de la diversidad diatópica y diatócnica del vocabulario. Sin embargo, ha sido el desarrollo de la sociolingüística a lo largo de las últimas décadas lo que ha permitido disponer de un soporte teórico y metodológico más ajustado para el conocimiento de los fenómenos de variabilidad lingüística, especialmente en el marco de los estudios de estratificación. Por último, y circunscrita básicamente al terreno de la variación diatócnica (derivada de la especialización académica y profesional de los hablantes), la investigación en terminología está ofreciendo también resultados de primer orden.

La orientación sociolingüística y la terminológica son las que a mi modo de ver, y de modo complementario, ofrecen mayores posibilidades de cara a la aplicación en el ámbito de la selección y secuenciación didáctica de los contenidos léxicos. A grandes rasgos, considero que la perspectiva sociolingüística se revela más adecuada para el diseño y aprovechamiento de *corpora* textuales con vistas a la elaboración de vocabularios de carácter general. Especialmente a la hora de atender aspectos tales como su representatividad y su tratamiento estadístico. La perspectiva terminológica, por su parte, resulta especialmente fructífera en la elaboración de vocabularios de especialidad, sumamente valiosos para la enseñanza con fines específicos.

3.3. Los recursos disponibles para el reconocimiento del vocabulario usual pueden ser habitualmente etiquetados con el marbete de *vocabularios básicos*, si bien bajo el mismo suelen aparecer obras bastante heterogéneas. Para la elaboración de estos vocabularios básicos se puede seguir un procedimiento objetivo (de base lexicoestadística) o subjetivo (a partir de la competencia e intuiciones del redactor). Tampoco falta un tercer cauce que yo llamaría *indirecto*: se adaptan y reajustan repertorios previamente existentes (en el mismo idioma o en otros) y luego se perfilan, si se estima necesario, con materiales propios de naturaleza objetiva y/o subjetiva. Así ocurre, por ejemplo, en el breve pero relevante trabajo del Equipo “Rosa Sensat” (1978), pensado para la enseñanza de la lengua materna a escolares de seis a nueve años.

Como testimonio del uso de procedimientos subjetivos pueden tomarse en consideración los listados léxicos que figuran en los volúmenes emanados del proyecto *Nivel Umbral*, que auspició el Consejo de Europa a partir de los años setenta. Como se sabe, el volumen correspondiente al español corrió a cargo de Peter J. Slagter (1979). Pese a las lógicas reservas que despiertan tales métodos subjetivos, lo cierto es que al menos en este caso concreto el resultado final fue bastante plausible. La selección del vocabulario, aunque recaía en última instancia en una decisión personal del autor, se efectuaba a partir de unas pautas previas muy bien delimitadas. Pautas que derivaban de un análisis nocional-funcional de los diversos dominios de discurso y de las situaciones comunicativas que, en términos de necesidades objetivas, tendrían que formar parte de la competencia efectiva del aprendiz. Como prueba

de la validez didáctica del trabajo de Slagter cabe argüir el amplio provecho que obtuvo de él el Equipo "Pragma" (1983), cuyo *Para empezar* marca, a mi modo de ver, un antes y un después en la historia de los manuales de enseñanza de español como lengua extranjera.

Por lo que se refiere a los vocabularios básicos obtenidos mediante procedimientos objetivos, es obligado admitir que responden a una línea de trabajo totalmente análoga y compatible con la seguida en la redacción de diccionarios de frecuencias. Al fin y al cabo, parten también de un corpus que se quiere suficientemente representativo. En teoría, la diferencia entre uno y otro tipo de repertorios estribaría en que mientras en un caso el dato estadístico más relevante sería el índice de frecuencia, en otro primaria el índice de uso. Sin embargo, dicha precisión es sumamente puntual y, en la práctica, baladí: nada impide que un diccionario de frecuencias ofrezca un cálculo del índice de uso (de hecho, es muy de agradecer⁸), como tampoco hay nada que impida que un vocabulario básico adopte como criterio de selección el índice de frecuencia, cuya obtención es previa a la del índice de uso. Por consiguiente, la cesura entre una y otra herramienta se difumina en buena medida, y, como ya he indicado, la obra de Juilland y Chang-Rodríguez (1964) podría ser entendida como un caso fronterizo.

A pesar de todo, considero que sigue habiendo elementos de juicio suficientes, siquiera sea en términos más operativos que teóricos, para mantener justificadamente la distinción entre ambas. En concreto, aprecio que los diccionarios de frecuencias aspiran a ofrecer un cómputo amplio y, en lo posible, general de las unidades léxicas de la lengua, mientras que los vocabularios básicos, como su nombre indica, tienden a aplicar unos criterios más restrictivos en la inclusión de unidades, que llevan aparejados un umbral estadístico más exigente. Del mismo modo, una notable proporción de vocabularios básicos acotan de antemano la variedad de lengua analizada. Aunque no contemplan las precisiones diatécnicas propias de los vocabularios de especialidad (por ejemplo, *léxico jurídico*), si es corriente que existan otras de índole diatópica, diastrática o diafásica (tales como: *vocabulario de Puerto Rico*, *vocabulario infantil*, *vocabulario coloquial*, etc.). Además, la elaboración de los vocabularios básicos suele venir ya marcada, desde el principio, por una finalidad didáctica, que en el caso de los diccionarios de frecuencias puede ser contemplada entre otras varias, o adoptada *a posteriori*, pero que rara vez constituye su objetivo primordial. Por último, y como ya he apuntado en otro momento, en los vocabularios básicos la información cualitativa adquiere una mayor relevancia. Lo importante es qué palabras han sido incluidas en el repertorio (y, por la recíproca, cuáles han sido excluidas). Todo lo más, y dependiendo de los objetivos concretos de la obra, cabe fijar una jerarquía por niveles de competencia o grados académicos. Pero, dentro de cada punto de la escala (por ejemplo, nivel inicial) todas las palabras son consideradas por igual, sin que en la práctica quepa reconocer un orden de prelación rígidamente sometido a los gradientes de uso o frecuencia. Así pues, la información cuantitativa interesa sólo como patrón clasificatorio. Las precisiones lexicoestadísticas que vayan más allá de eso pueden pasar a un segundo plano, e incluso omitirse por completo.

Algunos de los repertorios mencionados en los epígrafes relativos al léxico frecuente pueden ser incluidos también aquí como acercamientos, más o menos aproximados, al vo-

8 Y no sólo el cálculo, sino también secciones específicas en las que los lemas aparezcan ordenados de acuerdo con los gradientes de dispersión y de uso. De nuevo, el diccionario catalán dirigido por Rafel i Fontanals (1996-1998) se muestra como un ejemplo a seguir.

cabulario usual de nuestra lengua, pues, como ya he advertido anteriormente, la separación entre uno y otro tipo de herramientas distan mucho de ser impermeables. Lo dicho a propósito del carácter fronterizo de Juilland y Chang Rodríguez (1964) puede sostenerse también, *mutatis mutandis*, cuando nos acercamos a otros repertorios anteriores, y en particular a los de una más clara vocación didáctica. Por ejemplo, Buchanan (1927), Eaton (1940) o García Hoz (1953). De entre los publicados con posterioridad, quizá el dirigido por Márquez Villegas (1975) sea el que va más directamente encaminado hacia la enseñanza del español como lengua extranjera, si bien, desde una perspectiva actual, los presupuestos teóricos y metodológicos que lo inspiraron merecen ser puestos en cuarentena.

A mi modesto entender, de entre los vocabularios básicos de naturaleza objetiva, el más riguroso y mejor fundamentado es el que llevó a cabo Amparo Morales (1986) a propósito del español de Puerto Rico. Sería muy deseable que sirviera de estímulo e inspiración para repertorios análogos en otros territorios del dominio lingüístico hispánico. Claro que, incluso en un trabajo tan encomiable como el aludido, es posible advertir la sorprendente ausencia de palabras que difícilmente podemos considerar ajenas al vocabulario usual de cualquier hispanohablante nativo. Así, Bartol (2002: 64) constata la no aparición en el mismo de los vocablos *bicicleta*, *cuchara*, *lápiz*, *queso* y *uña*. ¿Se ha debido eso a una insuficiencia en el diseño o el desarrollo de la investigación? No lo creo así, aunque, claro está, todo es mejorable. Más bien me inclino a pensar que estos hechos denotan algo que va más allá del mero control de calidad de un repertorio concreto. Constituyen una evidencia directa de los límites a los que se enfrenta toda selección de contenidos léxicos derivada exclusivamente de muestras de actuación lingüística. Nos hallamos, en fin, ante el dilema entre competencia y corpus, que tan arduos debates ha provocado en el seno de la lingüística teórica desde Chomsky (1965). Una búsqueda de procedimientos lexicométricos objetivos más allá del horizonte impuesto por un corpus de partida nos aboca al tercero de los parámetros consignados al comienzo de este texto: la disponibilidad léxica.

4. El léxico disponible

4.1. El término *disponibilidad léxica* aparece en el seno de la lingüística francesa a mediados del siglo XX, si bien cabe rastrear nociones análogas en la psicolingüística anterior. Por léxico disponible hay que entender el conjunto de unidades léxicas que un hablante conoce realmente, independientemente de su frecuencia o del grado de uso que tengan. Se trata, en definitiva, del conjunto de palabras que se encuentran presentes en el lexicón mental del individuo. Son recuperables durante el proceso de producción lingüística y comprensibles durante el de percepción.

La necesidad de tomar en consideración este léxico viene dada por la evidencia de que los usuarios de una lengua conocen muchas más palabras que las que manejan habitualmente, y la diferencia se acentúa si las confrontamos con las verdaderamente frecuentes. Para corroborar esto, pueden compararse los testimonios aludidos por Pedro Benítez (1994: 10), comentados en 3.1, con los que ofrece Aitchison (1987: 7). Esta última autora, tras evaluar investigaciones precedentes para el inglés, afirma que el número de palabras conocidas por un hablante adulto con cierto nivel educativo difícilmente puede ser inferior a cincuenta mil (sumando formas básicas y derivadas), y puede alcanzar incluso las doscientas cincuenta mil. Quizá esta última cifra resulte excesiva y deba matizarse en virtud de un análisis de la

validez y fiabilidad de los cálculos. Pero, incluso rebajando en mucho tales estimaciones, las cantidades que postulan las investigaciones psicolingüísticas más plausibles siguen siendo considerables. Pérez Basanta (1999: 274) recoge la cifra de diecisiete mil quinientos como la media de vocablos que conoce un hablante de inglés de dieciocho años⁹. Este dato, como veremos, es significativo, pues atañe a sujetos de la misma edad que los encuestados en el proyecto panhispánico de disponibilidad léxica. 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.

Como confirmación de lo expuesto, voy a proponer a continuación un ejemplo quizá un tanto pedestre, pero que a mi juicio resulta suficientemente clarificador. Solicito al lector de estas páginas que tenga a bien pensar en la unidad léxica *trapezio*. Los datos objetivos permiten asegurar que la misma no forma parte en modo alguno del léxico frecuente de la lengua española; una consulta al correspondiente diccionario de frecuencias o el rastreo por cualquier corpus medianamente representativo confirman fehacientemente tal aserto. Del mismo modo, resulta plausible postular que tampoco forma parte del vocabulario usual de lo que en términos estadísticos cabría reconocer como un hablante estándar de nuestra lengua. Sólo alcanzaría el rango de palabra usual en aquellos individuos que, por profesión o afición, tratan habitualmente de cuestiones relacionadas con la geometría o con los espectáculos circenses. Si el lector no entra dentro de tal colectivo, le ruego intente recordar la última ocasión en que escuchó o emitió dicho vocablo. ¿Hace meses?, ¿años, quizá? Y, a pesar de todo esto, estoy convencido de que conoce perfectamente su significado, como lo conoce la práctica totalidad de hispanohablantes con un mínimo de formación y experiencia. Por lo tanto, se trata de un vocablo que se encuentra *disponible* en el lexicón mental; aparentemente dormido, pero dispuesto para ser recuperado en cuanto la ocasión lo requiera. Lo mismo ocurre con varios miles de unidades análogas a la aquí seleccionada. Gracias a esa amplitud del volumen de léxico disponible, el hablante nativo se encuentra capacitado para afrontar y resolver satisfactoriamente no sólo las situaciones de comunicación más comunes y cotidianas, sino también las que tienen un carácter excepcional.

4.2. El estudio de la disponibilidad léxica cuenta con la particularidad de que, a diferencia de lo que ocurre con el léxico frecuente y el usual, no puede emprenderse a partir de un corpus (éste necesitaría unas dimensiones tales que lo harían científicamente inabarcable). La disponibilidad se manifiesta, pues, como un fenómeno de competencia que difícilmente adquiere una plasmación cabal en la actuación espontánea. Por ello, su medición cuantitativa ha de acometerse por medio de procedimientos alternativos, entre los que destacan los *tests* y los cuestionarios. Al respecto, creo que merece ser reconocido como un avance pionero el *test* diseñado en su día por Seashore y Eckerson (1940) para el reconocimiento de vocabulario en lengua inglesa. En el ámbito francófono, y desde postulados metodológicos diferentes, Michéa (1953) y, sobre todo, Gougenheim (1967) sentaron las bases de lo que ha sido buena parte de la investigación posterior en disponibilidad léxica, que pronto se

9 Por su parte, Goulden, Nation y Read (1990) estiman que un estudiante universitario, hablante nativo de inglés, domina una media de diecisiete mil 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 veintisiete mil doscientas lexías diferentes.

extendió a Canadá y otros países. La plasmación cualitativa de los datos de disponibilidad, combinados con los de frecuencia, quedó reflejada en la segunda edición del diccionario destinado a recopilar el vocabulario fundamental¹⁰ del francés, cuya redacción corrió a cargo de Gougenheim (1971).

En esta misma línea, contamos actualmente para el español con un ambicioso proyecto (en adelante, *proyecto panhispánico*) que indaga en la disponibilidad léxica de los jóvenes preuniversitarios de diversos territorios de todo el dominio hispanohablante. Bajo la coordinación de Humberto López Morales, sin duda su principal impulsor (cf. López Morales, 1978, 1997; López Morales y García Marcos, 1995), este proyecto, aunque se encuentra todavía en curso, ha ofrecido ya resultados muy importantes. La Universidad de Salamanca ha habilitado en su sitio de internet una página (<http://www3.usal.es/dispolex>) dedicada específicamente al mismo, que ofrece igualmente información sobre otros trabajos de disponibilidad relacionados con el español, aunque no directamente vinculados al proyecto panhispánico.

4.3. La voluntad aplicada de los estudios de disponibilidad léxica a la enseñanza de lenguas resulta incuestionable. De hecho, el diccionario de francés fundamental, antes aludido, va destinado ante todo a usuarios no francófonos. En el marco del proyecto panhispánico, López Morales se ha interesado fundamentalmente por las implicaciones de estos estudios en lo relativo a la enseñanza de la lengua materna, mientras que otros autores, entre los que se cuentan Benítez (1994) y Bartol (2002), han ponderado sus posibilidades con vistas a la enseñanza del español como segunda lengua. En los últimos años, han aparecido ya trabajos que, como los de Carcedo (1998, 1999), han utilizado este procedimiento de análisis para estudiar la competencia léxica de los estudiantes de español como lengua extranjera.

No obstante lo dicho, quizá sea conveniente tomar en consideración el hecho de que las investigaciones comentadas, por sus propios condicionantes metodológicos, no dan cuenta de la totalidad del léxico disponible, sino sólo de una parte (relativamente pequeña, me atrevería a decir). Y eso limita ineludiblemente sus posibilidades didácticas. El *test* de Seashore y Eckerson utiliza como estímulo una muestra de entradas léxicas extraídas de un diccionario de lengua. Tras el correspondiente tratamiento estadístico, los resultados obtenidos pueden ser extrapolados al total del universo considerado (el diccionario en su conjunto), pero sólo en términos cuantitativos, no cualitativos. Es decir, podemos calcular cuántas palabras conoce un hablante, pero seremos incapaces de precisar cuáles son. La información cualitativa se limita a las propias unidades seleccionadas en la muestra. Una apreciación similar es válida para otros procedimientos psicolingüísticos de cálculo diseñados con posterioridad¹¹. Además, hay sobradas evidencias de que estas pruebas de reconocimiento dan lugar a unas estimaciones que adolecen de una excesiva dispersión, en virtud de decisiones metodológicas tales como el número total de *items* administrados a los sujetos (cf. Justicia, 1995: 8 y ss.).

10 El vocabulario fundamental, en tanto que correlato de lo que en la bibliografía anglosajona se conoce como *nuclear vocabulary* o, más recientemente, *core vocabulary*, estaría formado por todas aquellas palabras que, en un idioma dado, pueden ser consideradas como plenamente neutras desde el punto de vista pragmático y sociolingüístico (cf. Stein, 1978; Quirk, 1981).

11 Para el español, existen investigaciones de este tipo al menos desde García Hoz (1946).

Por lo que respecta al método seguido por Gougenheim, que es el adoptado en el proyecto panhispánico, cabe decir que parte de un cuestionario destinado a ofrecer una serie de respuestas asociativas a partir de un estímulo de partida. Se pide al encuestado que escriba, para cada estímulo, tantas palabras cuantas sea capaz de recordar, con una limitación de tiempo de dos minutos. Este procedimiento de listas abiertas me parece acertado, si bien otras investigaciones de disponibilidad optan por listas cerradas, solicitando un número fijo de palabras para cada estímulo. Por ejemplo, en el trabajo de Azurmendi (1983) se operaba sobre listas de veinte unidades. En cualquiera de los casos, la información obtenida es tanto cuantitativa como cualitativa: obtenemos palabras concretas y realmente disponibles en la mente del hablante, las cuales además pueden ser tabuladas de acuerdo con el correspondiente tratamiento estadístico. Por todo ello, y en una primera valoración, este procedimiento es muy superior al de Seashore y Eckerson de cara a su aplicabilidad didáctica. Ahora bien, tal aplicabilidad se ve amenazada por otros graves inconvenientes emanados del propio diseño de las investigaciones sobre disponibilidad. Y ello porque, aunque la calidad y validez de las respuestas resultan satisfactorias, no creo que pueda decirse lo mismo respecto de los estímulos de partida. Me explicaré con más detenimiento.

Gougenheim fijó un total de dieciséis estímulos, que suelen denominarse con el discutible término de *centros de interés* (yo, personalmente, habría preferido una alternativa como *ámbitos nocionales*). Los centros de interés pueden definirse, en principio, como macrodominios semánticos que agrupan diversos campos léxicos. Pues bien, tales centros de interés no focalizan su atención sobre el conjunto del caudal léxico del idioma, sino que se concentran, prácticamente en su totalidad, en palabras que pertenecen a la categoría léxica sustantivo. Los adjetivos y los adverbios quedan completamente preteridos¹², mientras que los verbos apenas si pueden tener una tímida presencia gracias al centro número trece (“trabajos del campo y del jardín”), donde, si tomamos como referencia los datos de la provincia de Almería, los primeros puestos los ocupan los verbos *regar* y *podar* (cf. Mateo, 1998: 185). La atención tampoco se extiende por igual a todos los sustantivos, sino sólo a aquellos que designan entidades concretas; la competencia léxica relativa a los nombres abstractos o de eventos, con muy puntuales excepciones, permanece opaca a esta encuesta. Eso significa, por ejemplo, que no tienen cabida palabras como *vez* y *vida*, que son los dos sustantivos más frecuentes del español de acuerdo con los datos de Alameda y Cuetos (1995: 590), en cuyo diccionario ocupan los puestos cincuenta y dos y sesenta y cuatro, respectivamente. Al margen de esto, también es obligado constatar que la caracterización de los dieciséis centros de interés resulta, cuando menos, irregular. Así, existe un centro específico para animales (el número catorce), pero no para plantas, que es un ámbito que queda disuelto entre varios es-

12 La heterogeneidad de la categoría *adverbio*, en su caracterización tradicional, y los debates que ello provoca aún en la actualidad acerca de su carácter léxico o gramatical, es sin duda un serio inconveniente a la hora de acometer su tratamiento lexicoestadístico. En cualquier caso, creo que al menos los adverbios de modo pueden ser reconocidos como unidades léxicas de pleno derecho. Y el que la inmensa mayoría de los mismos sean en español formas derivadas no puede servir de justificación para este olvido, pues según eso deberían también ser excluidos de los recuentos todos los sustantivos derivados, que no son pocos. Aunque la relevancia estadística de los adverbios de modo no sea comparable a la de las demás categorías léxicas, tampoco resulta desdeñable. De hecho, Justicia (1995) documentó un buen número de ellos entre los escolares de Andalucía oriental. Entre los que alcanzaban un peso estadístico de primer orden no sólo se encontraban los archifrecuentes *bien* y *mal*, sino también otros como *despacio*, *solamente* y *aproximadamente*, con más de veinte apariciones absolutas cada uno.

tímulos; frente a centros sumamente amplios y heteróclitos, como “la ciudad” y “el campo” (números diez y once, respectivamente), aparecen otros mucho más restringidos, entre los que sobresale el número seis: “objetos colocados sobre la mesa para la comida”. Sospecho que este último fue contemplado por Gougenheim como consecuencia de ciertas especificidades de la cultura francesa, difícilmente extrapolables a otras sociedades. Dado que el tiempo que se ofrece a los encuestados es el mismo en todos los casos (dos minutos), es legítima la sospecha de que en los centros más amplios ese tiempo resultará insuficiente para poder reflejar todas las lexías que forman parte del léxico activo del encuestado, mientras que, en los más restringidos, puede incluso resultar excesivo para tal fin, ya que permitiría actualizar unidades comparativamente poco relevantes.

Es de lamentar que el proyecto panhispánico haya adoptado de forma casi literal la secuencia de centros de interés del proyecto francés, sin haber procedido a los necesarios reajustes. Reajustes que sí llevó a cabo, por ejemplo, Azurmendi (1983), aunque de acuerdo con unos criterios establecidos básicamente *ad hoc*. En el proyecto panhispánico, la única modificación apreciable de la que tengo constancia (y que, según mis noticias, se introdujo a partir de la investigación en Las Palmas de Gran Canaria) se encuentra en el número nueve; a su formulación original (“calefacción e iluminación”) se le añadió la apostilla “medios para airear un recinto”, ante la evidencia de que en muchos lugares del dominio lingüístico hispanohablante las necesidades de clima interior se orientan hacia la refrigeración y no hacia la calefacción. Al margen de esto, algunos investigadores han tomado la iniciativa de incorporar en su trabajo algún centro de interés adicional, pero, probablemente debido a que el proyecto se encontraba por entonces en una fase relativamente avanzada, tales iniciativas no se han generalizado. Así, Mateo (1998) incluye acertadamente, en su estudio sobre la provincia de Almería, dos nuevos centros: “los colores” (que abre una pequeña puerta a los adjetivos) y “la mar” (imprescindible en un territorio litoral).

4.4. Quiero dejar meridianamente claro que las precisiones que estoy introduciendo en estos párrafos a propósito de los estudios de disponibilidad, al igual que las ya señaladas a propósito de los diccionarios de frecuencias, en absoluto pretenden minimizar su importancia e interés, que por otro lado van más allá de su aprovechamiento para la enseñanza. Simplemente quieren poner de relieve las enormes dificultades a las que, se quiera o no, han de enfrentarse los acercamientos objetivos al componente léxico de las lenguas. Dificultades que condicionan su proyección aplicada, y que aparecen redobladas en el caso de la disponibilidad por lo que tiene de acercamiento empírico a una realidad no directamente observable: el lexicón mental de los hablantes. Constituye un mérito innegable de estos estudios el haber conseguido afrontar ese reto de manera decidida y con garantías de éxito por lo que se refiere a ciertas parcelas del vocabulario. Pero, y aquí se encuentra la piedra de toque de mis argumentaciones, ese éxito dista mucho de ser ampliable a la competencia léxica en su globalidad. Si acaso he hecho excesivo hincapié en esa circunstancia es porque la considero una llamada de atención muy conveniente que, sinceramente, suelo echar en falta en las publicaciones dedicadas a esta línea de investigación. Pareciera en ellas como si se diera por sentado que los centros de interés tomados en consideración completasen a efectos prácticos todo el espectro de dominios temáticos y discursivos pertinentes para establecer (de forma combinada con los datos de frecuencia) el léxico fundamental de una lengua. Sin embargo, insisto una vez más en que los datos obtenidos a partir de tales centros

suponen solamente una muestra (no representativa, sino selectiva) del verdadero volumen del lexicón mental.

A poco que se indague en las posibilidades designativas de un idioma, emergen ante nosotros nuevos ámbitos nocionales no contemplados inicialmente por Gougenheim y que habrían podido formar parte, legítimamente, de las encuestas de disponibilidad léxica. Piénsese en las siguientes etiquetas a las que podría añadirse un largo etcétera: “relaciones sociales: parentesco, amistad...”, “clima y fenómenos meteorológicos”, “enfermedades”, “el tiempo (cronológico) y sus intervalos”.

La omisión de verbos, adjetivos y adverbios en las encuestas de disponibilidad tampoco me parece justificada. El argumento a veces esgrimido (cf. Bartol, 2002: 63 y ss.) de que la investigación sobre la frecuencia da poca cobertura a los nombres que se refieren a cosas concretas es cierto en el momento actual de primacía de los *corpora* escritos, pero la solución pasa por profundizar en los estudios de frecuencias a partir de *corpora* orales verdaderamente representativos. A lo sumo, ello haría comprensible una mayor atención proporcional hacia la disponibilidad de los sustantivos (a modo de *discriminación positiva*), pero no la plena exclusividad que se le ha asignado. Es algo que sesga excesivamente cualquier utilización posterior de los datos, tanto en el plano didáctico como en el lexicográfico. Y huelga decir que me niego a admitir, ni siquiera como hipótesis de trabajo, el apriorismo de que el léxico disponible en el lexicón mental de un hablante está formado mayoritariamente por nombres concretos, y muy escasamente por otras categorías léxicas. Es un supuesto que suena demasiado a eso de ‘hacer de la necesidad virtud’.

Tampoco considero suficiente excusa el que el recuerdo de verbos y adjetivos sea más difícil de estimular por medio de una encuesta, ya que no por ello dejan de resultar viables algunos cauces para atenuar, al menos parcialmente, dicha circunstancia. La inclusión del campo del color en el trabajo de Mateo (1998) da testimonio de ello. Otras posibilidades podrían haber sido, para los adjetivos, “los estados de ánimo” (*feliz, triste, etc.*) y, para los verbos, estímulos del tipo “cosas que sueles hacer cuando estás en casa (en la calle, etc.)”. Por último, tampoco conviene perder de vista la existencia de vocablos indiscutiblemente usuales que, por su condición de tabú, difícilmente tienen presencia en un corpus estándar. Al menos cuando se trata de la enseñanza de idiomas dirigidos a adultos, dichos vocablos son necesarios al menos para que los aprendices sepan reconocerlos y eviten su empleo en situaciones inapropiadas. Cabría proponer para los mismos un estímulo como “insultos y palabras que no te atreverías a decir en situaciones formales”.

En resumen, me parece que reconocer las limitaciones de la investigación empírica sobre disponibilidad léxica, y muy especialmente del método seguido en el proyecto panhispánico, no significa en absoluto infravalorar sus logros. Simplemente permite tomar conciencia de qué es lo que se le puede pedir y qué es lo que, en puridad, no se debe esperar de ella. Porque precisamente conociendo con claridad sus límites es como pueden ponderarse sus verdaderas posibilidades.

5. Hacia un replanteamiento de la selección de contenidos léxicos

5.1. Llegados a este punto de la exposición, se diría que el panorama que se nos presenta ante la vista ofrece unos tintes poco alentadores. Los tres parámetros que permiten una selección objetiva de los contenidos léxicos (frecuencia, uso y disponibilidad) resultan

manifiestamente problemáticos, dadas las insuficiencias de los instrumentos con los que contamos hoy día para su correcta delimitación. Sin embargo, sería injusto quedarse anclado en una valoración meramente negativa. Primero, porque las carencias aludidas son simplemente una consecuencia lógica tanto de la complejidad del objeto de estudio como de la falibilidad de la indagación científica, todo lo cual ha de servir de acicate más que de desánimo. Segundo, porque, pese a todos los reparos y matizaciones que correspondan, lo cierto es que los recursos derivados de la investigación cuantitativa del léxico siguen siendo de suma utilidad, lo que significa que el esfuerzo empleado en su elaboración merece cualquier calificativo menos el de baldío. Manejados con las debidas precauciones y en sus justos límites, ofrecen un considerable potencial para la planificación curricular y el diseño de materiales didácticos. Obviamente, a nadie se le oculta que la alternativa cifrada en una introducción meramente aleatoria o intuitiva de los contenidos léxicos constituye a todas luces una senda más insegura y conflictiva. Por decirlo con una fórmula inspirada en las verdades del barquero, cabe afirmar que, por muy deficiente que sea un diccionario de frecuencias, o cualquier otro recurso de selección cuantitativa del léxico, siempre será mejor tenerlo a nuestra disposición que carecer de él.

Procederé a continuación a plantear las vías que a mi modo de ver posibilitarían una proyección positiva de la investigación lexicostatística en el marco de una enseñanza comunicativa de segundas lenguas. Para tal fin, pienso que ha de partirse de tres corolarios que se desprenden de todo lo expuesto anteriormente. Su eje común es la conveniencia de huir tanto de un dogmatismo estéril como de un eclecticismo acrítico y poco riguroso.

a) Dado que ninguno de los tres parámetros de cálculo cuantitativo resulta satisfactorio por sí solo, lo más razonable es proceder a una selección que contemple de forma combinada los tres. Se ha de partir de aquel que se revele como más ajustado a los objetivos didácticos que se persiguen y, sobre esa base, utilizar los otros dos como mecanismos complementarios que subsanen los sesgos o lagunas advertidos en el primero.

b) Los defensores de la limitación objetiva plantean generalmente como horizonte curricular unos contenidos léxicos excesivamente restringidos. Para el caso concreto del francés fundamental, el repertorio cualitativo de referencia contempla unas tres mil quinientas palabras (incluidas las de índole gramatical), según indica su propio responsable (cf. Gougenheim, 1971: 6). Esta cifra se mueve dentro de los márgenes habitualmente calculados para el léxico usual de un hablante medio. Pero ya he advertido a lo largo del presente trabajo que la disponibilidad real desborda ampliamente tales límites. En consecuencia, un currículo que aspire a dar cuenta del grado de competencia propio de los niveles superiores está obligado a barajar un volumen de contenidos léxicos mucho más alto. De hecho, creo que esos planteamientos restrictivos constituyen un resabio conductista heredado de los enfoques didácticos audiolingual y situacional. Como ya observó atinadamente Zimmerman (1997: 11), durante el auge de la enseñanza audiolingual se expresaron claros pronunciamientos en contra de que los aprendices de niveles iniciales aprendieran un *volumen excesivo de vocabulario*. El manual de Rivers (1968: 208 y ss.), uno de los más representativos de aquel periodo, constituye un caso paradigmático al respecto. Los motivos esgrimidos para tal actitud variaban en cada caso, pero remitían, en última instancia, a la idea de que el aprendizaje de ese *vocabulario superfluo* incidiría negativamente sobre el aprendizaje lingüístico en su conjunto, y muy especialmente sobre el de otras vertientes de la lengua tenidas entonces por mucho más importantes: verbigracia, los patrones estructurales. Ni que decir tiene que

los argumentos que avalaban tales prejuicios se encuentran desde hace tiempo totalmente superados, y que una actitud como la descrita se revela totalmente incompatible con los modernos enfoques comunicativos de enseñanza de lenguas.

c) Se hace obligado reconocer que los instrumentos de selección objetiva están lejos de constituir una panacea. De ahí que, sin menoscabo de los requisitos de rigor exigibles a una planificación curricular, hay que mantener amplios márgenes de flexibilidad y admitir también a consideración los criterios de índole subjetiva. A mi modo de ver, dichos criterios encuentran su mejor expresión en el análisis de necesidades, que abren la puerta a un diseño curricular abierto y centrado en el alumno.

5.2. Por lo que se refiere a la primera cuestión, no es de extrañar que investigadores que trabajan en disponibilidad léxica, como Benítez (1994: 12) y Bartol (2002: 64), se pronuncien a favor de este parámetro como el cauce de mayor ayuda para los fines didácticos. Como se desprende de lo dicho hasta ahora, yo me muestro bastante más escéptico. Personalmente, y una vez analizados los diversos acercamientos lexicoestadísticos, me inclino a pensar que el parámetro objetivo más útil para la selección de contenidos léxicos sigue siendo el de frecuencia. Esta opinión se justifica por su mayor rentabilidad, derivada tanto de factores teóricos como de viabilidad material.

En términos de viabilidad, parece innegable que la investigación de la frecuencia se revela a todas luces como mucho más económica. Cualquier persona familiarizada con los estudios de disponibilidad es consciente de la enorme cantidad de recursos materiales y humanos que requiere. El proyecto panhispánico, en concreto, constituye una obra titánica que ha movilizado a numerosos especialistas y que, tras varios años de puesta en práctica, aún se encuentra lejos de ver su culminación. Además, sus características impiden que se puedan introducir, *in media res*, rectificaciones o mejoras en el diseño de la investigación (por ejemplo, una modificación sustancial del listado de centros de interés), pues eso alteraría la confrontación de los datos obtenidos antes y después de la aparición de los cambios. Ante ello, la pregunta que se me viene a la cabeza es: ¿realmente merece la pena? Desde el punto de vista de la lingüística teórica y descriptiva, en campos como la sociolingüística y la lexicología, indudablemente sí, y gracias a ello es de desear que el proyecto panhispánico siga adelante con todos los parabienes. Pero, por lo que respecta a la lingüística aplicada, me atrevo a afirmar que no ofrece la debida correlación entre costes y beneficios. Si se me permite una (odiosa) comparación, diré que encuentro aquí un profundo paralelismo con los estudios de lingüística contrastiva emprendidos en los años sesenta y setenta: proyectos ambiciosos y de costosísima ejecución de los que se obtuvo un rendimiento pedagógico desproporcionadamente pequeño.

Frente a todo lo anterior, los estudios de frecuencia optimizan mejor los esfuerzos y, lo que no es pequeña cosa, ofrecen unos prometedores márgenes de mejora. Es impensable una actualización continua de los datos de la disponibilidad (pongamos por caso, repitiendo cada diez años las encuestas en todos los lugares investigados). En cambio, dicha actualización es perfectamente factible para los datos de frecuencia, y los progresos técnicos agilizan cada vez más tal empeño. Eso quiere decir que, aunque los diccionarios de frecuencias actualmente existentes para el español sean deficientes (que lo son, y mucho), podemos confiar en que los que se elaboren en el futuro resulten más consistentes. Así, el grave problema del reconocimiento de homógrafos y homónimos cuenta ya con instrumentos para su solución

ya la especificación de las unidades pluriverbales también es una posibilidad cada vez más contemplada. En cuanto a las deficiencias de representatividad del corpus, son subsanables incorporando paulatinamente nuevos materiales que atiendan las correspondientes variables sociolingüísticas. Hoy en día, la difusión por internet posibilita la aparición de *corpora* potencialmente abiertos, con continuas adiciones y reajustes. Los datos lexicoestadísticos emanados de tales *corpora* son susceptibles de una actualización regular de manera automatizada y sin especiales exigencias materiales.

Desde el punto de vista teórico, no puede olvidarse la relación implicativa existente entre léxico frecuente, usual y disponible: los cálculos de frecuencias ofrecen información acerca del uso y, siquiera sea de manera parcial e indirecta, también de la disponibilidad. Por supuesto, esa información se puede ver sensiblemente incrementada mediante el trabajo con *sub-corpora* específicos en función de las distintas variables (diatópicas, diastráticas, diatécnicas y diafásicas), con el consiguiente beneficio para la enseñanza orientada a fines específicos. La relación implicativa no funciona en sentido contrario, por lo que los datos de disponibilidad no aportan una información verdaderamente fiable ni del uso ni de la frecuencia. Al margen de esto, el cálculo de frecuencias es aplicable no sólo al léxico conceptual, sino también al nomenclador (antropónimos, topónimos, etc.), el cual forma parte también de la competencia léxica. En última instancia, un umbral de disponibilidad mínimamente exigente nos ofrece unos repertorios léxicos excesivamente reducidos, muy poco operativos por sí solos. Por ejemplo, Bartol (2002: 71) da la cifra de cuatrocientos diecinueve vocablos con un índice de disponibilidad superior a 0'1 en las encuestas de Zamora. Yo he comprobado que en los datos de la provincia de Cádiz (cf. González, 2002: 119-269) se reduce a trescientos noventa y nueve, que oscilan entre diecisiete y treinta y ocho por centro de interés. Si comparamos estos números con las tres mil quinientas entradas del diccionario de Gougenheim (1971), advertiremos que la proporción no llega siquiera al 15%. Así pues, la investigación de la disponibilidad puede ser una herramienta complementaria para perfilar y mejorar los repertorios léxicos obtenidos por otras vías, pero no parece dar lugar por sí sola a recursos pedagógicos autónomos.

5.3. Los datos ofrecidos por los *corpora* lingüísticos parecen mostrar que un número de palabras con altos niveles de frecuencia o de uso es relativamente reducido si lo comparamos con el caudal léxico total de un idioma. Por su parte, las investigaciones psicolingüísticas han puesto de relieve que la competencia léxica de los hablantes nativos es muy superior a la que sería predecible de su actuación. Ante tal dilema, los defensores de la limitación objetiva del léxico han optado generalmente por conceder más crédito a los datos de corpus. Con ello, tienden a abogar por un horizonte de aprendizaje muy restrictivo, que rara vez supera las cinco mil unidades. Lahuerta y Pujol (1993: 117 y ss.) indican cifras no muy alejadas de ésta a propósito de los libros de texto y de lecturas graduadas para la enseñanza del inglés. Creo que una valoración sopesada de los datos de la investigación empírica del léxico obliga a replantear una selección tan restrictiva. Las cifras que maneja Aitchison (entre cincuenta mil y doscientas cincuenta mil) pueden ser desorbitadas, pero la de diecisiete mil quinientas, que señala Pérez Basanta (1999: 274), en absoluto debe ser percibida como una barrera infranqueable para un aprendiz de nivel avanzado.

El objetivo que inspiró las investigaciones de Gougenheim para el francés debe ser reconocido hoy día como un antecedente de lo que posteriormente fue el proyecto Nivel

Umbral. En ambos casos, se buscaba definir un nivel de lengua alejado ya de la mera fase de *supervivencia*, pero aún lejos de la competencia del hablante nativo. Sin entrar en pormenores, cabe identificarlo con un nivel intermedio. Es bajo tales coordenadas donde hay que situar un diccionario como el de Gougenheim (1971), que puede servir de referencia para los niveles inicial e intermedio, pero no para los más elevados. Por otra parte, aun admitiendo que los vocablos incluidos en dicho repertorio constituyan el léxico fundamental de la lengua en cuestión, eso no significa de ningún modo que las palabras ausentes del mismo sean todas ellas *superfluas*.

Está sobradamente demostrado que son precisamente las palabras menos frecuentes, y por tanto las más específicas de un determinado dominio, las que soportan en mayor medida el contenido global de un texto (cf. Arnaud y Savignon, 1997). O sea, que aparecen pocas veces, pero, cuando lo hacen, dejan su impronta. Appel (1996: 397), en consonancia con la tendencia general en los nuevos enfoques didácticos, postula como primer principio de la adquisición del léxico el que la competencia perceptiva precede a la productiva. Es la comprensión la que abre las puertas a la producción. Así que, si se aspira a que los estudiantes alcancen un dominio activo de ese umbral de tres mil quinientas palabras, entonces hay que reconocer que el objetivo de aprendizaje, en términos de *disponibilidad receptiva*¹³, ha de estar varias veces por encima de ese límite. Dicho objetivo deja de parecer una utopía si se tienen en cuenta diversos factores. Uno de ellos es que las unidades léxicas no se almacenan aisladas en el lexicón mental, sino que forman agrupaciones tanto por su forma como por su significado. Los campos léxicos que maneja la semántica estructural parecen mostrarse excesivamente constreñidos a unos vínculos paradigmáticos cerrados. Por ello, creo que resultan más plausibles desde el punto de vista psicolingüístico agrupaciones más proteicas, como las apuntadas en su día por Alinei (1974), y que en la actualidad suelen ser denominadas *constelaciones*: cada palabra constituye un nudo del que parten múltiples conexiones asociativas. Así, y atendiendo sólo al plano semántico, una palabra como *cenicero* se asocia, entre otras, con la palabra *cigarrillo*, y ésta con *cáncer*, que te lleva a *hospital*, y así sucesivamente. Una enseñanza orientada no sólo a la cantidad, sino también a la calidad, ha de promover en el aprendiz el uso de estrategias que lo capaciten para establecer esa clase de asociaciones, que multiplican la operatividad de los contenidos disponibles en el lexicón mental.

Otro factor relevante viene dado por las concordancias etimológicas y morfo-funcionales entre las palabras de la lengua meta y las correspondientes en la lengua materna del aprendiz. Los enfoques tradicionales y conductistas se interesaban por dichas concordancias poco más que cuando eran fuente de error: los cognados o los llamados falsos amigos. Por ejemplo, el español *realizar* y el inglés *realize* ('darse cuenta'). Esa actitud no es sostenible hoy día; los falsos amigos constituyen la excepción, no la regla de las concordancias morfo-funcionales. Yo suelo decir, coloquialmente, que los amigos, mientras no se demuestre lo contrario, son verdaderos, y están para ayudarte. En el caso de las lenguas europeas de mayor difusión internacional, que son las que ocupan buena parte del *mercado lingüístico* contemporáneo, la existencia de un trasfondo cultural común hace que existan varios miles de raíces léxicas

13 Este término me parece preferible al de *disponibilidad pasiva*, que se documenta a menudo en la bibliografía.

compartidas, con un valor semántico esencialmente idéntico. Los tecnicismos y los términos cultos de origen grecolatino, así como los anglicismos que se extienden por doquier en la sociedad contemporánea, son los más abundantes y representativos, pero no los únicos. En este marco, cabe asegurar que, en los contextos oportunos, el significado de palabras como *raqueta*, *chequeo*, *diabetes*, *poema* y otras muchas puede ser fácilmente colegido, de forma espontánea y sin necesidad de instrucción previa, por parte de un amplísimo colectivo de aprendices de español. Es decir, que los estudiantes de nuestro idioma cuentan, ya desde los niveles iniciales, con un volumen de léxico disponible más que considerable. Y eso hay que aprovecharlo.

5.4. Si un manejo combinado de los distintos procedimientos objetivos amplía la validez de la selección de contenidos léxicos, lo mismo vale decir, a mi entender, para el uso combinado de los criterios objetivos y subjetivos. Los procedimientos subjetivos han sido justamente criticados a menudo debido a sus problemas de fiabilidad, y ciertamente parecen poco defendibles como cauce único para una planificación curricular de amplio alcance. Ahora bien, no por ello deben ser rechazados por completo, ya que ofrecen otras contrapartidas que merecen ser tenidas en cuenta. La intuición lingüística de los profesionales de idiomas debidamente cualificados suele ser bastante aguda. De ahí que pueda ofrecer unos resultados prácticos razonablemente satisfactorios y no particularmente alejados de los obtenidos por criterios objetivos. Claro es que no faltan ejemplos a todas luces discordantes, como el que refiere López Morales (2003: 81) a propósito de la inclusión de la voz *conquiliólogo* en el currículo escolar puertorriqueño. Pero, en términos globales, me parece que la incidencia de estos hechos es puramente anecdótica. Por su parte, los procedimientos objetivos no siempre garantizan su erradicación. Valga un botón de muestra. No le falta razón a Benítez (1994: 10) cuando critica en un manual de español para extranjeros el que se actualice la palabra *perdiz* y no su hiperónimo *pájaro*. Ahora bien, en los datos de disponibilidad de la provincia de Cádiz, *pájaro* tiene un índice inferior al de su hipónimo *canario*¹⁴. Asimismo, *escalímetro*, voz que no figura de momento en el Diccionario de la Real Academia, cuenta con una disponibilidad relativamente pequeña, pero que aun así triplica la de una palabra tan común como *enciclopedia* (cf. González Martínez, 2002: 166, 168, 229).

Al margen de lo anterior, creo que no debe olvidarse que la competencia llega a rincones de la lengua que, al menos actualmente, un corpus es incapaz de ilustrar. Por tanto, no me parece descabellado el apelar a la competencia de los expertos, siquiera sea como complemento de los acercamientos objetivos. Esa competencia de los expertos puede hacerse necesaria cuando se trata de seleccionar unidades comunes en el registro coloquial, pero rara vez representadas en *corpora* escritos. Igualmente, puede desempeñar una importante labor cuando entran en juego aspectos que escapan a los análisis cuantitativos al uso. Verbigracia, a la hora de decidir el tratamiento didáctico de ciertas palabras, comunes y no marcadas en el español peninsular, que han adquirido la condición de tabú en muchos lugares de Hispanoamérica: *madre*, *concha*, *bolsa*, *coger*, *venirse*, etc.

14 Este dato, u otros similares, no tienen por qué causar extrañeza. Como explicaré un poco más abajo, lo que se constata aquí es que la mayor o menor rentabilidad de un hiperónimo respecto de sus hipónimos es algo que depende de los datos de uso mucho más que de los de disponibilidad o de frecuencia.

Una sacralización del dogma objetivista es fácilmente esperable en los modelos cerrados de currículo, en los que se procede con un patrón rígido en la secuenciación de contenidos. Una especie de *café para todos* en el que las decisiones curriculares están firmemente selladas, independientemente de las características de los alumnos o del contexto de aprendizaje. Para el léxico, una secuencia típica sería la que estableciera los niveles de vocabulario a partir de sucesivos umbrales de frecuencia, tal y como, de forma simplificada, expongo a continuación:

1. *Vocabulario frecuente.*
2. *Vocabulario académico.*
3. *Vocabulario técnico (especializado).*
4. *Vocabulario poco frecuente.*

Esta gradación tiene su origen en los cálculos lexicométricos llevados a cabo por Nation (1990), y desde entonces se ha manejado frecuentemente en los estudios sobre la enseñanza del vocabulario de la lengua inglesa (cf., por ejemplo, Nation y Newton, 1997: 239; Pérez Basanta, 1999: 274; etc.). El punto de partida es un corpus de inglés académico escrito. El primer bloque, vocabulario frecuente, consta de dos mil vocablos de aparición general (independientemente de las variables diafásicas o diastráticas), y acumula un 87% de las ocurrencias totales del corpus. Es decir, este conjunto de voces formaría parte del vocabulario fundamental de la lengua. El segundo bloque está formado por ochocientos vocablos, que suponen un 8%. Aquí 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 decir, sería un vocabulario marcado diafásicamente como característico del registro formal escrito. En cuanto al vocabulario que Nation denomina *técnico* (que yo preferiría llamar *especializado*), implica a un total de dos mil vocablos, con una incidencia del 3% en el total del corpus. Aquí encontraríamos las unidades léxicas marcadas diastráticamente como testimonio de las nomenclaturas y terminologías propias de cada especialidad académica. El último nivel engloba a más de ciento veinte mil vocablos, que apenas si alcanzan una presencia del 2% en el corpus de referencia. Se trataría de un vocabulario marcado diastráticamente, puesto que, a efectos de producción lingüística, sólo estaría al alcance, y en una proporción variable, de la población con mayor formación cultural o académica.

No niego la utilidad de una secuenciación de esta naturaleza. Al fin y al cabo, he reconocido explícitamente que, dentro de los criterios objetivos para la selección del léxico, el de frecuencia me parece con mucho el más operativo. Tampoco voy a repetir las limitaciones que, pese a todo, siguen lastrando una jerarquización basada exclusivamente en el gradiente de frecuencia, de las cuales me he ocupado ampliamente en páginas anteriores, y sobre las que ya existe una amplia bibliografía. Lo que me interesa destacar ahora es lo que tiene esta secuencia de listado cerrado, concebido como único e invariable para todos los contextos de aprendizaje.

En primer lugar, lo que se observa en la secuencia anterior es que se produce una equiparación práctica entre la rentabilidad didáctica de un vocablo y su frecuencia, lo cual es, cuando menos, reduccionista. La rentabilidad responde a la potencial utilidad comunicativa en las distintas situaciones de interacción. Por tanto, el criterio objetivo más ajustado para su delimitación es el de uso, que ciertamente está condicionado por la frecuencia, pero también

por la dispersión¹⁵. Esta incidencia de la dispersión es la que determina que normalmente, como apuntan entre otros Benítez (1994: 10) o Pastor (2004: 215), un hiperónimo se perciba como más rentable que sus correspondientes hipónimos. Un hipónimo aporta al intercambio comunicativo una información más rica y específica, por lo que será preferible en los contextos que requieran ese grado de concreción. Piénsese, sin ir más lejos, en una situación típica de compraventa en una frutería. Resultaría totalmente inadecuado que el cliente se limitase a solicitar *un kilo de fruta*. En dicho contexto, se hace obligado especificar la clase de fruta que se desea comprar. Por su parte, el hiperónimo cuenta con la ventaja de su mayor versatilidad, pues se adecúa a un mayor número de contextos en los que, en consonancia con las máximas conversacionales de Grice (1975), la utilización de unidades significativas con valor más genérico es suficiente, e incluso preferible, para el correcto progreso del intercambio comunicativo. Por todo ello, no puede postularse una correspondencia biunívoca entre rentabilidad y frecuencia. De nuevo, el ejemplo de la fruta, al que he apelado antes, sirve como ilustración. De acuerdo con los datos de Justicia (1995: 92, 113), la voz *naranja* (como sustantivo), con doscientas sesenta y tres apariciones absolutas, tiene entre los escolares andaluces una frecuencia muy superior a la de su hiperónimo *fruta*, con ciento veintinueve ocurrencias. Y la frecuencia de esta última no es significativamente superior a la de otros hipónimos que también superan el centenar de apariciones: *fresa*, *uva*, *plátano*, etc. Todo ello es perfectamente explicable dadas las condiciones socioculturales en las que se inscribe el corpus de referencia. Una aplicación mecánica del criterio de frecuencia obligaría a situar *naranja* y *fruta* en niveles distintos de la secuenciación del sílabus. Pero esa decisión desatendería el potencial comunicativo del hiperónimo, que se revela como herramienta lingüística básica desde el primer momento en que el aprendiz entra en contacto con un *input* asociado al campo designativo de la fruta.

En otro orden de cosas, la secuencia de contenidos léxicos inspirada en Nation (1990) adolece de otras inadecuaciones si la analizamos desde una perspectiva propiamente comunicativa. Dado que se parte de un corpus de textos escritos situados en un registro formal, propios de una norma idiomática muy concreta, transmite la impresión errónea de que el aprendizaje lingüístico, y particularmente el del vocabulario, debe entenderse fundamentalmente como una actividad académica y, de algún modo, filológica. Ello puede coincidir con las condiciones propias de ciertos contextos escolares de aprendizaje, pero en absoluto puede elevarse a patrón general. De hecho, todo apunta a que la demanda de lenguas modernas en la sociedad contemporánea apunta mayoritariamente hacia otros derroteros. Por tanto, situaciones ajenas a la prevista inicialmente requerirían una atención a muestras de lengua de diferente procedencia, con la correspondiente adaptación del sílabus. Ahora bien, la secuencia descrita deja, en la práctica, muy escaso margen de maniobra. Se permite, en el tercer nivel, la concreción de los contenidos correspondientes a cada especialidad técnica, y eventualmente queda abierta la posibilidad de sintetizar niveles¹⁶. Pero no está contemplada una alteración del orden lineal de las distintas fases. Según eso, el vocabulario infrecuente

15 A mi modo de ver, la delimitación entre el vocabulario académico y el vocabulario especializado ha de basarse en el índice de dispersión, que debe ser sustancialmente más elevado en el primero de ellos. Sólo subsidiariamente se puede apelar a la frecuencia.

16 Verbigracia, en contextos de aprendizaje cuyos objetivos no contemplen ni una finalidad profesional específica ni una atención prioritaria al lenguaje académico, cabe clasificar el vocabulario únicamente en dos grupos: frecuente e infrecuente.

siempre debe figurar en el último lugar de prelación. No se contempla la posibilidad de que parte de ese vocabulario vaya asociado a dominios discursivos acordes con los gustos o aficiones de los estudiantes, y que estos podrían estar interesados en otorgarle una atención prioritaria incluso en los momentos iniciales de aprendizaje. Más allá de la delimitación de las necesidades *objetivas* de índole profesional o académico, en forma de vocabulario especializado, un formato de esta naturaleza a la hora de secuenciar contenidos no hace sino asignar al aprendiz un papel totalmente marginal y subordinado.

5.5. En los modelos curriculares cerrados que han hecho un uso prioritario o exclusivo de los criterios subjetivos, estos se han apoyado simplemente en el principio de autoridad. Ello conduce hacia patrones al menos tan rígidos, y además mucho menos fiables, que los obtenidos por medios objetivos. O sea, que la subjetividad que se ha venido considerando en tales planteamientos es la de los *responsables autorizados* de la docencia: profesores, gestores de centros de enseñanza y autores de materiales. Rara vez se toma en consideración el papel de los estudiantes, que son los protagonistas auténticos de todo el proceso. Ese secular olvido no puede ser prolongado ya en nuestros días, al menos en el marco de los enfoques comunicativos de enseñanza.

La defensa que hago aquí de los criterios subjetivos reivindica, ante todo, la subjetividad de los aprendices. Si importante es saber el léxico que se debe enseñar, mucho más importante es saber qué léxico quieren aprender los estudiantes. Ello pasa por la debida atención curricular al análisis de necesidades, que, como bien apunta Susana Pastor (2004: 215), es el que condiciona probablemente la pertinencia de otros posibles criterios de selección. A la incipiente consideración de las necesidades objetivas que guió el proyecto Nivel Umbral (cf. Slagter, 1979: 1) ha seguido a partir del nacimiento del enfoque por tareas una línea de trabajo basada en la negociación y la autonomía del aprendizaje, que sólo es posible a partir de la información relativa a la necesidades subjetivas. En consonancia con esto, hace algunos años diseñé un cuestionario de análisis de necesidades léxicas (Salazar, 1996 y 1999) destinado a obtener información acerca de las creencias y actitudes que sobre la adquisición del léxico manifiestan los estudiantes de segundas lenguas. El *item* situado en noveno lugar preguntaba directamente a los encuestados el tipo de vocabulario que deseaban aprender. Para ello, se enunciaban seis grandes dominios que debían ser numerados en orden de preferencia. Los reproduzco aquí sintéticamente, entre otras cosas para que pueda ser comparado con la secuenciación de contenido reproducida en 5.4:

1. *La vida cotidiana.*
2. *Ocio y tiempo libre.*
3. *Comunidad hispanohablante.*
4. *Temas de actualidad.*
5. *Intereses profesionales.*
6. *Otros temas.*

Lo primero que hay que advertir es que la ordenación de los seis dominios es meramente convencional, y no presupone de antemano ninguna secuencia didáctica concreta. Lo segundo es que el criterio utilizado no es cuantitativo (la frecuencia), sino cualitativo: el valor semántico. Tales dominios abarcan, con trazos necesariamente gruesos, todas las potencialidades comunicativas del vocabulario, y permiten una primera delimitación que habrá de irse con-

cretando paulatinamente, a lo largo del proceso negociador, hasta llegar al diseño de tareas relacionadas con ámbitos nocionales específicos. Este listado está pensado para un currículo abierto en el que la respuesta de los aprendices tenga incidencia decisiva en el desarrollo posterior de la práctica docente. Carece de sentido preguntar por los deseos del aprendiz si existe de antemano una secuenciación de contenidos presuntamente objetiva, impuesta desde instancias externas. La lógica hace pensar que los estudiantes de nivel inicial necesitarán ante todo el vocabulario del primer dominio, mientras que los estudiantes de niveles avanzados se orientarán hacia otras preferencias. La administración experimental de este cuestionario reveló que los estudiantes son capaces de responder *lógicamente* por sí solos (cf. Salazar, 1999: 390 y ss.). Pero una administración honesta del análisis de necesidades está obligada a asumir de principio que unos resultados contrarios a las expectativas del docente son igualmente válidos de cara a su plasmación en el desarrollo posterior del curso.

6. A modo de conclusión: contenidos supeditados a la tarea

En el enfoque por tareas, prototipo de una enseñanza de segundas lenguas de corte decididamente comunicativo y con vocación de fomentar la autonomía del aprendizaje, la indagación en las necesidades subjetivas del aprendiz actúa como requisito previo a la aplicación de procedimientos objetivos de selección de contenidos. Para el caso particular de los contenidos léxicos, dicha indagación es mucho más relevante, ya que, dentro de los diferentes componentes del sistema lingüístico, es probablemente el que guarda una relación más directa con el entorno y los intereses de los aprendices. Por tanto, es el que está sometido a un mayor margen de fluctuación en virtud de tales factores. A la hora de acometer una planificación didáctica dentro de este marco, los contenidos son inseparables de las tareas que entran en juego durante la actividad pedagógica. En consecuencia, la selección de unidades léxicas destinadas a formar parte del sílabus resulta pertinente solamente después de que, en el marco específico del aula, se haya decidido, por parte de los protagonistas del proceso, el ámbito nocional que vertebrará cada unidad didáctica. Es entonces, y no antes, cuando puede hacerse uso, con toda legitimidad, de los criterios objetivos que (en términos de frecuencia, uso y disponibilidad) permitan la puesta en práctica efectiva de dicha selección, destinada a especificar las unidades léxicas más acordes con la correcta realización de la tarea correspondiente.

De entre los criterios objetivos contemplados, ya he puesto de relieve a lo largo del presente trabajo que el de frecuencia se revela como el más operativo a estos efectos. Los datos obtenidos por ese medio pueden ser perfilados en las proporciones oportunas con los procedentes de estudios de uso y de disponibilidad. Asimismo, no hay por qué cerrar la puerta a la eventual intervención de las apreciaciones subjetivas del docente, que pueden contribuir a un mejor ajuste de los materiales lingüísticos a las condiciones de aprendizaje. En cualquier caso, no resulta en ningún modo procedente postular una selección de contenidos única y universalmente válida para todos los contextos de instrucción. Los recursos estandarizados, ya sean diccionarios de diversa índole, vocabularios graduados, etc., han de ser contemplados como materiales de apoyo para la enseñanza-aprendizaje, en ningún caso como componentes constitutivos del currículo. El aprovechamiento didáctico que se haga de ellos será muy variable en cada caso, dependiendo de las necesidades de los aprendices, los objetivos del curso y los resultados de la negociación en el aula.

Creo que las diferencias entre un currículo abierto y un currículo cerrado pueden ser ilustradas brevemente con un sencillo ejemplo, con el que pondré fin a estas páginas. Un currículo cerrado constata primeramente que existe en español cierto número de vocablos relativos a la jardinería que los estudiantes extranjeros están obligados a aprender si quieren alcanzar un determinado nivel de competencia. Diseña luego tareas en las que se debe practicar dicho vocabulario, con el fin de que se cumpla el objetivo de su aprendizaje. Un currículo abierto, por el contrario, comienza indagando en qué dominios de discurso están más interesados los estudiantes. Una vez que se ha comprobado que uno de esos dominios es la jardinería y se ha negociado dedicar a él una serie de tareas comunicativas, se procede a seleccionar los contenidos (léxicos, socioculturales, etc.) más rentables para alcanzar el propósito de aprendizaje acordado por el grupo. En aras de ese propósito, los procedimientos objetivos de selección léxica, tanto en términos de frecuencia como, en la medida de lo posible, de uso y de disponibilidad, están llamados a desempeñar un papel de capital importancia. Pero siempre teniendo en cuenta que el camino a seguir parte del diseño de la actuación didáctica, no de la especificación de los contenidos lingüísticos.

Referencias bibliográficas

- Aitchison, Jean (1987): *Words in the Mind. An Introduction to the Mental Lexicon*. Oxford, Blackwell.
- Alameda, José Ramón y Cuetos, Fernando (1995): *Diccionario de frecuencias de las unidades lingüísticas del castellano*. 2 vols. Oviedo, Universidad de Oviedo.
- Alinei, Mario (1974): *La struttura del lessico*. Bolonia, Il Mulino.
- Arnaud, Pierre J. L. y Savignon, Sandra J. (1997): "Rare words, complex lexical units and the advanced learner". En James Coady y Thomas Huckin (eds.), *Second Language Vocabulary Acquisition. A Rationale for Pedagogy*. Cambridge, Cambridge University Press, págs. 157-173.
- Appel, René (1996): "The lexicon in second language acquisition". En Peter Jordens y Josine Lalleman (eds.), *Investigating Second Language Acquisition*. Berlín, Mouton de Gruyter, págs. 381-403.
- Azurmendi Ayerbe, María José (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.
- Bartol Hernández, José Antonio (2002): "El léxico que enseñamos (aportaciones desde la disponibilidad léxica)". En Vicente González Martín (coord.), *Hacia la unidad en la diversidad: difusión de las lenguas europeas*. Salamanca, Universidad de Salamanca y Diputación de Salamanca, págs. 62-74.
- Bellón Cazabán, Juan Alfredo (1973): *La poesía de Luis Cernuda: estudio cuantitativo del léxico de LA REALIDAD Y EL DESEO*. Granada, Universidad de Granada.
- Benítez Pérez, Pedro (1994): "¿Qué vocabulario hay que enseñar en las clases de español como lengua extranjera?". En Lourdes Miquel y Neus Sans (coords.), *Didáctica del español como lengua extranjera 2*. Madrid, Fundación Actilibre, págs. 9-12.
- Buchanan, Milton A. (1927): *A Graded Spanish Word Book*. Toronto, Toronto University Press.
- Carcedo González, Alberto (1998): "Sobre disponibilidad léxica en estudiantes de español como lengua extranjera". *RILCE*, 14/2, págs. 204-224.
- Carcedo González, Alberto (1999): "Estudio comparativo del vocabulario español(LE) disponible de estudiantes finlandeses y el de la sintopía madrileña: propuestas didácticas". *Documentos de Español Actual*, 1, págs. 73-87.
- Chomsky, Noam (1965): *Aspectos de la teoría de la sintaxis*. Madrid, Aguilar, 1970.

- Eaton, Helen S. (1940): *An English – French – German – Spanish Word Frequency Dictionary*. Nueva York, Dover.
- García Hoz, Víctor (1946): “Evolución cuantitativa del vocabulario en escolares de 9 a 18 años”. *Revista Española de Pedagogía*, 16, págs. 403-437.
- García Hoz, Víctor (1953): *Vocabulario usual, común y fundamental*. Madrid, CSIC.
- González Martínez, Adolfo E. (2002): *La disponibilidad léxica de los alumnos preuniversitarios de la provincia de Cádiz*. Cádiz, Universidad de Cádiz.
- Gougenheim, Georges (1967): “La statistique du vocabulaire et son application dans l’enseignement des langues”. *Les langues modernes*, 61, págs. 137-144.
- Gougenheim, Georges (1971): *Dictionnaire fondamental de la langue française*. 2ª ed. París, Didier.
- Goulden, Robin, [I. S.] Paul Nation y John Read (1990): “How large can a receptive vocabulary be?”. *Applied Linguistics*, 11/4, págs. 341-363.
- Grice, H. P[aul] (1975): “Logic and conversation”. En P. Cole y J. Morgan (eds.), *Speech Acts (Syntax and Semantics)*, vol. 3). Nueva York, Academic Press, págs. 41-58.
- Imbs, Paul (1971): *Études statistiques sur le vocabulaire français. Dictionnaire des fréquences. Vocabulaire littéraire des XIXe et XXe siècles*. París y Nancy, Centre de Recherches pour un Trésor de la langue française (CNRS) y Librairie Marcel Didier.
- Inalf (Institut National de la Langue Française) (1984): *Dictionnaire des fréquences: table de répartition des homographes*. Nancy, CNRS e INALF.
- Jenkins, Joseph R. y Robert Dixon (1983): “Vocabulary Learning”. *Contemporary Educational Psychology*, 8, págs. 237-260.
- Juilland, Alphonse y Eugenio Chang-Rodríguez (1964): *Frequency Dictionary of Spanish Words*. La Haya, Mouton.
- Juilland, Alphonse, Dorothy Brodin y Catherine Davidovitch (1970): *Frequency Dictionary of French Words*. La Haya, Mouton.
- Justicia, Fernando (1985): *El vocabulario usual del niño en el ciclo inicial y el ciclo medio de la E.G.B.* Granada, Universidad de Granada.
- Justicia, Fernando (1993): “Metodología en la investigación del vocabulario infantil”. *RESLA*, 9, págs. 125-136.
- Justicia, Fernando (1995): *El desarrollo del vocabulario. Diccionario de frecuencias*. Granada, Universidad de Granada.
- Keniston, Hayward (1933): *Basic List of Spanish Words and Idioms*. Chicago, Chicago University Press.
- Keniston, Hayward (1941): *A Standard List of Spanish Words and Idioms*. Lexington, D. C. Heath.
- Lahuerta Galán, Javier y Mercè Pujol Vila (1993): “La enseñanza del léxico: una cuestión metodológica”. En Lourdes Miquel y Neus Sans (coords.), *Didáctica del español como lengua extranjera 1*. Madrid, Fundación Actilibre, págs. 117-138.
- Laufer, Batia (1997): “The lexical plight in second language reading”. En James Coady y Thomas Huckin (eds.), *Second Language Vocabulary Acquisition. A Rationale for Pedagogy*. Cambridge, Cambridge University Press, págs. 20-34.
- López Morales, Humberto (1978): “Frecuencia léxica, disponibilidad y programación curricular”. En *Aportes de la lingüística a la enseñanza del español como lengua materna*, número especial del Boletín de la Academia Puertorriqueña de la Lengua Española, 6/2, págs. 73-86.
- López Morales, Humberto (1997): *Léxico disponible de Puerto Rico*. Madrid, Arco Libros, 1997.
- López Morales, Humberto (2003): “La enseñanza-aprendizaje del vocabulario: de la teoría a la práctica”. En Elisa Viseras Soler y Antonio Castro Díaz (eds.), *Actas del VIII Simposio de Actualización Científica y Didáctica de Lengua Española y Literatura*. Málaga, Asociación Andaluza de Profesores de Español “Elio Antonio de Lebrija” y Excma. Diputación Provincial de Málaga, págs. 79-89.

- López Morales, Humberto y Francisco J. García Marcos (1995): "Disponibilidad léxica en Andalucía. Proyecto de investigación". *REALE*, 3, págs. 65-76.
- Marcos Marín, Francisco (1991): "Corpus lingüístico de referencia de la lengua española". *Boletín de la Academia Argentina de Letras*, 56, págs. 129-155.
- Márquez Villegas, Luis (Dir.) (1975): *Vocabulario del español hablado*. Madrid, SGEL.
- Mateo García, M^a Victoria (1998): *Disponibilidad léxica en el COU almeriense*. Almería, Universidad de Almería.
- McCarthy, Michael y Felicity O'Dell (1994): *English Vocabulary in Use. Upper-Intermediate & Advanced*. Cambridge, Cambridge University Press.
- Michéa, René (1953): "Mots fréquents et mots disponibles. Un aspect nouveau de la statistique du langage". *Les Langues Modernes*, 47, págs. 338-344.
- Morales, Amparo (1986): *Léxico básico del español de Puerto Rico*. San Juan, Academia Puertorriqueña de la Lengua Española.
- Nation, I. S. Paul (1983): *Learning and Teaching Vocabulary*. Wellington (Nueva Zelanda), Victoria University Press.
- Nation, [I. S.] Paul y Jonathan Newton (1997): "Teaching vocabulary". En James Coady y Thomas Huckin (eds.), *Second Language Vocabulary Acquisition. A Rationale for Pedagogy*. Cambridge, Cambridge University Press, págs. 238-254.
- Ortega Olivares, Jenaro (1985): "Algunas observaciones sobre los diccionarios de frecuencias". *Actas del 2º Congreso Nacional de Lingüística Aplicada*. Madrid, AESLA, págs. 429-439.
- Pastor Cesteros, Susana (2004): *Aprendizaje de segundas lenguas. Lingüística aplicada a la enseñanza de idiomas*. Alicante, Universidad de Alicante.
- Paribakht, T. Sima y Marjorie WESCHE (1997): "Vocabulary enhancement activities and reading for meaning in second language vocabulary acquisition". En James Coady y Thomas Huckin (eds.), *Second Language Vocabulary Acquisition. A Rationale for Pedagogy*. Cambridge, Cambridge University Press, págs. 174-200.
- Pérez Basanta, Carmen (1999): "La enseñanza del vocabulario desde una perspectiva lingüística y pedagógica". En Sagrario Salaberri Ramiro (ed.), *Lingüística aplicada a la enseñanza de lenguas extranjeras*. Almería, Universidad de Almería, págs. 262-306.
- Pérez Vidal, Carmen y Eliseo Picó Mas (1995): "Impacto de un programa de lectura extensiva en la adquisición del inglés lengua extranjera (parte II)". En J. M. Ruiz Ruiz et al. (eds.), *XI Congreso Nacional de Lingüística Aplicada*. Valladolid, Universidad de Valladolid, págs. 591-596.
- "Pragma", Equipo (1983): *Para empezar* (Libro del alumno, 2 vols.; cuaderno de ejercicios, 2 vols.; libro del profesor y material de audio). Barcelona, Edi-6.
- Quirk, Randolph (1981): "International communication and the concept of nuclear English". En Larry E. Smith (ed.), *English for Cross-Cultural Communication*. Londres, Macmillan, págs. 151-165.
- Rafel i Fontanals, Joaquim (dir.) (1996-1998): *Diccionari de Freqüències*. 3 vols. Barcelona, Institut d'Estudis Catalans.
- Richards, Jack C., John Platt y Heidi Platt (1985): *Diccionario de lingüística aplicada y enseñanza de lenguas*. Adaptación de Carmen Muñoz y Carmen Pérez. Barcelona, Ariel., 1997.
- Rivers, Wilga M. (1968): *Teaching Foreign Language Skills*. Chicago, The University of Chicago Press.
- "Rosa Sensat", Equipo (1978): *Vocabulario básico infantil*. Barcelona, Bibliograf.
- Salazar García, Ventura (1996): "Una propuesta para el análisis de necesidades léxicas en el aula de español". En Carlos Segoviano (ed.), *La enseñanza del léxico español como lengua extranjera*. Frankfurt am Main, Vervuert, págs. 13-21.
- Salazar García, Ventura (1999): "Indagaciones en la enseñanza del léxico a través del análisis de necesidades". En Fermín Sierra Martínez y Carmen Hernández González (eds.), *Las lenguas en la Europa Comunitaria III*. Amsterdam y Atlanta, Rodopi, págs. 379-394.
- Samper Padilla, José Antonio (1995): "Macrocorpus de la norma lingüística culta de las principales ciudades de España y América". *Lingüística (ALFAL)*, 7, págs. 263-293.

- Seashore, R. H. y L. D. Eckerson (1940): "The measurement of individual differences in general English vocabularies". *Journal of Educational Psychology*, 31, págs. 14-38.
- Seco, Manuel (1978): "Problemas formales de la definición lexicográfica". En *Estudios ofrecidos a Emilio Alarcos Llorach*, vol. 2. Oviedo, Universidad de Oviedo, págs. 217-239.
- Slagter, Peter Jan (1979): *Un nivel umbral*. Estrasburgo, Consejo de Europa.
- Stein, Gabriele (1978): "Nuclear English: reflections on the structure of its vocabulary". *Poetica*, 10, págs. 64-76.
- Thorndike, Edward e Irving Lorge (1944): *The Teacher's Word Book of 30000 Words*. Nueva York, Columbia University Press.
- Ullmann, Stephen (1962): *Semántica. Introducción a la ciencia del significado*. Madrid, Aguilar, 1976.
- Zimmerman, Cheryl Boyd (1997): "Historical trends in second language vocabulary instruction". En James Coady y Thomas Huckin (eds.), *Second Language Vocabulary Acquisition. A Rationale for Pedagogy*. Cambridge, Cambridge University Press, págs. 5-19.