

Antonio Manuel Ávila Muñoz /
Juan Andrés Villena Ponsoda

Variación social del léxico disponible en la ciudad de Málaga

Diccionario y análisis

Segunda edición con epílogo
de Natividad Hernández Muñoz



PETER LANG

Antonio Manuel Ávila Muñoz / Juan Andrés Villena Ponsoda

Variación social del léxico disponible en la ciudad de Málaga

En este libro se presentan los fundamentos y los resultados de la investigación sobre el vocabulario disponible de los hablantes de la ciudad de Málaga. Esta investigación reúne y hace compatibles la línea tradicional de disponibilidad léxica y la línea específica de este trabajo, que se centra en la capacidad léxica de los hablantes. La primera conduce a la elaboración del diccionario de disponibilidad de la ciudad de Málaga obtenido a partir de cuestionarios basados en centros de interés; ello permite establecer comparaciones con trabajos precedentes realizados con estudiantes preuniversitarios y nos provee de un material de gran interés para el mejor conocimiento de la lengua española hablada en Málaga. La segunda línea pretende confirmar determinadas hipótesis sobre la determinación que la posición social y otras variables de hablante ejercen sobre el vocabulario virtual de los individuos.

Los autores

Antonio Manuel Ávila Muñoz es profesor de Lingüística General en la Universidad de Málaga. Investiga en áreas relacionadas con el estudio de la variación léxica y dialectología social, esencialmente. Sus intereses de investigación también abarcan ámbitos relacionados con la lingüística cognitiva y el cálculo de la riqueza léxica a través de propuestas novedosas como la (des)centralización léxica.

Juan Andrés Villena Ponsoda es Catedrático Emérito de Lingüística General en la Universidad de Málaga. Su investigación ha supuesto importantes avances en el campo de la variación lingüística, la dialectología social y la teoría sociolingüística en general. Entre sus principales contribuciones al desarrollo de estas áreas caben mencionarse decenas de artículos científicos, así como algunos libros publicados en editoriales internacionales de primer nivel.

ISBN 978-3-631-88442-3



9 783631 884423

www.peterlang.com

Variación social del léxico disponible
en la ciudad de Málaga

Antonio Manuel Ávila Muñoz /
Juan Andrés Villena Ponsoda

Variación social del léxico disponible en la ciudad de Málaga

Diccionario y análisis

Segunda edición con epílogo
de Natividad Hernández Muñoz



PETER LANG

Información bibliográfica publicada por la Deutsche Nationalbibliothek

La Deutsche Nationalbibliothek recoge esta publicación en la Deutsche Nationalbibliografie; los datos bibliográficos detallados están disponibles en Internet en <http://dnb.d-nb.de>.

Catalogación en publicación de la Biblioteca del Congreso

Para este libro ha sido solicitado un registro en el catálogo CIP de la Biblioteca del Congreso.

ISBN 978-3-631-88442-3 (Print)
E-ISBN 978-3-631-88654-0 (E-PDF)
E-ISBN 978-3-631-89511-5 (EPUB)
DOI 10.3726/b20490

© Peter Lang GmbH
Internationaler Verlag der Wissenschaften
Berlin 2023

Todos los derechos reservados.

Peter Lang – Berlin · Bruxelles · Lausanne · New York · Oxford

Esta publicación no puede ser reproducida, ni en todo ni en parte, ni registrada en o transmitida por un sistema de recuperación de información, en ninguna forma ni por ningún medio, sea mecánico, fotoquímico, electrónico, magnético, electroóptico, por fotocopia, o cualquier otro, sin el permiso previo por escrito de la editorial.

www.peterlang.com

ÍNDICE

	Página
PRÓLOGO	9
0. INTRODUCCIÓN	15
Aproximación a los estudios estadísticos del léxico.	
Teoría y principios. Antonio M. Ávila Muñoz.	15
0.1. El índice de frecuencia	17
0.2. El índice de disponibilidad	24
0.3. Otros recuentos	29
PRIMERA PARTE. LA INVESTIGACIÓN SOBRE LA DISPONIBILIDAD LÉXICA.	35
Capítulo 1. La disponibilidad léxica. Antecedentes y fundamentos. Antonio M. Ávila Muñoz, y José María Sánchez Sáez.	37
1.1. Antecedentes.	37
1.1.1. La tradición francesa	37
1.1.2. La tradición hispana	39
1.1.3. Aplicaciones	42
1.2. Fundamentos.	45
1.2.1. Fundamentos cognitivos	45
1.2.2. Fundamentos matemáticos	50
1.2.2.1. Revisión del modelo clásico	52
1.2.2.2. Esbozo de un modelo difuso	55
1.2.2.2.1. Paso 1. Concepto de ‘prototipicidad’ y disponibilidad léxica	55
1.2.2.2.2. Paso 2. Cálculo de la capacidad léxica: Índice de Descentralización	66
1.2.2.2.3. Final. Formulación	76
Capítulo 2. Estudio del condicionamiento social del léxico dispo- nible. Objetivos, metodología y criterios de edición del diccionario. Antonio M. Ávila Muñoz y María de la Cruz Lasarte Cervantes ...	83
2.1. Antecedentes.	84
2.2. Objetivos.	85
2.3. Metodología	87

2.3.1. Estandarización de los datos	88
2.3.2. Criterios de edición generales	93
2.3.3. Criterios de edición específicos	99
Capítulo 3. Análisis estadístico comparativo y estudio cualitativo.	
Antonio M. Ávila Muñoz	113
3.1. Análisis estadístico comparativo	113
3.1.1. Promedio de palabras. Generalidades	116
3.1.2. Promedio por palabras. Estudio comparativo	118
3.1.3. Promedio de vocablos. Generalidades.	122
3.1.4. Análisis por vocablos. Estudio comparativo	124
3.1.5. Índice de cohesión	126
3.1.6. Índice de disponibilidad	129
3.1.6.1. Estudio general	129
3.1.6.2. Léxico de frecuencia y léxico disponible	131
3.2. Estudio cualitativo	141
3.2.1. Sexo	142
3.2.2. Edad	156
3.2.3. Educación	164
Capítulo 4. Análisis cuantitativo. Variación social del léxico disponible. Juan A. Villena Ponsoda y Antonio M. Ávila Muñoz	
4.1. Hipótesis	177
4.2. Variables	178
4.2.1. Variables lingüísticas	179
4.2.1.1. Variables léxicas iniciales	179
4.2.1.2. Variables léxicas corregidas	183
4.2.1.2.1. Gradación de nuclearidad	183
4.2.1.2.2. Nuclearidad y desplazamiento	187
4.2.1.3. Identificación de tecnicismos	192
4.2.2. Variables de hablante	199
4.2.2.1. Descripción general	199
4.3. Análisis	202
4.3.1. Exploración	202
4.3.1.1. Prestigio ocupacional	202
4.3.1.2. Educación	206
4.3.2. Distribución. Análisis bivalente	211
4.3.2.1. Efecto de la posición social	211
4.3.2.1.1. Educación y ocupación	211

4.3.2.1.2. Interacción de la educación y la ocupación.	214
4.3.2.1.3. Posición social y léxico especializado	217
4.3.2.1.4. Descentralización léxica e identificación de tecnicismos.	218
4.3.2.2. Efecto de la edad	219
4.3.2.2.1. Patrón curvilíneo de edad	219
4.3.2.2.2. Patrones fónicos curvilíneos	222
4.3.2.2.3. Estratificación educacional de los contrastes de edad	223
4.3.2.2.4. Efecto de la edad en el ICoDD de los hablantes universitarios.	225
4.3.2.3. Efectos de las diferencias de sexo.	228
4.3.2.3.1. Sexo y edad	228
4.3.2.3.2. Educación y sexo.	232
4.3.2.3.3. Sexo, edad y educación.	236
4.3.2.3.4. Grupo interactivo (Educaredad)	241
4.3.3. Determinación multicausal de la disponibilidad léxica. Análisis multivariante	243
4.3.3.1. Efecto combinado de la educación y la intensidad reticular. Análisis de regresión lineal	243
4.3.3.2. Modelo causal (<i>path analysis</i>).	246
4.3.3.2.1. Modelo sociológico (1)	250
4.3.3.2.2. Modelo socioléxico (2)	251
4.3.3.3. Reducción de datos. Dimensiones subyacentes	255
4.3.3.3.1. Análisis de Componentes Principales. Modelo 1	257
4.3.3.3.1.1. Dimensiones de nuclearidad y desplazamiento.	257
4.3.3.3.1.2. Dimensión de la nuclearidad.	260
4.3.3.3.1.3. Dimensión del desplazamiento	262
4.3.3.3.2. Análisis de Componentes Principales. Modelo 2	265
4.3.3.3.2.1. Dimensiones de la nuclearidad ampliada y la desviación. ACP 2	265
4.3.3.3.2.2. Continuos corregidos de nuclearidad ampliada y desviación	268
4.3.3.3.3. Control de las dimensiones subyacentes. Análisis multivariante	272
4.3.3.3.4. Conclusiones	274

4.3.3.3.4.1. Modelo metodológico	274
4.3.3.3.4.2. Modelo de la comunidad de hablantes	276
4.3.3.3.4.3. Interacción de modelos	277
4.3.3.3.4.4. Consecuencias.	279
Capítulo 5. Final. El estudio de la variación del léxico disponible. Juan A. Villena Ponsoda y Antonio M. Ávila Muñoz	281
5.1. Disponibilidad y centralidad.	281
5.2. Posición social y capacidad léxica	283
5.3. Redes sociales.	284
5.4. Edad	284
5.5. Sexo	285
5.6. Perspectivas	285
Referencias bibliográficas	287
SEGUNDA PARTE. APÉNDICES.	299
Apéndice 1: Cuestionario sociológico y reticular	301
Apéndice 2: Apéndice ortográfico	315
Apéndice 3: Test de léxico especializado	321
TERCERA PARTE. DICCIONARIO DEL LÉXICO DISPONIBLE	325
Presentación de las listas	327
Lista de palabras ordenadas por índice de disponibilidad	329
Lista de palabras ordenadas por orden alfabético	465
EPÍLOGO A CARGO DE NATIVIDAD HERNÁNDEZ MUÑOZ..	535
AGRADECIMIENTOS.....	537

Prólogo

En este libro se presentan los fundamentos y los resultados de la investigación sobre el vocabulario disponible de los hablantes de la ciudad de Málaga. El trabajo ha sido posible gracias a la financiación de la Consejería de Innovación, Ciencia y Empresa de la Junta de Andalucía a través de la primera convocatoria de Proyectos de Excelencia (2006-2010).

En gran medida, por lo tanto, el contenido de este volumen responde a la mayor parte de los objetivos del Proyecto de Investigación de Excelencia HUM-315: Condicionamiento Social del Léxico Disponible en la Ciudad de Málaga (CONSOLEX).

Por otra parte, y dado el carácter fuertemente relacionado de ambos proyectos, nuestro estudio se ha beneficiado asimismo de la ayuda de la DGICYT al Proyecto de Investigación Estudios sobre el Corpus PRESEEA de la Ciudad de Málaga (HUM2007-65602-C0703).

I

Es cierto que los estudios sobre la disponibilidad léxica cuentan con una tradición sólida y arraigada, sobre todo en el ámbito hispano. Sin embargo, en el presente trabajo se han intentado desarrollar vías poco transitadas con anterioridad.

En primer lugar, frente a lo que ha sido más común hasta ahora, hemos trabajado con una muestra representativa de una ciudad, en vez de centrarnos en el estudio del sector de los jóvenes estudiantes. Si bien hemos aprovechado los conocimientos, los métodos de trabajo de campo y las herramientas analíticas surgidas y perfeccionadas en la investigación de sectores parciales de las poblaciones, el campo de nuestro estudio puede considerarse relativamente nuevo. Hasta ahora, no se ha considerado oportuno desarrollar una investigación de este tipo, que pretende no solo construir el diccionario del vocabulario disponible de una ciudad, sino asi-

mismo buscar los factores estratificacionales, reticulares e individuales que condicionan su uso por parte de los hablantes. En este sentido, los resultados obtenidos son —como el lector podrá apreciar— de cierto interés y, sobre todo, animarán a la realización de investigaciones similares en otras comunidades¹.

En segundo lugar, este trabajo plantea una revisión de los modelos cognitivos y matemáticos sobre los que se sostienen los estudios de disponibilidad léxica. Hemos reflexionado acerca de los fundamentos y los límites de las listas de disponibilidad (qué representa la lista de disponibilidad léxica de un hablante o qué estructura reproduce) y hemos repasado las bases de las fórmulas utilizadas para la obtención de los listados de palabras. Ahora bien, dada la naturaleza de nuestro estudio, centrado en el condicionamiento social del léxico disponible, era necesario definir la disponibilidad no tanto como propiedad de los elementos léxicos, sino como variable a disposición de los hablantes individuales. En este trabajo proponemos, en consecuencia, una aproximación matemática pensada para desentrañar la estructura de acceso al vocabulario por parte de los hablantes. Hemos querido explicar el proceso objeto de estudio y, por extensión, reutilizar la información obtenida para construir nuevos aspectos cuantificables. De esta manera, se avanza una propuesta de explotación de las listas de disponibilidad que utiliza el tradicional modelo de cuantificación de vocablos para crear un índice que representa el concepto de disponibilidad individual a partir de la capacidad léxica de los hablantes, más allá del simple cómputo de elementos léxicos mencionados en cada centro de interés. Para la elaboración de este índice individual hemos aprovechado los conceptos y las herramientas conceptuales de la teoría de los conjuntos difusos, de modo que fuera posible la explotación cuantitativa en el plano individual de los datos que contienen las listas de disponibilidad.

De este modo, hemos desarrollado una investigación que reúne y hace compatibles la línea tradicional de disponibilidad léxica y la línea específica de este trabajo, que se centra en la capacidad léxica de los hablantes. La primera conduce a la elaboración del diccionario de disponibilidad de la ciudad de Málaga obtenido a partir de cuestionarios basados en centros de interés; ello permite establecer comparaciones con trabajos pre-

¹ De hecho, en el momento de la redacción de este prólogo, cumple una estancia como profesora visitante en nuestro seminario Celia María Pérez Marqués, investigadora del Centro de Lingüística Aplicada y profesora de la Universidad de Oriente (Cuba), que plantea la realización de un estudio similar al nuestro en su comunidad de origen.

cedentes realizados con estudiantes preuniversitarios y nos provee de un material de gran interés para el mejor conocimiento de la lengua española hablada en Málaga. La segunda línea pretende confirmar determinadas hipótesis sobre la determinación que la posición social y otras variables de hablante ejercen sobre el vocabulario virtual de los individuos.

La estructura del libro es la siguiente: por un lado, la introducción y las dos primeras partes, que se centran en los fundamentos, el método y los resultados del análisis cualitativo y cuantitativo de los datos obtenidos. Por otro lado, la tercera parte, donde se ofrecen las listas de disponibilidad léxica obtenidas, ordenadas, primero, según el índice de disponibilidad de cada vocablo en cada uno de los centros de interés estudiados y, después, siguiendo un criterio alfabético. Cada uno de los capítulos que componen el libro va firmado por su autor o autores. De este modo, queremos dejar constancia de la contribución de cada miembro del grupo de investigación al resultado final, sin por ello dar la impresión de una colección de capítulos sin unidad.

En la Introducción (Aproximación a los estudios estadísticos del léxico. Teoría y principios), Antonio M. Ávila efectúa un breve repaso de las principales investigaciones en el ámbito de la estadística léxica, con especial atención por aquellas que se ocupan de los recuentos de formas. La idea es situar al lector en el marco de investigación en el que surgieron este tipo de trabajos en la primera mitad del siglo pasado.

En la Parte Primera (La investigación sobre la disponibilidad léxica) se desarrolla el núcleo teórico, metodológico y analítico. En la Parte Segunda (Apéndices) se incluyen algunos documentos útiles para una mejor comprensión de los datos manejados y de las pruebas complementarias realizadas.

En el capítulo 1 (La disponibilidad léxica. Antecedentes y fundamentos), Antonio M. Ávila y José María Sánchez se centran en la historia y los fundamentos de este tipo de estudios. Primero, desarrollan un recorrido panorámico por el devenir de los estudios sobre la disponibilidad léxica, desde su génesis en el ámbito francés hasta los últimos trabajos aparecidos en el contexto hispano. Se exponen asimismo las principales aplicaciones de este tipo de investigación (Antecedentes). A continuación, se ocupan de los principios cognitivos y matemáticos de los estudios sobre el léxico disponible (Fundamentos). Llevan a cabo una revisión del modelo matemático tradicional y presentan una propuesta de cálculo de la capacidad

léxica individual a partir de una consideración de la disponibilidad léxica basada en modelos de centralización/descentralización.

En el capítulo 2 (Estudio del condicionamiento social del léxico disponible. Objetivos, metodología y criterios de edición del diccionario) Antonio M. Ávila y María de la Cruz Lasarte detallan los principios y métodos del mencionado Proyecto de Estudio del Condicionamiento Social del Léxico Disponible en la Ciudad de Málaga. Exponen asimismo los criterios de edición generales y particulares de las listas de léxico disponible que se incluyen en la Tercera Parte.

En el Capítulo 3 (Análisis estadístico comparativo y estudio cualitativo) Antonio M. Ávila ofrece la descripción estadística general de los resultados obtenidos y la comparación con los presentados en trabajos anteriores. Lleva a cabo asimismo un estudio cualitativo en el que se sirve del programa *LexiDisp* para obtener los veinte vocablos con mayor índice de disponibilidad de cada centro de interés. Con ellos desarrolla algunas comparaciones de los resultados en función del número reducido de variables sociales que permite el programa. Se accede así al conocimiento de ciertas tendencias educacionales, de edad y de sexo en el léxico mencionado por los hablantes de la muestra estudiada.

En el Capítulo 4 (Análisis cuantitativo. Variación social del léxico disponible) Juan Andrés Villena y Antonio M. Ávila analizan los patrones socioléxicos de variación a través del estudio bivalente y multivalente de los índices de capacidad léxica individual. La información social, personal y reticular obtenida de los informantes mediante cuestionario (Apéndice 1) ha permitido la construcción de modelos de regresión, multicausales y de componentes principales que permiten afirmar que la variación del léxico virtual de los hablantes estudiados tiene una estructura que está determinada por los efectos directos y, sobre todo, indirectos de las variables de posición social, la intensidad de los vínculos reticulares que unen a los hablantes con sus pares más cercanos y el grado de exposición a la información especializada. Para desarrollar este estudio se han construido tres medidas que constituyen las variables dependientes del estudio: además del número de palabras aportadas por cada informante en sus listas (IAL), el Índice de Descentralización Léxica (IDD) o capacidad léxica individual y la Identificación de Tecnicismos o grado de conocimiento de léxico especializado (*Pident*) a partir de una prueba complementaria a la de disponibilidad (Apéndice 3).

Una de las conclusiones más interesantes de este estudio sobre el léxico disponible en Málaga es positiva y seguramente estimulante para el

lector. Los modelos multivariantes construidos sobre la idea de un condicionamiento directo del estatus social sobre la capacidad léxica, según los cuales el nivel educacional obtenido por el hablante en el período de instrucción formal determinaría esencialmente la riqueza de su vocabulario individual, resultan menos satisfactorios que aquellos que tienen en cuenta el incremento de capacidad léxica que depende de factores distintos, como pueden ser la actitud personal ante el conocimiento o la exposición a los medios de comunicación. El grado de reconocimiento de léxico especializado se revela así como una variable de intervención que multiplica la capacidad predictora de la educación y de otras variables de posición social y destruye la imagen estereotipada de dos códigos compartimentados que separan o segregan a la comunidad.

II

La realización de este estudio no habría sido posible sin la ayuda desinteresada, paciente y animosa de nuestros setenta y dos informantes, que representan las diferentes clases sociales, generaciones y sexos de la comunidad investigada. Ellos nos suministraron cuantos datos personales, reticulares, de costumbres y actitudes les solicitamos. Se sometieron a las pruebas de disponibilidad y nos ayudaron con los tests de reconocimiento del léxico especializado. Sin ellos esta investigación no habría podido llegar a buen puerto. A ellos dedicamos este libro.

El trabajo de campo, la codificación y la informatización de los datos la realizaron, con diversos grados de participación, Antonio M. Ávila, que se encargó de la tarea de coordinación en estas etapas iniciales de la investigación, María de la Cruz Lasarte Cervantes, Marina Ariza Oriol y Pedro Villena Oliver. Su esfuerzo continuado en el tiempo y su tenacidad en los momentos difíciles han sido un ejemplo y un acicate para nosotros.

José María Sánchez Sáez ha trabajado en la revisión de los modelos matemáticos sobre los que se sostiene el trabajo clásico del léxico disponible. Acostrumbrado a colaborar con los lingüistas, ha tenido la paciencia necesaria para hacernos accesibles algunos conceptos y herramientas que, procedentes del campo de la teoría de los conjuntos difusos, están llamados a constituir el fundamento de los modelos de la capacidad léxica individual. Estamos convencidos de que esta aportación es solo el principio.

Agradecemos, por último, a Félix Requena Santos, sociólogo especializado en el análisis de redes sociales, sus consejos y su impagable magisterio. En particular, dedicó parte de su tiempo libre durante algún que

otro fin de semana a revisar y discutir con nosotros los modelos de análisis multicausal que rematan la parte de análisis cuantitativo de este estudio.

Juan Andrés Villena Ponsoda
Antonio M. Ávila Muñoz

Grupo de Investigación Vernáculo Urbano Malagueño (VUM)
Universidad de Málaga

Introducción

Aproximación a los estudios estadísticos del léxico. Teoría y principios

Antonio M. Ávila Muñoz

Los modelos lingüísticos relacionados con procedimientos estadísticos se consolidan, en general, en la década de los setenta, inspirados en los principios sociolingüísticos basados en la probabilidad de William Labov (1969). Al comienzo, estos esquemas teóricos recibieron un rechazo contundente por parte de ciertos sectores de la comunidad de estudiosos de la lengua, al considerar que los análisis estadísticos y probabilísticos eran inaplicables en sus formulaciones más corrientes y usuales a la estructura algebraica concebida por los lingüistas, desde Chomsky (1957), como un entramado subyacente al sistema. Por este motivo, algunos investigadores mediaron en la polémica sugiriendo que un estudio estadístico de las lenguas aceptable desde una perspectiva lingüística consistiría en la construcción de extensiones probabilísticas de los modelos lingüísticos vigentes¹. Es de esta manera como hoy día se entiende y se acepta la existencia de una materia que combine ambas disciplinas, la lingüística y la estadística².

En la actualidad podemos hablar de una auténtica lingüística estadística aplicada al estudio del léxico debido a la intervención generalizada de los procedimientos estadísticos en todos los ámbitos lingüísticos. Hasta hace unos años, esta intervención era tan tenue que, a pesar de trabajos variados de lexicoestadística, como los orientados a calcular la frecuencia, la riqueza o la disponibilidad léxica, o los preocupados por determinar las leyes de la estructura léxica, la extensión de los vocabularios y otras investigaciones similares basadas en la estadística³, no existía una justificación

¹ Cedergren y Sankoff (1974); Sankoff (1978); Rousseau y Sankoff (1989).

² *Vid.* Charniak (1993); Bod (1995), entre otros.

³ Müller (1968).

de peso suficiente para considerar una disciplina constituida, como podemos proclamar hoy. Las causas de este retraso son diversas pero, quizás, una de las razones que pueden ayudarnos a entenderlo sean, de nuevo, las limitaciones que la propia gramática generativa imponía a la disciplina. Valgan, en este sentido, las siguientes palabras del considerado máximo representante de esta corriente lingüística: «A pesar de su innegable interés e importancia, los estudios semánticos y estadísticos sobre el lenguaje parece que no atañen directamente al problema de determinar y de caracterizar el conjunto de oraciones gramaticales⁴».

La situación hoy día es bien distinta. En el ámbito de la estadística léxica, en particular, la organización de los datos y el recuento de formas adquieren una singular importancia. En función de los objetivos específicos se determina el tipo de recuento, las unidades consideradas y la forma final que tendrán los listados. El manejo de grandes bases de datos léxicos ha sido, sin lugar a dudas, esencial para el desarrollo de la disciplina. En este sentido, la irrupción de las nuevas tecnologías en el ámbito que nos ocupa supuso una considerable ayuda que facilitó el tratamiento informático de enormes archivos léxicos. El proceso seguido, aunque muy intenso, se ha llevado a cabo en poco tiempo.

De hecho, al lector del siglo XXI podrían parecerle anacrónicas las palabras que hace apenas veinticinco años aparecían en una de las obras de referencia de la lexicografía contemporánea:

Es de suponer que, en el futuro, todas las posibilidades técnicas se irán perfeccionando y serán cada vez más accesibles no sólo a entidades que disponen de recursos económicos importantes. Mientras tanto, muchos lexicógrafos tendrán que trabajar con los métodos tradicionales, es decir con ficheros⁵.

Si bien es cierto que, solo un año después, Sáez-Godoy (1983) editaba una bibliografía que recogía las principales publicaciones de estudios de lengua y literatura españolas realizados mediante procedimientos electrónicos o mecánicos, también es verdad que a principios de los ochenta el uso de la informática, al menos en nuestro país, era poco menos que un privilegio. El uso de la electrónica en la elaboración de diccionarios es común en la actualidad; sin embargo, lo hoy

⁴ Chomsky (1957: 17).

⁵ Haensch *et alii* (1982: 433). La cita se refiere, naturalmente, a ficheros compuestos por fichas de papel.

tenido por evidente hace muy poco tiempo entraba de lleno en el ámbito de la ficción científica. Quince años después de las palabras de Haensch y sus colaboradores, la revista *Cuadernos Cervantes* dedicaba íntegramente un número especial a la evolución experimentada en la tarea de elaboración de diccionarios electrónicos⁶. En la actualidad, en muchos de los departamentos de nuestras universidades especializados en lingüística computacional se realizan trabajos sobre diccionarios electrónicos del español y sobre su aplicación, por ejemplo, al análisis del léxico y la traducción automática.

Como vemos, los avances de las nuevas tecnologías en el desarrollo del estudio del léxico son muy importantes. La utilización práctica inmediata más fácil y frecuente la realizan expertos en lexicografía para determinar los artículos que debe contener un repertorio léxico, sus acepciones más importantes, las combinaciones semánticas y sintácticas que no pueden faltar, etcétera⁷. El desarrollo tecnológico ha facilitado el avance de la lexicografía computacional, gracias a que se han relacionado grandes bases de datos integradas por unidades léxicas con la reflexión teórica. A partir de este proceso se han potenciado algunas teorías hasta ahora difícilmente demostrables⁸.

En los próximos apartados exponemos algunos de los proyectos de investigación más fructíferos en el ámbito lexicográfico, con especial atención por los recuentos léxicos, como el que ordena las unidades en función del índice de frecuencia y el índice de disponibilidad.

0.1. El índice de frecuencia

Un índice de frecuencias del léxico de una lengua está formado por un listado de vocablos al que se adjunta una determinada frecuencia de

⁶ *Cuadernos Cervantes*, 11: noviembre-diciembre de 1996. Unos años antes se publicaron los trabajos de Vidal Beneyto (1991) o de Marcos Marín (1994).

⁷ *Vid.* Alvar Ezquerro *et alii* (1994: 10-11). En realidad, los análisis estadísticos basados en los sistemas informáticos aplicados al ámbito lexicográfico ofrecen enormes posibilidades en el cálculo de las frecuencias relativas de las unidades léxicas, con el objetivo de considerar o no su inclusión en diccionarios o manuales de español para extranjeros, lo que representa solo una de las posibles aplicaciones que se nos ofrecen.

⁸ Por ejemplo, sabemos que, por lo general, el aumento de palabras en un corpus no implica necesariamente un aumento del número de unidades léxicas diferentes. En cualquier corpus de textos, más o menos la mitad de las palabras solo aparece una vez, mientras que la mayor parte de las restantes aparece menos de diez veces.

aparición en el corpus considerado⁹. La procedencia de esos elementos léxicos y el tratamiento estadístico aplicado variarán en función de los intereses planteados por el autor o autores del trabajo. Los estudiosos tendrán en cuenta el destino que quieran darle a la lista de frecuencias, aunque normalmente se obtendrá por medio de la aplicación de fórmulas de frecuencia, dispersión y uso, referencias estadísticas que facilitan la elaboración de los denominados vocabularios básicos, es decir, los aproximadamente cinco mil vocablos más usados en la comunidad estudiada.

Sabemos que, a pesar de la riqueza del lenguaje, tendemos a utilizar un reducido número de unidades tanto léxicas como estructurales que son, evidentemente, las más frecuentes¹⁰. Los léxicos de frecuencia más completos, además, suelen estudiar la frecuencia del léxico común en función de ciertas variables sociolingüísticas y situacionales. Se suele indagar el uso que los hablantes seleccionados hacen del sistema léxico de su lengua, en su modalidad escrita o hablada, determinando, si es posible, ciertos subsistemas léxicos que puedan ser considerados como variedades sociolingüísticas¹¹.

⁹ Martínez de Sousa (1995: 149) lo define como «diccionario que registra el índice de frecuencia con que una palabra es utilizada en una lengua».

¹⁰ Es sabido desde hace ya más de cincuenta años que el hombre culto y considerado intelectual maneja entre cuatro mil y cinco mil vocablos, en tanto que esta cifra disminuye hasta dos mil en el menos favorecido culturalmente, que no quiere decir analfabeto sino, como lo califica López Morales (1986: 59), «el hombre común».

¹¹ En nuestro país, Guillén Sutil (1987) fue la primera investigadora en plantearse unos objetivos como los descritos. No obstante, en aquella ocasión el estudio del léxico se limitaba a un estrato sociocultural denominado culto, agrupándose los sujetos informantes en función de solo dos variables sociales: generación y sexo. Además, en dicho estudio se partía de un corpus total de tan solo 55.753 palabras obtenidas de veinticuatro encuestas. A pesar de ello, este estudio puede considerarse pionero en el ámbito de la socioestadística léxica en España, porque pretendía establecer diferencias de uso léxico en el español hablado en Andalucía a partir de unas variables sociológicas determinadas. Así mismo, Raúl Ávila (1986, 1988), lleva a cabo un estudio lexicoestadístico con atención a ciertas variables sociológicas en la ciudad de México. Aunque este autor estaba interesado fundamentalmente en establecer determinados índices de riqueza léxica, en especial la densidad léxica, la metodología empleada en la recolección del material procedente de grabaciones realizadas en México con informantes de ambos sexos, distinta edad y condición social, así como procedentes de distintas localidades, puede considerarse pionera de este tipo de investigación en el ámbito hispano.

A pesar de lo anterior, aún hoy, no existe demasiada bibliografía en torno a la variación léxica¹². Desde siempre se han realizado distintas clasificaciones del lenguaje que permiten asomarnos a su estructura sistemática de manera coherente y ordenada. Sin embargo, disponemos de pocos trabajos que clasifiquen las unidades lingüísticas por su frecuencia de uso¹³, y esto a pesar de que esta variable ha demostrado ser una de las de mayor peso en el procesamiento del lenguaje, sea desde una perspectiva de producción o de descodificación¹⁴. No obstante, existen diccionarios de frecuencias léxicas de las principales lenguas occidentales y las nuevas tecnologías están demandando nuevos trabajos en este campo para ser aplicados de manera muy diversa.

Los diccionarios de frecuencias más desarrollados, es decir, aquellos en los que se emplean fórmulas estadísticas ponderadoras como la desviación típica, no se limitan a presentar simples colecciones de palabras sino a mostrar una selección de las que a través de los métodos estadísticos empleados resultan ser las más estables y básicas de la muestra en cuestión.

Los primeros diccionarios encargados de medir la frecuencia de aparición de los distintos elementos léxicos en los textos manejaban el concepto de ‘frecuencia’ como base para realizar sus cálculos¹⁵. Sin embargo, la mera frecuencia, sea esta considerada de modo absoluto o relativo¹⁶, no resulta una

¹² López Morales (1989: 105), ya lo advirtió hace años. De entre los pocos trabajos específicos, Vincent (1982; 1983); Horvath (1985); Lemieux *et alii* (1986). En nuestro país, los estudios son aún más escasos: Guillén Sutil (1987); Benítez (1992); García Marcos (1994); Mateo García (1996), Escoriza Morera (1999, 2002, 2004a, 2004b, 2006) o Ávila Muñoz (1998).

¹³ Y así lo denuncian Alameda y Cuetos (1995) en la introducción de su obra.

¹⁴ Ellis (1985) puso de manifiesto la dificultad, e incluso imposibilidad, de producción lingüística cuando el emisor debe componer mensajes con unidades que le son desconocidas. De Vega *et alii* (1990) demostraron también que existe una proporción inversa en el sentido de que cuanto menor es la frecuencia de utilización de las palabras de un determinado texto, mayor es la dificultad de comprensión por parte del receptor.

¹⁵ En el caso de los diccionarios de frecuencia del español, Rodríguez Bou (1952) utiliza la frecuencia ponderada, en tanto que Keniston (1941) primero y García Hoz (1953) después emplean el rango como factor de cálculo básico. Juilland y Chang Rodríguez (1964) incorporan con modificaciones el rango o dispersión simple para alejar el riesgo de realizar un cálculo en el que no se tenga en cuenta la distribución de las frecuencias en el texto.

¹⁶ Está claro que para el análisis solo pueden emplearse las relaciones de valor, y no los valores absolutos, lo que supone que dispongamos de auténticos puntos de comparación calculados según criterios idénticos. Solo podremos interpretar unos datos estadísticos si pueden ser fraccionados y las partes resultantes pueden reagruparse a su vez de otro modo.

medida especialmente significativa, ya que podemos encontrar dos vocablos con la misma frecuencia, pero con distribución irregular entre los géneros establecidos, lo que significa que el elemento está sujeto a ciertas variables tipológicas, mientras que una distribución regular entre los géneros denota que la palabra es independiente de esas variables, y por tanto, su peso y utilidad dentro del idioma debe ser mayor¹⁷. Los listados obtenidos a partir del simple cómputo de frecuencia de aparición no pueden controlar las irregularidades surgidas de modo circunstancial en la elección de muestras escogidas al azar. Evidentemente, podría ocurrir que, de modo casual, se seleccionen varios textos monográficos sobre un mismo tema, por lo que el vocabulario correspondiente daría la sensación de ser muy frecuente¹⁸, por lo que se ponderó rápidamente este factor. El índice de dispersión compleja mide la distribución de la frecuencia de un modo muy sencillo: se parte de la distribución del universo léxico considerado en tipos o «mundos» delimitados por el contenido de los textos o por las condiciones formales de su recolección. La función de la dispersión es determinar la estabilidad de la frecuencia en las categorías establecidas, lo que favorece que la tipología considerada se convierta en representativa y básica en la ponderación¹⁹. La fórmula definitiva propuesta por Juilland y Traversa en 1973 tras diversas correcciones, se basa en el promedio de los desvíos de la frecuencia de cada vocablo en cada subfrecuencia respecto a su frecuencia teórica, con lo que se obtiene el coeficiente de variación de cada palabra en la muestra escogida, con independencia de la frecuencia:

Lo importante, en todo caso, es que se garantice la posibilidad de regresar a la consulta de los datos iniciales en cualquier momento con el fin de volver a emplearlos de nuevo con otros objetivos o, simplemente, controlarlos. Cf. de Kock (1983: 14-15).

¹⁷ Lara y Chande (1974: 35). Por esta razón, se buscaron cálculos cuyos índices estabilizaran la frecuencia en el texto considerado, llegándose a los índices de dispersión compleja y uso ya utilizados en la serie *The Romance Language and their Structures* inaugurada por A. Juilland y sus colaboradores.

¹⁸ Más acuciante es el problema al seleccionar intencionadamente un determinado tipo de texto con un campo léxico muy particular; por ejemplo, si tomásemos como base del recuento el mundo de la economía, elementos léxicos como *divisa* o *bursátil* alcanzarían una frecuencia mayor que el índice real de uso de estos elementos en la comunidad general.

¹⁹ Partimos del supuesto de que dichas categorías son intrínsecamente homogéneas y lo suficientemente heterogéneas entre sí para ser representativas de la tipología a la que se refieren.

$$D = 1 - \frac{\sqrt{nX^2 - T^2}}{2T},$$

donde $T = \sum xi$.

Posteriormente, otro cálculo introduce el factor del número de categorías consideradas, lo que permite obtener un índice que oscila entre 0 —dispersión pésima— y 1 —dispersión óptima—.

Diferentes variantes de esta fórmula han venido aplicándose a los distintos recuentos léxicos a partir de ese momento, a pesar de los varios errores en la presentación del tratamiento estadístico que han advertido algunos autores²⁰.

Una vez conocido si el lexema se reparte o no uniformemente, es decir, si los hablantes, (independientemente del registro utilizado, de la variable sociolingüística que lo identifica o del tema del discurso lingüístico) hacen un uso uniforme o no de aquel, solo resta obtener el listado de vocablos que se ha de incluir en el diccionario básico de la lengua. Para ello se recurre a un nuevo índice, el factor de uso léxico, cuya fórmula pondera al alza los repartos equitativos de frecuencias, en detrimento de aquellas palabras que tienen mayor frecuencia pero una distribución desigual, de tal manera que el índice de uso se acercará más a la frecuencia a medida que se regularice la repartición de las unidades en las categorías escogidas²¹. La fórmula de uso léxico es el resultado de multiplicar la frecuencia por la dispersión:

$$U = F \times D.$$

Normalmente, como hemos señalado, los léxicos básicos recogen los aproximadamente cinco mil vocablos más usados por la comunidad estudiada²². A través de los procedimientos metodológicos anteriores se obtiene un léxico que posee una estabilidad estadística determinada, es decir, el léxico al que los hablantes recurren más a menudo para construir sus mensajes, independientemente del tema del discurso. El esquema del

²⁰ Müller (1965: 35) ya da cuenta del problema.

²¹ Müller (1977: 68-76).

²² Haensch *et alii* (1982: 159).

listado básico de una lengua carece de microestructura, o sea, de definiciones, ejemplos, acepciones, etcétera. Su aplicación suele ser multifuncional, aunque de modo general, se aprovechan intensamente en las planificaciones de la enseñanza del léxico de lenguas maternas²³ y de segundas lenguas²⁴. En este sentido, hay que tener presente que con los materiales lingüísticos obtenidos, tanto si se aplican en la enseñanza de lenguas maternas o de lenguas extranjeras, no estamos solo facilitando la enseñanza o el aprendizaje de una asignatura concreta, es decir, una materia con una frontera de acción claramente delimitada y propia. La formación lingüística parece interesar a una concepción global de la educación.

A continuación (Tabla 1) mostramos un fragmento del listado de frecuencia elaborado por Amparo Morales (1986: 201). Realmente, el modelo estructural que se observa es el mismo que siguen casi todos los léxicos básicos y de frecuencia que se han elaborado hasta el momento²⁵: resaltado en negrita el lema correspondiente; a continuación, su categoría gramatical; columnados, también en negrita, tres índices correspondientes a: uso, frecuencia y dispersión respectivamente. En la fila siguiente, en cursiva, las apariciones totales del lema. En las filas restantes se presentan las frecuencias particulares de cada forma flexiva.

²³ Sensat (1978: 6), Lorenzo Delgado (1980: 9-26) o López Morales (1986: 57). Aplicaciones prácticas en el ámbito del español encontramos en García Hoz (1953 y 1976), Sensat (1978), Lorenzo Delgado (1980), Neggers (1981), Jiménez Jiménez (1986), Mesanza López (1987), Sarramona (1988), Ministerio de Educación y Ciencia (1989), Barbadillo (1991) o Guerrero Ruiz y López Valero (1993). Para una exposición bibliográfica de diccionarios básicos aplicados fundamentalmente a la enseñanza de lenguas maternas, *vid.* Sarramona (1988: 32-41).

²⁴ En este aspecto de enseñanza de segundas lenguas, las listas de palabras básicas realizadas sobre textos orales son imprescindibles. Si lo que deseamos es enseñar a hablar, parece lógico buscar la objetivación del léxico en datos procedentes de documentos hablados. En español tenemos una importante aplicación en este sentido realizada por Rojo Sastre (1980), quien elaboró un Español Fundamental para la enseñanza de nuestra lengua como LE, cuyos materiales fueron recogidos de encuestas orales y tests de disponibilidad léxica. Con estos datos, y bajo la supervisión de una comisión de «jueces» tan prestigiosos como Alvar, Lapesa, Criado de Val, Green, Lorenzo Criado, Merimée, Pottier, Zamora Vicente o Pérez Villanueva, se procedió a la creación de un método de español para extranjeros (*Vida y Diálogo de España*) cuya importancia es todavía hoy incuestionada.

²⁵ Formato que, por otro lado, es idéntico al que propusieron Juilland y Chang Rodríguez (1964).

Tabla 1. Fragmento de listado de frecuencia léxica. Fuente: A. Morales (1986)

jamás av.	38.25	57	0.67	judicial aj.	4.70	9	0.52
jamás	25	11	4	2	10	1	0
jardín n.	22.09	44	0.50	judicial	1	2	4
jardín	25	11	4	2	2	1	1
jardines	24	5	4	2	judío n.	4.84	13
jaula n.	7.13	9	0.79	judío	9	2	0
jaula	3	2	1	2	judíos	3	1
jaulas	3	2	1	2	1	1	1
jefe n.	55.89	68	0.82	juego n.	63.77	107	0.60
jefe	7	10	16	14	29	14	12
jefes	5	9	15	7	1	10	1
jíbaro n.	27.82	47	0.59	juegos	28	11	10
jíbaro	9	17	19	0	1	3	2
jíbaros	7	10	18	2	jueves n.	4.92	24
jíbaro aj.	11.51	19	0.61	jueves	1	1	2
jíbaro	4	4	9	0	juez n.	37.70	61
jíbara	2	1	4	1	9	4	6
jíbaras	1	1	2	1	5	3	5
jíbaros	1	1	2	1	4	1	1
jornada n.	6.08	9	0.68	jugador n.	5.97	17	0.35
jornada	3	1	3	0	1	3	1
jornal n.	3.64	8	0.45	jugador	1	1	1
jornal	0	1	1	5	jugadores	2	8
jornales	1	1	5	1	jugar v.	81.38	130
joven aj.	101.0	116	0.87	58	35	9	4
joven	31	23	20	14	24	1	1
jóvenes	28	18	9	2	7	1	7
joven n.	64.54	43	101	4	0.64	11	3
joven	9	32	14	2	5	2	5
jóvenes	5	11	4	2	21	31	0.63
joya n.	5.64	13	0.43	jugan	1	31	0.63
joya	8	0	3	2	0		
joyas	1	2	2	1			

0.2. El índice de disponibilidad

Existen otros tipos de análisis léxicos que buscan inventariar el vocabulario más usual en determinados contextos comunicativos. Tal es el caso de los índices de disponibilidad que, en cierta manera, vienen a suplir las deficiencias que se observan en las listas de vocabulario básico²⁶. Al examinar estos últimos listados, podemos observar que ciertas palabras muy comunes y conocidas pueden faltar o alcanzar unos índices de frecuencia muy bajos, de modo que no se considerarían integrantes del repertorio básico. Esta conclusión estadística puede, en determinadas ocasiones, conducirnos a error²⁷.

Se observó en seguida que el criterio lexicoestadístico de selección basado en la frecuencia era insuficiente, pues las listas de frecuencia léxica contenían casi exclusivamente vocablos «atemáticos» y faltaban otros de contenido léxico, vocablos «temáticos» muy comunes y usuales para cualquier hablante, siempre que se den las condiciones comunicativas necesarias para su aparición²⁸. En realidad, las palabras «temáticas» están relacionadas con un tema o asunto determinado, designan seres y objetos y son, en su mayor parte, palabras que aparecen casi siempre en agrupaciones, de modo que el empleo de alguna favorece el de otras y casi siempre se repiten cuando se abordan determinados temas.

El empleo de las palabras «atemáticas», por otro lado, no está relacionado con una decisión consciente del hablante, sino que estas se actualizan en el discurso de manera más o menos automática, pues su función es polivalente²⁹. La múltiple función que pueden llegar a alcanzar estos elementos léxicos hizo que Rojo Sastre los llamase «palabras ganzúa»,

²⁶ En lo que sigue presentamos los principios elementales de los listados de disponibilidad, pues dedicaremos los capítulos siguientes a explicar con profusión sus fundamentos epistemológicos.

²⁷ Gougenheim (1967: 137-144) llegó a la conclusión de que el hecho de que no aparezcan en el *Français Fondamental* (Gougenheim *et alii*, 1964) palabras de sobra usadas en ambientes francófonos, como *métro*, *lettre* o *timbre*, se debía fundamentalmente a cuestiones ajenas al conocimiento de estos vocablos por parte de los hablantes. Michéa (1953: 338-344) inauguró la tradición de la elaboración de los léxicos disponibles puesto que, hasta su incursión en el mundo estadístico, la frecuencia era el único factor manejado para establecer el orden de los vocablos en las listas (*vid.* capítulo 1).

²⁸ La diferencia «temático»-«atemático» está muy relacionada con la discusión de origen aristotélico entre palabras llenas (con significado) y palabras vacías (sin significado).

²⁹ De Kock (1983: 59-60).

pues, tal y como él demostró, se utilizan en la expresión lingüística para dar entrada a diversas combinaciones enunciativas³⁰. En definitiva, estas palabras «atemáticas», muy usadas y de aparición común en los listados de frecuencia léxica, dependen directamente más de los hábitos y de la competencia lingüística de los usuarios que del tema que se esté tratando; la gramática, a fin de cuentas, se muestra ajena, entre otros, a los cambios de tema.

Para compensar estas irregularidades, los léxicos disponibles pueden resultar el complemento adecuado por ser el reflejo del caudal léxico usado en una situación comunicativa determinada. Este tipo de listado explota el concepto de ‘situación frecuente’, que contrasta con el manejado por los listados de frecuencia léxica, centrados en el objetivo del ‘léxico frecuente’. En realidad, el léxico de disponibilidad encuentra su sentido en la máxima de que ciertas palabras muy usadas en una lengua concreta están estrechamente relacionadas con la aparición o no de determinados temas³¹. Para la recolección del vocabulario disponible, aunque no necesariamente frecuente, se parte de la elaboración de unas pruebas asociativas que giran en torno a unos estímulos o centros de interés³². Alrededor de estos núcleos temáticos surge determinado vocabulario relacionado que, se supone, es el léxico potencial que pertenece al vocabulario activo de los sujetos y del que estos hacen uso si una conversación en un momento dado discurre

³⁰ Rojo Sastre (1980: 202).

³¹ De hecho, hay una baja probabilidad, a menos que se hable sobre los medios de transporte, de que se mencionen palabras como *metro* o *autobús* en situaciones comunicativas corrientes. Del mismo modo, como indica López Morales (1983: 213), es «poco probable [...] que salgan en nuestra expresión palabras como *carta* y *sello* si no nos referimos específicamente a asuntos del correo».

³² En los distintos estudios sobre disponibilidad léxica, los centros de interés son tan variados que pueden abarcar temas tan dispares como el cuerpo humano, el correo, la prensa, la agricultura, los sindicatos, las finanzas, la meteorología, el vestuario, los transportes, el circo, las profesiones, la ganadería, etcétera. En realidad, la metodología de recogida de materiales aprovecha una tradición bien consolidada en el ámbito de la psicología experimental. De hecho, el asociacionismo de palabras está ligado en sus orígenes a las teorías de Freud. Aunque la psicología posterior modificó los planteamientos de base, el método ha seguido empleándose hasta nuestros días para la investigación en la construcción de redes léxicas (Aitchison, 1987; Nelson, Bennet y Leibert, 1997; Nelson, McEvoy y Denis, 2000). Para un análisis crítico de la prueba usada en la producción del léxico disponible, sus mecanismos y su evolución en la psicología experimental, así como sus repercusiones en la lingüística, véase Hernández Muñoz (2006: 85-125).

por los cauces necesarios³³. El número total de vocablos que componen estos listados de léxico disponible está normalmente limitado, bien por el tiempo de reacción de los hablantes, bien por el cierre de las listas tras una cantidad determinada de palabras. Se supone que son más disponibles las palabras que acuden antes a la memoria, es decir, aquellas que aparecen en los primeros lugares de las listas³⁴. El léxico disponible forma parte, por tanto, del lexicón mental de los hablantes, pero no suele actualizarse en los intercambios lingüísticos cotidianos, a menos que exista una especialización temática. La Tabla 2 es una muestra de la lista de disponibilidad léxica elaborada por Mateo García (1996: 151) en el estudio del léxico de la alimentación y las áreas temáticas próximas en la localidad de Motril. Como se aprecia, el rango establece una gradación de los vocablos, ordenados según su frecuencia de aparición, en tanto que el índice de disponibilidad proviene de una fórmula de uso generalizado, en virtud de la cual la disponibilidad de un vocablo se calcula sumando las ocasiones en que aparece referido en primer lugar, en segundo, etcétera, compensando la frecuencia por un coeficiente, inferior a la unidad, que va restando valor a la frecuencia a medida que la palabra se aleja del primer lugar de la lista (*vid. infra*, § 1.2.2.).

³³ Para una crítica del concepto de ‘léxico disponible’, véase Hernández Muñoz (2006: 29-36). Allí se pone en duda que el léxico disponible sea realmente el caudal de léxico utilizable en una situación comunicativa dada, pues, tal y como señala la autora, no podemos olvidar que la tarea de producir el léxico disponible constituye un método artificial de activación y enumeración de palabras. Siguiendo el esquema teórico de Marconi (2000), Hernández Muñoz acepta que los hablantes manejamos dos tipos de destrezas diferentes a la hora de producir léxico: las inferenciales y las referenciales. Para la producción del léxico disponible se aplican las destrezas inferenciales basadas en el conocimiento de la red de conexiones establecidas entre una palabra y el resto de los elementos del mismo sistema lingüístico. Ello provoca que no sepamos con seguridad si los hablantes son capaces de utilizar con eficiencia esas palabras en el discurso cotidiano, ya que para incluirlas en las listas solo requieren el conocimiento básico de que pertenecen a cierta clase de elementos, pero no la información necesaria para su aplicación referencial, necesaria para actualizar las palabras en situaciones comunicativas reales. Cf. Hernández Muñoz (2006: 46).

³⁴ Hacemos una revisión crítica de este planteamiento en § 1.2.2.2., Paso 2.

Tabla 2. Fragmento de listado de disponibilidad léxica.
Fuente: García Mateo (1996)

RANGO	VOCABLO	ÍNDICE de disponibilidad
87	ensaladilla rusa	.466
88	congelar	.465
88	guinda	.465
88	pestiño	.465
88	zumo	.465
89	breve	.464
89	despensa	.464
89	espinacas	.464
89	jurel	.464
89	torta “al-hajú”	.464
90	café	.463
90	caqui	.463
90	espagueti	.463
90	guisado	.463
90	limpiar	.463
90	ron	.463
91	acelgas	.462
91	asado	.462
91	fregar	.462
91	hígado	.462
92	café negro	.461
92	carne de vacuno	.461
92	“li-chis”	.461

El grado de disponibilidad resulta un elemento corrector eficaz de la frecuencia léxica, y la consideración de uno y otro índice puede ofrecernos una imagen bastante cercana de la realidad léxica estudiada. Esta queda especificada en el léxico fundamental, definido como el resultante de la combinación del léxico básico y el léxico disponible de una comunidad³⁵. En la Tabla 3 se comparan los datos producidos por la frecuencia (tanto relativa como absoluta) con otra lista donde se relativiza esa frecuencia a partir del grado de disponibilidad³⁶.

³⁵ Evidentemente, este concepto de ‘léxico fundamental’ no se entiende ya hoy en el sentido que estableció García Hoz (1953), para quien este léxico estaría constituido por aquellos vocablos con una frecuencia superior al 100 por 100.000 de aparición. Como ya hemos señalado (*supra*, nota 24), en nuestro país ya se han realizado y aplicado a la enseñanza de segundas lenguas léxicos fundamentales (Rojo Sastre, 1980).

³⁶ Los datos proceden de una cala realizada por López Morales (1983: 214) entre alumnos de primer grado de escuelas públicas de la zona metropolitana de San Juan de Puerto Rico.

Tabla 3. Comparación de los datos de frecuencia y disponibilidad léxica.
Fuente: López Morales (1983)

rango		f. absoluta	f. relativa	rango	índice de disponibilidad	
1	mano	6	100	1	mano	.572
1	pie	6	100	2	cabeza	.562
2	cabeza	5	83.3	3	pie	.541
3	ojos	4	66.6	4	pierna	.339
3	pierna	4	66.6	5	ojos	.305
4	cuello	3	50	6	cuello	.212
4	nariz	3	50	7	nariz	.191
5	dedo	2	33.3	8	oreja	.146
5	hueso	2	33.3	9	hueso	.140
5	oreja	2	33.3	10	dedo	.138
5	pelo	2	33.3	11	pelo	.118
5	brazo	2	33.3	12	brazo	.100
6	carne	1	16.6	13 ³⁷	carne	.090
6	rodilla	1	16.6	13	rodilla	.090
6	pecho	1	16.6	13	piel	.090
6	piel	1	16.6	14	hombro	.081
6	hombro	1	16.6	15	ombbligo	.072
6	ombbligo	1	16.6	15	rabo	.072
6	barriga	1	16.6	16	pecho	.065
6	boca	1	16.6	16	barriga	.065
6	corazón	1	16.6	16	boca	.065
6	rabo	1	16.6	16	corazón	.065

A simple vista observamos cómo el poder discriminatorio del índice ponderado es mayor que el de las demás medidas. Como se constata, se producen hasta 16 rangos allí donde el índice de la frecuencia sin ponderar arroja solo 6. Ello se debe a que el índice de disponibilidad valora al alza aquellos vocablos aparecidos en primer lugar frente a los que aparecen en otras posiciones. Es de suponer, por tanto, que ante vocablos con idéntica frecuencia como *mano* y *pie* (frecuencia absoluta 6 y frecuencia relativa

³⁷ Debido a lo limitado de la muestra empleada, los elementos que presentan igual frecuencia y grado no pudieron ser discriminados por el índice de disponibilidad, aunque con muestras lo suficientemente amplias, la compensación de la frecuencia con el grado hace que los vocablos suban o bajen de rango.

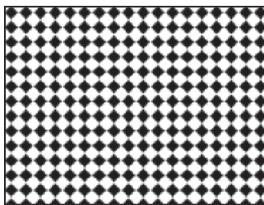
100), el índice de disponibilidad valora más *mano* por haber sido citado alguna vez en primera posición frente a *pie*, que según este criterio no ha aparecido en dicha posición o lo ha hecho en menos ocasiones³⁸. Normalmente, la compensación de la frecuencia con el grado de disponibilidad modifica el rango de los vocablos.

0.3. Otros recuentos

También resultan significativos los léxicos de la norma culta: colección de índices de frecuencia con especificaciones de competencia léxica, del vocabulario conocido y utilizado en el habla culta media (habitual) de cada ciudad investigada³⁹.

En realidad, el léxico culto se sitúa en una tercera etapa del idioma, cuyas dos fases anteriores corresponden al léxico básico y al léxico disponible⁴⁰. El léxico básico se suele situar en la primera fase por ser el más necesario, el que utilizamos corrientemente. En alguna ocasión (Ávila Muñoz, 1998; 2006) hemos representado de modo muy simple el léxico básico de un individuo tal y como se muestra en la Figura 1.

Figura 1. Representación del léxico básico individual



Se trata de un núcleo siempre activo que está integrado, principalmente, por palabras gramaticales, algunos adjetivos y verbos comunes, y muy pocos sustantivos, estos siempre de carácter muy general con re-

³⁸ De hecho, López Morales asegura que en la muestra empleada *mano* aparece cinco veces en primer lugar, es decir, fue la palabra más disponible para cinco sujetos, mientras que *pie* apareció solo una vez en esa posición inicial (López Morales, 1983: 215).

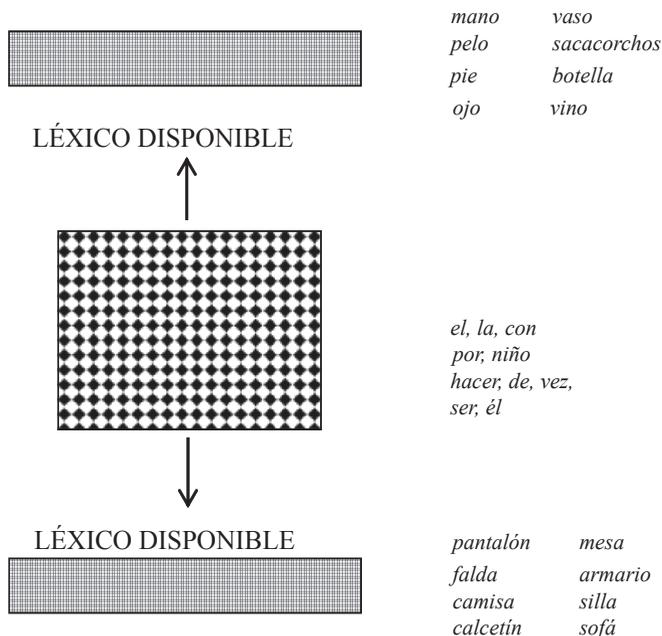
³⁹ Lope Blanch (1986: 22).

⁴⁰ Sensat (1978: 5-6).

ferencia a múltiples realidades⁴¹: *año, cosa, casa, día, niño, vez, gente y tiempo*⁴².

En un segundo estadio (Figura 2) se sitúan las palabras también conocidas por todos los hablantes pero solo utilizadas en ocasiones muy determinadas (léxico disponible). La combinación de estos dos primeros niveles produce el léxico fundamental:

Figura 2. Representación del léxico fundamental individual (frecuencia y disponibilidad)



⁴¹ Ávila Muñoz (2006: 27).

⁴² *Niño* es un ser humano en edad infantil, pero también puede referirse a cualquier adulto:

—*Señora, esto le va a costar 100 euros*

—*¡Ojú niño!, eso es muy caro.*

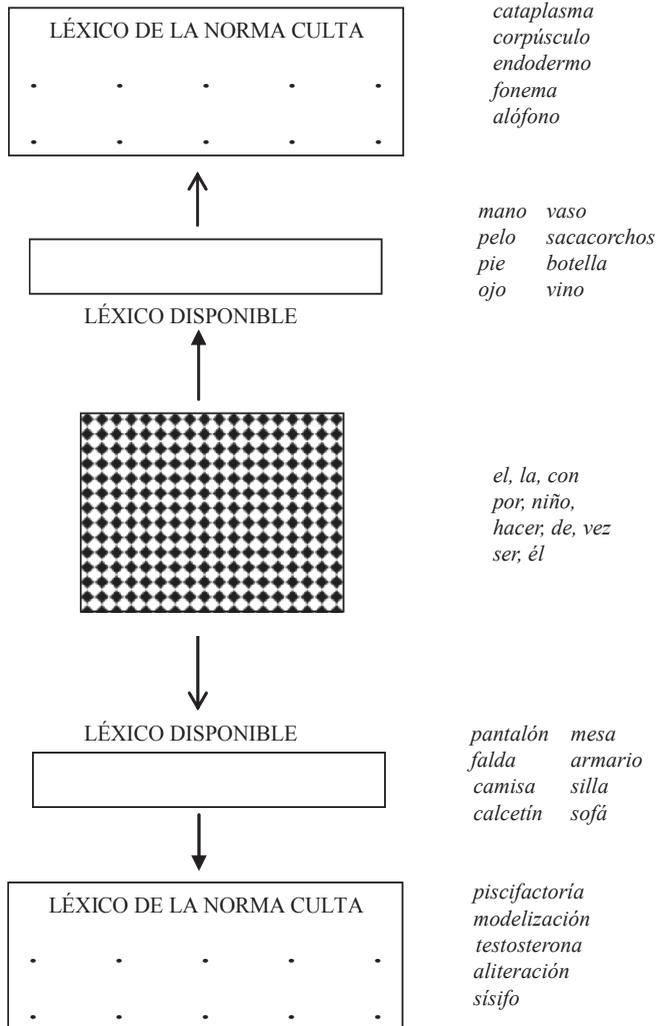
Gente hace referencia a un colectivo humano, pero en ocasiones se usa como sustitutivo de expresiones más precisas:

En la sala de espera de la consulta de un médico; conversación telefónica:

—*No voy a llegar a tiempo. Voy a tardar porque esto está lleno de gente* (en vez de *pacientes, enfermos, usuarios*, etcétera).

En la tercera y última etapa, en definitiva, está el léxico de la norma culta, cuyo dominio depende de la proyección social y de la base cultural de cada persona; se llega así a la representación del léxico completo (Figura 3).

Figura 3. Representación del léxico completo individual (frecuencia, disponibilidad y norma culta)



Normalmente, los listados de norma culta hacen referencia al habla habitual, con referencias al estilo formal (habla esmerada) e informal (habla familiar) y su objeto básico es obtener un corpus representativo y fidedigno del vocabulario activo considerado de disponibilidad espontánea e inmediata de los hablantes estudiados⁴³. Para ello se estratifica la comunidad de estudio atendiendo a niveles generacionales, genéricos y educacionales principalmente y, una vez seleccionados los informantes, se les pasa un cuestionario lingüístico dividido en determinados campos léxicos. El resultado es una lista en la que aparecen las entradas y las respuestas registradas, normalmente en orden decreciente de frecuencia.

En la Tabla 4 se muestra un ejemplo presentado por López Morales sobre San Juan de Puerto Rico⁴⁴. Se manejó una muestra de 48 sujetos divididos en cuatro submuestras paralelas, pre-estratificada atendiendo a los niveles señalados en el párrafo anterior. El cuestionario léxico que se empleó estaba integrado por 4.452 puntos repartidos en 21 campos léxicos. La muestra que ofrecemos en la Tabla 4 corresponde al campo temático del cuerpo humano.

⁴³ Uno de los trabajos más ambiciosos en este sentido es el *Proyecto de Estudio del Habla Culta de las principales ciudades de lengua española*, nacido en 1964 durante la celebración del *II Simposio del Programa Interamericano de Lingüística y Enseñanza de Idiomas (PILEI)*. Vid. Lope Blanch (1967: 255-264). Este Proyecto pretende llevar a cabo un estudio de la fonología, morfosintaxis y el léxico de la norma culta del español hablado en todas las ciudades que se unieron al macroproyecto. Para realizar el estudio de los dos primeros niveles de lengua se utilizó material grabado en cinta magnetofónica. Para el estudio del léxico se obtuvieron los datos con la ayuda de un cuestionario (Alvar (coord.), 1972). Los primeros resultados del estudio sobre el léxico pueden verse desde los trabajos pioneros de Lope Blanch (1977). Además, el lector interesado puede actualizar los datos con los informes trianuales que la *Asociación de Lingüística y Filología de la América Latina (ALFAL)* publica en los boletines informativos de esta asociación, que lleva las riendas del Proyecto. En cada una de las ciudades donde se han realizado las encuestas también se han publicado los resultados (Bogotá, Buenos Aires, Caracas, Madrid, Sevilla, Granada, Las Palmas, entre otras). Por último, en <http://www.mundoalfal.org/Proyecto%20norma%20culta.htm> el usuario accede a información sobre la historia de la investigación y noticias actualizadas sobre su desarrollo [última consulta: 6/3/2010].

⁴⁴ López Morales (coord., 1986: 16). La fuente original de los datos es el estudio llevado a cabo en San Juan de Puerto Rico por Huyke Freiría, Pérez del Valle, Rodríguez Cabrera y Cardona Santana como parte de sus respectivas tesis de maestría dirigidas por Amparo Morales, María Vaquero y López Morales.

Tabla 4. Fragmento de listado del léxico de la norma culta.

Fuente: López Morales (coord. 1986)

LAS COSTILLAS	el muslo (6)
las costillas (12)	el hueso del muslo (1)
EL TÓRAX	LA TIBIA
el pecho (9)	la espinilla (7)
el tórax (4)	la tibia (6)
EL COXIS	EL PERONÉ
el coxis (8)	el peroné (8)
el hueso del gusto (3)	el hueso de la pierna (1)
el huesito del gusto (3)	la pierna (1)
el hueso de la alegría (1)	LA RÓTULA
el huesito de la alegría (1)	la rodilla (8)
EL PUBIS	la rótula (2)
el pubis (4)	la patela (1)
el hueso púbico (3)	LA CABEZA
el área púbica (3)	la cabeza (12)
la pelvis (2)	el coco (9)
la ingle (1)	el casco (7)
EL FÉMUR	la chola (7)
el fémur (8)	el cráneo (4)

En la ciudad de Granada (Tabla 5) el estudio estuvo coordinado por F. Salvador, y también dio resultados muy interesantes, sobre todo si lo comparamos con los ofrecidos en las ciudades hispanoamericanas⁴⁵.

Tabla 5. Fragmento de listado del léxico de la norma culta en la ciudad de Granada. Fuente: F. Salvador (1991)

LAS COSTILLAS	EL FÉMUR
costillas (25)	fémur (25)
EL TÓRAX	LA TIBIA
tórax (4)	tibia (24)
cavidad torácica (5)	NS / NC (1)
caja torácica (3)	espinilla (1)
pecho (1)	EL PERONÉ
EL COXIS	peroné (24)
coxis (21)	NS / NC (1)
NS / NC (3)	LA RÓTULA
sacro (1)	rótula (24)
EL PUBIS	rodilla (1)
pubis (12)	LA CABEZA
NS / NC (5)	cabeza (24)
pelvis (5)	testa (1)
cavidad (1)	...

⁴⁵ En Granada se aplicó el cuestionario léxico sobre una muestra de 25 sujetos cultos (estudios universitarios), nacidos en la ciudad, de padres granadinos y solo con breves estancias fuera de la ciudad, en algunos casos. *Vid.* F. Salvador (1991).

Es preciso citar otro proyecto de investigación que se desarrolla de manera oficial desde 1993, fecha de su presentación en la ciudad de Veracruz (México). En el *X Congreso Internacional de la Asociación de Lingüística y Filología de América Latina (ALFAL)* se hizo público el proyecto *Español del Mundo*, coordinado por Hiroto Ueda y Toshihiro Takagaki. Este ambicioso proyecto tiene como objetivo general conocer la situación actual del léxico español de todo el mundo a través de la elaboración de una red internacional de investigación. Los objetivos específicos son⁴⁶:

- Formar una red internacional de investigación sobre el léxico español moderno-urbano. Con ello se pretende una forma de cooperación internacional para la investigación lingüística.
- Reunir los datos del español moderno (con especial atención a la variante urbana) para proceder después a su análisis con el método multivariado y cartografiarlos en múltiples mapas de grandes ciudades hispanohablantes.
- Distribuir los datos obtenidos a todos los participantes en el proyecto para su utilización y explotación.
- Establecer un centro de recogida de datos en la Universidad de Tokio. El total de los datos es procesado desde esta Universidad y desde allí se envían a todos los miembros.

Por lo tanto, se trata de formar un archivo del léxico distintivo, con objeto de obtener informaciones útiles para la investigación. Los estudiosos interesados podrán, a partir de aquí, realizar sus trabajos dialectológicos, lexicográficos, etcétera.

En estrecha relación con este proyecto se encuentra el de *Difusión Internacional del Español por Radio, Televisión y Prensa (DIES-RTP)*, coordinado desde El Colegio de México por Raúl Ávila. Ambos proyectos han establecido un convenio de colaboración y cooperación⁴⁷.

Lo más interesante de este proyecto es que se está elaborando un archivo libre, abierto, relacional y flexible en continua renovación que, según afirman sus compiladores, será capaz de observar el ritmo frenético de los cambios lingüísticos tanto diacrónicos como diatópicos del Mundo Hispano.

⁴⁶ Para una detallada exposición, Ueda y Takagaki (1993). Además, puede consultarse la página oficial del Proyecto en la siguiente dirección electrónica: <http://gamp.c.u-tokyo.ac.jp/~ueda/varilex/index.php>. En ella puede encontrarse información general sobre el proyecto, ciudades investigadas, índice de conceptos, cuestionarios usados, etcétera. [última consulta, 6/3/2010].

⁴⁷ <http://www.colmex.mx/academicos/cell/ravila/index-archivos/page0003.htm> [última consulta: 2/9/2009].

PRIMERA PARTE

LA INVESTIGACIÓN SOBRE LA DISPONIBILIDAD LÉXICA

Capítulo 1

La disponibilidad léxica. Antecedentes y fundamentos

Antonio M. Ávila Muñoz
José María Sánchez Sáez

1.1. Antecedentes

Los estudios sobre disponibilidad léxica surgieron hace más de cincuenta años para paliar los errores de método observados en el diseño de un ambicioso programa lingüístico auspiciado por la UNESCO para facilitar el aprendizaje de francés a los no nativos. Se pretendía dotar tanto a los habitantes de las colonias como a los inmigrantes en territorio francés de una «lengua base» simplificada desde el punto de vista léxico y gramatical con el objeto de posibilitar su integración social¹. Este programa educativo se centraba, esencialmente, en el vocabulario, por lo que se insistió pronto en la importancia de seleccionar los vocablos que debían integrar el soporte de enseñanza.

1.1.1. *La tradición francesa*

Los trabajos previos a la edición en 1954 de *Le Français Élémentaire* de Gougenheim y sus colaboradores pueden considerarse el punto de partida de la investigación en disponibilidad léxica². Téngase en cuenta que en aquella época el factor de selección léxica más conocido y generaliza-

¹ No está de más recordar que tras este objetivo, aparentemente benéfico, se escondía una absoluta desigualdad en el sistema educativo francés desde los orígenes republicanos. *Vid.* Balibar y Laporte (1974).

² La obra se reeditó con algunas revisiones en 1964 con el título de *Français Fondamental (1er degré). Étude sur l'établissement d'un vocabulaire et d'une grammaire de base*, por la que citan la mayor parte de los trabajos posteriores.

do era la frecuencia, por lo que en estos trabajos iniciales se empleó este índice por considerarlo el más apropiado a los objetivos: las palabras más frecuentes debían ser las más útiles y usadas³.

Sin embargo, como hemos visto en el capítulo anterior, había que cambiar de estrategia, pues las listas de frecuencia no eran suficientes para obtener la «lengua base» que pretendían los lexicógrafos franceses de mediados del siglo XX. En consecuencia, la tradición lexicostatística francesa fue poco a poco perfeccionando tanto las fórmulas estadísticas de base como otros muchos asuntos relacionados con los planteamientos metodológicos que se aplicaban en esos estudios pioneros⁴. Se comienza a entender ya, en estos años, que algunas palabras comunes y usuales no eran frecuentes sencillamente por el hecho de que hay determinados elementos lingüísticos (v. gr. los sustantivos) que solo se actualizan si se abordan determinados temas.

Michéa (1953) fue uno de los primeros en rechazar la metodología basada en el cálculo de frecuencia, impropia para la obtención de los fines perseguidos. De hecho, con él surgió la idea de oponer los conceptos de ‘palabra frecuente’ y ‘palabra disponible’. A partir de este momento, la disponibilidad léxica comenzó a considerarse como el caudal léxico potencial que se actualiza cuando es necesario comunicar una información específica. Gracias a estos planteamientos se supuso que determinados estímulos verbales (*centre d'intérêt*) facilitaban la actualización de este tipo de palabras disponibles⁵.

El trabajo que sirvió de base, modelo y punto de partida para marcar la pauta de los estudios posteriores sobre el léxico disponible fue el realizado por Gougenheim, Michéa, Rivenc y Sauvageot en 1964: *L'élaboration du Français Fondamental*. Para empezar, este trabajo sirvió para que en 1969, Dimitrijêvic diseñara una investigación sobre el léxico disponible del inglés hablado en Escocia; poco después, en 1971, Mackey y sus colaboradores compararon los resultados del trabajo de Gougenheim y su equipo con los que ellos obtuvieron al analizar el léxico disponible del francés hablado en Canadá. Como es lógico, para efectuar las oportunas comparaciones, Mackey y sus ayudantes hicieron suyas las pautas meto-

³ Michéa (1949).

⁴ Fue así como, por primera vez, los recuentos emplearon estrategias que no solo tenían en cuenta el número de veces en el que aparece una palabra, sino asimismo el orden de aparición (vid. § 1.2.2).

⁵ Vid. nota 32 de la introducción.

dológicas de los pioneros europeos. El resultado fue la publicación de dos volúmenes en los que se establece una comparación intralingüística que abarca ámbitos de intraculturalidad⁶.

1.1.2. *La tradición hispana*

La influencia del trabajo de Gougenheim y sus colaboradores fue, como vemos, muy importante en la bibliografía posterior. Pronto se observó que los estudios sobre disponibilidad léxica tenían aplicaciones pedagógicas que fueron aprovechadas por algunos estudios pioneros en el ámbito hispano. En nuestro contexto, esta línea de investigación comenzó a tomar entidad aplicada al estudio del bilingüismo. De hecho, en 1971 Bailey Victory estudió la disponibilidad léxica de hablantes bilingües de español e inglés en Tejas y María José Azurmendi (1983) se interesó por la población escolar bilingüe de San Sebastián.

Posteriormente, Max S. Echeverría y sus colaboradores (1987) trabajaron con diez centros de interés en su investigación en Chile. Existen, asimismo, varias memorias de licenciatura elaboradas en diversos países del continente americano que tienen como objeto el estudio de la disponibilidad léxica de determinados sectores poblacionales: B. Román (1985); H. Justo Hernández (1986); M. Mena Osorio (1986); A. V. Cañizal Arévalo (1987); A. Ruiz Basto (1987); R. M. Mesa Canales (1989). Por otro lado, la importancia de estos trabajos queda manifiesta en el hecho de que, por ejemplo, debemos a las investigaciones de López Chávez y Strassburguer Frías en México (1987) la presentación y desarrollo de la fórmula matemática para el cálculo de la disponibilidad que posteriormente ha sido aceptada en los diversos trabajos (*vid. infra*, § 1.2.2).

Sin embargo, los trabajos que más han ayudado al asentamiento definitivo de la tradición de la disponibilidad léxica en el contexto hispano han sido los elaborados y amparados por Humberto López Morales⁷. A ellos debemos la puesta en marcha del *Proyecto Panhispánico de estudio sobre la Disponibilidad Léxica (PPHDL)*, cuyo fin último es obtener un diccionario del léxico disponible del español a partir de la elaboración de listas

⁶ La investigación se llevó a cabo desde el *Centre International de Recherches sur le Bilinguisme de L'Université Laval* de Quebec. Esta misma institución auspició, en 1979, el estudio sobre el francés y el basaa de niños africanos del Camerún (Njock, 1979).

⁷ López Morales (1973, 1978, 1979 y, sobre todo, 1999).

de disponibilidad léxica en diferentes lugares de habla hispana repartidos por todo el mundo⁸.

La importancia de este macroproyecto coordinado reside, por encima de cualquier otra consideración, en la propuesta de homogeneización y unificación de los criterios de diseño metodológico comunes que permiten el intercambio de datos y el desarrollo de estudios comparados entre los resultados de los diferentes proyectos locales.

Los distintos grupos adscritos al macroproyecto asumen unas directrices metodológicas que pueden sintetizarse en los siguientes puntos:

⁸ En la actualidad, esta tarea se encuentra muy avanzada. Desde los primeros trabajos de López Morales (1979) y Orlando Alba (1995) han surgido grupos de trabajo que, poco a poco, se han ido uniendo a dicho proyecto. Muchos de ellos ya han concluido su tarea, sobre todo los que se han desarrollado en España: Pedro Benítez trabajó la disponibilidad léxica en el área metropolitana de Madrid (1992). José Antonio Samper y Clara Eugenia Hernández (1997) coordinan la tarea en Gran Canaria; fruto de su empeño son los diversos trabajos que nos llegan desde allí; uno de los más representativos fue la Tesis Doctoral de Marta Samper Hernández (2003). En Andalucía, López Morales y García Marcos (1995) presentaron las bases metodológicas para el estudio del léxico disponible en esta región; estos presupuestos coinciden con los generales del Proyecto Panhispánico, y se manifestaron en las investigaciones de M^a Victoria Mateo García en Almería (1998), Adolfo González Martínez en Cádiz (2002), Juan José Bellón en Córdoba (2003), Ávila Muñoz en Málaga (2006), Ayora Esteban en Ceuta (2006), Trigo Ibáñez en Sevilla (2007), Pastor Milán y Sánchez García en Granada (2008), Ignacio Ahumada en Jaén (2006), Prado Aragonés y Galloso Camacho en Huelva (2008) y Fernández Smith en Melilla (2008). Maitena Etxebarria Arostegui (1996 y 1999) centró su interés en correlacionar el léxico disponible y el modelo de enseñanza cursado en el País Vasco. M^a Victoria Galloso Camacho (2003) ha trabajado en el distrito universitario de Salamanca (Ávila, Salamanca y Zamora). En realidad, en Castilla y León existe uno de los grupos más activos en este campo de estudio, en el que se integran J. A. Bartol, J. Borrego Nieto, N. Hernández Muñoz, C. Fernández Juncal, y F. J. de Santiago Guervós. En Asturias ha estudiado la disponibilidad léxica Alberto Carcedo González (2001), quien ha centrado sus esfuerzos en observar la importancia de este campo en la enseñanza de español como LE. Gran parte de los resultados de la investigación que se viene desarrollando en la región valenciana han sido publicados por Gómez Molina y Gómez Devis (2004). María Luisa Arnal (2004) coordinó la tarea en Aragón. Asimismo, Hernández Muñoz nos ofreció en 2004 los listados referidos a la provincia de Cuenca. Entre los últimos trabajos publicados se encuentra el de Fernández Juncal en Burgos (2008), de Santiago Guervós en Segovia (2008) o López Meirama en Galicia (2008). Maribel Serrano (2003) prepara sus materiales sobre la zona de Lleida. Otros trabajos se encuentran en fase de elaboración más o menos avanzada como los que se desarrollan en Barcelona, Panamá, Cuba, Chicago, Extremadura, Alicante, Castellón. En cualquier caso, como se observa, los trabajos colectivos más adelantados se vienen desarrollando en España, donde próximamente se publicará el Diccionario de léxico disponible de España con datos de más de cincuenta provincias.

1. Se trabaja con estudiantes preuniversitarios. De esta manera, se evita la contaminación «técnica» o específica de un área profesional determinada, a la vez que el estudio se centra en sujetos que se suponen suficientemente «maduros» desde el punto de vista léxico⁹.
2. Los materiales se obtienen mediante pruebas asociativas a partir de 16 centros de interés o núcleos temáticos sobre los que el informante debe aportar los elementos léxicos que considere relacionados¹⁰. Los centros de interés propuestos son los mismos que se utilizaron en los trabajos pioneros de Gougenheim: 01. El cuerpo humano; 02. La ropa; 03. Partes de la casa (sin los muebles); 04. Los muebles de la casa; 05. Alimentos; 06. Objetos colocados en la mesa para la comida; 07. La cocina y sus utensilios; 08. La escuela: muebles y materiales; 09. Calefacción, iluminación y medios de airear un recinto; 10. La ciudad; 11. El campo; 12. Medios de transporte; 13. Trabajos del campo y del jardín; 14. Los animales; 15. Juegos y distracciones; 16. Profesiones y oficios.
3. Las listas son abiertas en el sentido de que no se limita el número de palabras aportadas, sino que el informante debe escribir todas las posibles relacionadas con el asunto propuesto.
4. La única limitación en este sentido se refiere al tiempo de que disponen los individuos para confeccionar sus listas, pues la condición es que no sobrepase los dos minutos por centro de interés.

⁹ Benítez (1992: 73), López Morales (1999: 28).

¹⁰ No obstante, como hemos señalado, siempre han existido discrepancias sobre el número y la clase de los centros de interés apropiados para la tarea (Butrón, 1987). Quizás el planteamiento más crítico y realista pueda leerse en Hernández Muñoz (2006: 127-188) quien fundamenta su exposición en: 1) no todos los centros de interés responden a la concepción tradicional de las categorías naturales organizadas en torno a un prototipo (organización interna de los centros de interés); 2) los centros de interés pueden ser de carácter inclusivo (los vocablos son elementos del nivel de base o subordinado) o de carácter relacional: los vocablos no establecen una relación de pertenencia directa (niveles de inclusividad). Como hay ocasiones en las que esta identificación se dificulta, los sujetos pueden desarrollar estrategias de respuesta que manipulen los resultados. Todo ello lleva a Hernández Muñoz (2006: 185-186) a afirmar que ni cualitativa ni cuantitativamente existe unidad entre los centros de interés seleccionados en los estudios de léxico disponible, y, lo más importante, hacen que la disponibilidad de los vocablos pueda deberse a diferentes factores. A partir de aquí, es fácil deducir que la disponibilidad no es una propiedad fija de las palabras «(probablemente ni siquiera en su naturaleza), sino que depende, en gran medida, del marco categorial». *Vid.* Hernández Muñoz (2006: 186). Para una explicación de la verdadera naturaleza del índice de disponibilidad, véase *infra*, § 1.2.2.2.1, Paso 1.

5. Una vez obtenidas las listas léxicas se editan los materiales mediante unos criterios consensuados de lematización¹¹ y las bases de datos se someten a un proceso de análisis estadístico que nos proporciona, en última instancia, las listas de disponibilidad léxica. Para ello, los investigadores cuentan con un software de distribución gratuita desarrollado por la Universidad de Alcalá y aprobado por el Instituto Cervantes: *LexiDisp* es una aplicación para Windows© que puede calcular la disponibilidad léxica a partir de la fórmula matemática desarrollada por López Chávez y Strassburger en 1987¹².
6. Al disponer de información referida a las características sociales y personales de los sujetos se pueden establecer correlaciones sociolingüísticas particulares que, posteriormente, son comparadas con las obtenidas en diferentes sintopías, con lo que los estudios contrastivos enriquecen considerablemente las posibilidades de partida.

1.1.3. Aplicaciones

Son incuestionables las múltiples posibilidades de aplicación práctica que nos ofrecen los estudios de disponibilidad léxica. Como hemos mencionado, uno de los ámbitos en los que se han desarrollado más extensamente estas posibilidades ha sido el de la enseñanza de la lengua, sea materna o extranjera, pues no en vano los estudios de disponibilidad nacieron para dar respuesta a preocupaciones surgidas en este ámbito de estudio.

¹¹ *Vid.* § 2.3.1.

¹² *Vid.* § 1.2.2. Tanto el programa como las instrucciones de manejo pueden obtenerse en el siguiente enlace: <<http://www.linguas.net/Proyectos/LexiDisp/tabid/73/Default.aspx>> [última consulta: 6 de marzo de 2010]. En § 2.3.1 presentamos los requisitos básicos de los archivos que maneja el programa. Además, el PPHDL cuenta con un lugar de encuentro electrónico desde el cual se informa de las últimas novedades a todas las personas que investigan la disponibilidad léxica o se interesan por el estado de la cuestión: <http://www.dispolex.com/> [última consulta: 6 de marzo de 2010]. En realidad, Dispollex es algo más que un simple medio de contacto, pues gracias a su interfaz, el usuario, debidamente registrado, puede incorporar sus materiales a un banco de datos general. Según sus creadores (Bartol, Hernández Muñoz, Universidad de Salamanca) con ello se contribuye a configurar un almacén panhispánico de amplia capacidad, configurado de modo que se adecue a las características de las diferentes investigaciones particulares. Por otra parte, el investigador registrado puede acceder de manera gratuita a las herramientas necesarias para realizar los cálculos más habituales de disponibilidad léxica: índice de disponibilidad, frecuencia, porcentaje de aparición, número de palabras, número de vocablos, promedios por informantes, índice de cohesión y comparaciones entre proyectos.

Sin embargo, la explotación de los materiales de léxico disponible se ha realizado desde disciplinas adyacentes, que han encontrado en sus listas una fuente fiable y relativamente fácil de estudiar a través de la cual llevar a cabo diferentes experimentos aplicados.

No es nuestro objetivo el examen minucioso de la explotación de las listas de léxico disponible en cada una de esas disciplinas¹³. Nos limitamos, en lo que sigue, a citar algunas posibilidades de interés.

En primer lugar, la sociolingüística encontró desde el principio en el campo de la disponibilidad léxica unas líneas de análisis que salvan fácilmente los inconvenientes teóricos que dificultan el estudio cuantitativo del léxico¹⁴. De hecho, existen estudios de amplia difusión e impacto en los que se demuestra que, por encima de cualquier otro condicionante socio-cultural, la edad y el tipo de educación formal de los padres son los factores que determinan las diferencias encontradas en las listas de disponibilidad léxica que generan los distintos grupos de hablantes, tanto desde el punto de vista cuantitativo como cualitativo¹⁵.

En segundo lugar, no cabe duda de que cuando un hablante produce una lista de léxico disponible lleva a cabo una tarea cognitiva compleja en la que intervienen determinados procesos psicológicos que, en gran medida, condicionan la mencionada lista¹⁶. Por tanto, no es de extrañar que la psicolingüística haya aprovechado los resultados de estas pruebas que resultan, incluso, más naturales que las pruebas psicolingüísticas tradicionales¹⁷.

En tercer lugar, las diferencias culturales que se observan entre las comunidades de habla estudiadas quedan recogidas en las listas que se obtienen en cada sintopía. Por consiguiente, la etnolingüística ha encontrado siempre una rica fuente de datos sobre determinados aspectos culturales de los grupos a través de estos listados de vocablos disponibles¹⁸. Ya en los

¹³ Para ello remitimos al lector a los diversos trabajos que abordan en profundidad este asunto como, por ejemplo, Samper *et alii* (2003: 105 y ss.).

¹⁴ Morales (1999: 25) ; Escoriza Morera (1999, 2002, 2003, 2004).

¹⁵ Por señalar alguno de los más conocidos, Samper Padilla (2006), Samper Padilla y Hernández Cabrera (1997), Gómez Molina (2006), Gómez Molina y Gómez Devís (2004), Mateo García (1996, 1998) o Samper Hernández (2003).

¹⁶ *Vid.* § 1.2.1.

¹⁷ Aitchison (1987), López Morales (1995: 252-253), Hernández Muñoz (2006).

¹⁸ *Vid.* Paredes García (2005).

primeros trabajos, Mackey (1971) hizo una comparación entre las bebidas que aparecían en las listas de los lexicógrafos franceses y las que aparecían en las obtenidas con hablantes franceses de la *Acadie* canadiense. Al estudiar ambas listas advirtió, por ejemplo, que la disponibilidad de un vocablo como *vin* ‘vino’ es prácticamente nula en Canadá (índice de disponibilidad = 1.4), situación esta que contrasta con la obtenida en Francia para el mismo vocablo (índice de disponibilidad = 29.9)¹⁹.

En cuarto lugar, se ha desarrollado una cierta tradición consistente en llevar a cabo comparaciones diatópicas de los materiales proporcionados por estos estudios. La dialectología comparada, por tanto, es otra de las disciplinas que aprovechan los estudios sobre disponibilidad como fuente de datos cuantitativamente numerosos y cualitativamente diversos²⁰. Desde que López Chávez (1992, 1993) desarrollase índices matemáticos que permiten la comparación de las listas de disponibilidad obtenidas en diferentes sintopías, todos los estudios que las han empleado nos muestran que junto a los vocablos comunes existen otros exclusivos, lo que nos permite obtener conclusiones sobre el grado de aproximación o compatibilidad entre los diferentes dialectos²¹.

En quinto lugar, por último, los materiales surgidos a partir de la realización de los trabajos de disponibilidad léxica nos permiten llevar a cabo otros análisis sobre el vocabulario. Así, por ejemplo, se han desarrollado estudios sobre la penetración y el alcance de los extranjerismos o la presencia de faltas de ortografía en estas listas. Uno de los trabajos de mayor difusión e impacto en este último sentido lo realizó Galloso Camacho en las provincias de Ávila, Salamanca y Zamora²². En él se advertía del «panorama desolador» en materia ortográfica y se hacía una llamada de atención a todos los «organismos competentes» para empezar a tomar medidas sin demora²³.

¹⁹ Como era de esperar, el mismo vocablo *vin* obtiene índices diferentes en función de si la región francesa en la que se pasaron los cuestionarios era vitivinícola o no lo era. Cf. Mackey (1971, vol. I: 141).

²⁰ *Vid.* Samper Padilla (1999), Hernández Cabrera y Samper Padilla (2007) o Samper Padilla y Hernández Cabrera (2003).

²¹ *Vid.* López Morales (1999: 20-21). Para una muestra aplicada de este tipo de trabajos, véase López Chávez (1993) o Samper Padilla (1998).

²² Galloso Camacho (2003: 185-286).

²³ Galloso Camacho (2003: 285-286).

1.2. Fundamentos

A pesar de su tradición, los estudios sobre disponibilidad léxica adolecen de una teoría general que abarque todas sus dimensiones y sea capaz de responder a los interrogantes esenciales que, a veces, han desmerecido el enorme esfuerzo que supone la tarea de elaborar un listado de estas características. De hecho, la raíz del problema se encuentra en los trabajos fundacionales (franceses, en su mayoría) de la tradición de la disponibilidad léxica, pues en ellos se echa en falta, incluso, una explicación teórica detenida del fundamento léxico del estudio²⁴. Nadie, en aquel momento, se preocupó por explicar cómo está organizado el lexicón, qué representa la lista de disponibilidad léxica de un hablante o qué estructura reproduce. No cabe duda de que las respuestas a estas claves habrían ayudado a deducir cuestiones del estudio de disponibilidad aplicables a la estructura del vocabulario y al uso del léxico en las comunidades de habla.

Los estudios posteriores, incluido el nuestro de 2006, lejos de responder a cuestiones tan elementales, asumieron como propias esas limitaciones epistemológicas de partida y perpetuaron unas deficiencias que han dado como resultado la impresión general de que estos estudios son de importancia menor dentro del ámbito de investigación lingüística²⁵.

1.2.1. Fundamentos cognitivos

Para establecer una definición adecuada del léxico disponible debemos adoptar una perspectiva interdisciplinaria que integre el sistema lingüístico en el entramado complejo de la cognición humana. Cuando un individuo actualiza listas de léxico disponible pone en marcha una tarea cognitiva individual en la que se activan diversos mecanismos del procesamiento de la información. Disciplinas como la psicolingüística, la etno-

²⁴ La tradición sobre los estudios de disponibilidad léxica arranca, como comentamos en el apartado anterior, en la década de los cincuenta del siglo pasado en el ámbito francófono. Para una exposición de los inicios y evolución de estas investigaciones, *vid. supra* § 1.1.

²⁵ No obstante lo anterior, algunos han tratado de dar mayor solidez a los fundamentos teóricos de la disponibilidad léxica. Tal es el caso de Hernández Muñoz (2006), quien se ha esforzado en aclarar los procedimientos psicológicos que sustentan los fundamentos cognitivos del léxico disponible, con objeto de explicar los fenómenos lingüísticos que se manifiestan. Todo ello, con el propósito esencial de conseguir una teoría integrada de la disponibilidad en todas sus dimensiones científicas. El planteamiento que seguimos aquí está inspirado en ese trabajo.

lingüística o la lingüística cognitiva se convierten, de este modo, en esenciales para captar el mecanismo básico de nuestro objeto de estudio²⁶. De entre todas, quizás sea la psicolingüística la disciplina que más nos ayude a entender la verdadera naturaleza del léxico disponible, pues como es sabido, resulta difícil hallar un límite nítido entre la lingüística y la psicología, al menos cuando se consideran los hechos lingüísticos desde la perspectiva del hablante individual.

Sin entrar ahora en la polémica sobre las dicotomías clásicas en el ámbito de la psicolingüística, parece existir un consenso acerca del hecho de que existen dos niveles de representación que contienen la información necesaria para actualizar las palabras asociadas a una categoría semántica: el primer nivel es el lexicón mental, concepto que ha estado presente siempre en las explicaciones de los investigadores en este campo destinadas a construir (generalmente a posteriori) los fundamentos psicológicos de sus listas; el segundo nivel es la memoria semántica, que es donde parece producirse una primera selección del material léxico implicado. El concepto de ‘lexicón mental’ suele ser entendido como el almacén de las formas de una palabra²⁷, en tanto que el concepto de ‘memoria semántica’ se asocia a la representación del significado de las palabras y podría entenderse como un gran almacén cognitivo²⁸. Entre estos dos elementos esenciales se articula el concepto de ‘*priming* semántico’²⁹, que hace referencia a un proceso de accesibilidad a las respuestas particulares que reside en la memoria implícita y que se manifiesta en el léxico disponible³⁰. Aunque la función de la memoria semántica ha sido ampliamente aceptada por la comunidad científica como la encargada de codificar toda la información perceptiva y conceptual que refleja la visión implícita que el individuo tiene del mundo,

²⁶ Hernández Muñoz (2006: 47) no descarta la participación de más disciplinas integradas en las ciencias cognitivas: neurología, filosofía del lenguaje, inteligencia artificial, entre otras.

²⁷ Levelt, Roelofs y Meyer (1999).

²⁸ Sartori, Coltheart, Miozzo y Job (1994: 538).

²⁹ Troyer, Moscovitch y Wincour (1997).

³⁰ «Por ejemplo, si la palabra *azafrán* ha sido estudiada recientemente, es mucho más probable que la persona genere *azafrán* cuando se le pida que enumere flores, incluso aunque no recuerde explícitamente que ha estudiado la palabra» (Holyoack, 2002: 37).

su organización y composición es motivo de un debate acalorado en el que no vamos a entrar en este momento³¹.

Mediante la adaptación de las teorías y los modelos de producción léxica de Levelt³² y de Caramazza³³ a la producción del léxico disponible, Hernández Muñoz considera que los procesos mentales implicados en la tarea de generación de listas de disponibilidad se suceden del siguiente modo³⁴: cuando un individuo recibe el estímulo de un centro de interés dado, se produce una activación de la memoria semántica correspondiente al concepto léxico (v. gr. para el centro de interés ropa, el concepto ‘ropa’) y de otros asociados a él (‘falda’, ‘pantalón’, ‘camisa’, etcétera). Posteriormente, en el nivel del lexicón mental se activan los *lemmas* correspondientes de cada una de esas representaciones semánticas³⁵. En la siguiente fase, dependiendo de si el mensaje adopta una forma escrita u oral, la activación de estos *lemmas* se propaga hacia las formas fonológicas u ortográficas; en ellas se producirá el proceso de selección definitivo que culmina con la actualización de la palabra más disponible para cada sujeto. Es posible que en el proceso de escritura de las listas de disponibilidad no sea necesario el paso por el nivel de representación fonológico, ya que parece que el nivel

³¹ A grandes rasgos, y siguiendo la exposición de Hernández Muñoz (2006), los modelos más importantes de organización de la memoria semántica son: la teoría de los rasgos o teoría descomposicional (*feature theory*: Katz y Fodor, 1963; Chomsky, 1965; Johnson-Laird, 1983); la teoría de las redes semánticas (*semantic networks*: Collins y Quilian, 1969; Collins y Loftus, 1975); la teoría de los prototipos (*Prototypes theory*: Wittgenstein, 1953; Rosch, 1978; Lakoff, 1987); los postulados de significado (*Meaning postulates*: Carnap, 1952). De estos cuatro modelos de representación de la información de la memoria semántica, los estudiosos del léxico disponible han prestado especial atención a la teoría de las redes semánticas (López Morales, 1999; Urrutia, 2001; Galloso, 2003; Gómez Devis, 2003) y a la teoría de los prototipos (Romero Rubilar, 2000; Hernández Muñoz, 2006).

³² Levelt (1999, 2001), Levelt *et alii* (1999).

³³ Caramazza (1997).

³⁴ Hernández Muñoz (2006, 77 y ss.).

³⁵ Los *lemmas* son definidos como unidades funcionales que median entre conceptos y formas y, además, contienen la información sintáctica de las palabras. Dado que un mismo concepto puede activar diferentes *lemmas*, se supone que cada individuo elegirá en cada momento aquel que se adecue mejor al contexto en función de múltiples parámetros. Aunque en las pruebas de disponibilidad los sujetos no tienen que integrar la información sintáctica, parece que esta se activa de manera inconsciente, si bien en este tipo de pruebas no se necesita poner en marcha todo el proceso sintáctico, sino que los rasgos gramaticales se trasladan al momento en que escribimos o emitimos una cadena de habla (Hernández Muñoz, 2006: 80).

semántico está directamente comunicado con el almacén grafémico de la lengua³⁶.

Una vez que se activa la palabra más disponible, comienza el proceso de asociación de términos. Como señala Hernández Muñoz (2006: 87 y ss.), en el proceso de producción de las listas de léxico disponible se ponen de manifiesto dos de las tareas de activación de léxico más aceptadas en el ámbito de la psicología experimental: la fluencia categorial semántica (*category fluency*)³⁷ y la asociación libre de palabras (*free word association*)³⁸. La fluencia semántica consiste en asociar palabras relacionadas con una dada; la asociación libre consiste en actualizar las palabras que primero vengan a la mente después de escribir o escuchar otro término que actúa como estímulo. Hernández Muñoz (2006: 88) cree que la producción del léxico disponible, tal y como se concibe en la actualidad, es una prueba de fluencia semántica, aunque en el proceso subyacen algunos de los presupuestos de la asociación de palabras. En cualquier caso, como veremos más adelante, el planteamiento cognitivo tiene repercusiones metodológicas muy importantes, pues si consideramos que en la edición de los materiales solo tienen cabida las palabras que pertenecen a un espacio de categorización (experiencia) estaremos asumiendo que la disponibilidad léxica es una tarea de asociaciones categoriales; si, por el contrario, aceptamos que en nuestras listas se editen palabras asociadas a una categoría cualquiera, nuestra concepción de la disponibilidad léxica será entendida como una actividad que acepta también la asociación libre, pues se admitirían palabras no necesariamente incluidas en el campo de significado definido por el centro de interés³⁹. Las repercusiones metodológicas señaladas pueden incidir directamente tanto en la elaboración de las listas de disponibilidad como en los cálculos cuantitativos posteriores, tal y como se observa a

³⁶ Otras corrientes (Ellis, 1982) consideran que el nivel fonológico media entre el semántico y el grafémico. Para una teoría conciliadora de ambos planteamientos, *vid.* Cuetos (1991: 38), donde se asegura que en la escritura productiva ambas rutas se activan de manera conjunta.

³⁷ Warrington y McCarthy (1987); Damasio (1990); Saffran y Schwartz (1994); Garrard, Perry y Hodges (1997); Caramazza (1998).

³⁸ Deese (1962), Cramer (1968), Clark (1970).

³⁹ Evidentemente, la creación definitiva de un protocolo de instrucciones comunes para la realización de las encuestas y la edición de los materiales aclararía esta cuestión y solucionaría algunos de los problemas clásicos del proceso de edición (Hernández Muñoz, 2006: 122).

continuación. El Ejemplo 1.1 se refiere a la producción léxica del sujeto 015 en el centro de interés 14. Los animales:

Ejemplo 1.1 Procedimientos de edición

Lista de palabras original:

22224 015 14 perro, gato, pájaro, ratón, rata, cucaracha, saltamontes, elefante, tigre, león, conejo, doméstico, salvaje, vertebrado, invertebrado, rana, sapo, tortuga, salamanquesa, oso, lobo, leopardo, guepardo, koala, marítimo, ballena, delfin, tiburón, caballo, caballito de mar, cerdo, vaca, buey, yegua, camello, mamífero, vivíparo, ovíparo, insectívoro, omnívoro, carnívoro, depredador, sangre fría, sangre caliente, animal de compañía, zoo(lógico), selva, insecto

V1, donde se aceptan las asociaciones indirectas: (48 palabras)

22224 015 14 perro, gato, pájaro, ratón, rata, cucaracha, saltamontes, elefante, tigre, león, conejo, doméstico, salvaje, vertebrado, invertebrado, rana, sapo, tortuga, salamanquesa, oso, lobo, leopardo, guepardo, koala, marítimo, ballena, delfin, tiburón, caballo, caballito de mar, cerdo, vaca, buey, yegua, camello, mamífero, vivíparo, ovíparo, insectívoro, omnívoro, carnívoro, depredador, sangre fría, sangre caliente, animal de compañía, zoo(lógico), selva, insecto

V2, donde no se aceptan las asociaciones indirectas: (31 palabras)

22224 015 14 perro, gato, pájaro, ratón, rata, cucaracha, saltamontes, elefante, tigre, león, conejo, rana, sapo, tortuga, salamanquesa, oso, lobo, leopardo, guepardo, koala, ballena, delfin, tiburón, caballo, caballito de mar, cerdo, vaca, buey, yegua, camello, insecto

Si el investigador adopta una postura abierta y acepta asociaciones directas e indirectas, considerará que el sujeto 015 participa en el estudio con una aportación de cuarenta y ocho palabras; por el contrario, un planteamiento estrictamente categorial de asociaciones directas eliminará diecisiete elementos (*doméstico, salvaje, vertebrado, invertebrado, marítimo, mamífero, vivíparo, ovíparo, insectívoro, omnívoro, carnívoro, depredador, sangre fría, sangre caliente, animal de compañía, zoo(lógico), selva*) y atribuirá a la producción léxica del mismo sujeto solo treinta y una palabras⁴⁰.

⁴⁰ Desde luego, entre las dos posturas expuestas caben otras: por ejemplo, que se acepten tipos de animales, pero no sus hábitats, naturales o no, en cuyo caso solo eliminaríamos *zoo(lógico), marítimo y selva*, o que se acepten modos de alimentación, o constituciones anatómicas, etcétera.

En cualquier caso, como vemos, la producción del léxico disponible ha heredado los problemas intrínsecos que poseen las pruebas de asociación de palabras⁴¹. En realidad, nunca estaremos seguros de si los informantes van a desarrollar estrategias que falseen los procesos cognitivos de fondo, ya que los mismos resultados pueden proceder de estrategias dispares. No olvidemos que la tarea de actualización del léxico disponible procede de una actividad artificial que puede no ser exactamente igual al proceso de acceso al léxico en el lenguaje cotidiano⁴².

1.2.2. *Fundamentos matemáticos*

Las diversas aproximaciones cuantitativas orientadas a realizar clasificaciones del léxico disponible han tratado de poner de manifiesto que, al abordar determinados temas, las palabras que primero acuden a nuestra mente son más disponibles que aquellas otras que no hacen su aparición de forma inmediata⁴³. El índice de disponibilidad es, a grandes rasgos, un índice numérico que trata de poner en relación criterios de frecuencia y orden. El cálculo matemático en el que se basa pretende ponderar de manera adecuada la frecuencia con la que se actualiza una palabra en un centro de interés con su posición en las diversas listas.

En consecuencia, desde los trabajos de Lorán (1983) y Lorán y López Morales (1983) se han diseñado estrategias de ponderación de la frecuencia que, en un primer momento, consistieron en introducir coeficientes con valor inferior a la unidad (valor λ), que era multiplicado por la frecuencia alcanzada por la unidad en cuestión. Como este coeficiente no actuaba cuando las unidades aparecían en primer lugar, la frecuencia original quedaba inalterada. Si la palabra aparecía en segundo lugar, el

⁴¹ Hernández Muñoz (2006: 124).

⁴² «Por consiguiente, toda interpretación de los datos obtenidos en los estudios de disponibilidad léxica que conlleve una reflexión sobre la organización del lexicón mental o la memoria semántica ha de hacerse con suma cautela» (Hernández Muñoz, 2006: 124).

⁴³ En efecto, ya hemos señalado que pronto se observó que el criterio lexicostatístico de selección basado únicamente en la frecuencia (absoluta o relativa) era insuficiente. Michéa (1953), primero, y Gougenheim *et alii* (1956), después, hacen referencia a la importancia de usar un factor corrector de la simple frecuencia que tuviese en cuenta la presencia más o menos inmediata de las palabras en la memoria (*vid.* § 1.1). A pesar de estas advertencias previas y de la insistencia con la que Müller (1968) retomó la cuestión, hasta 1983 no se diseñan fórmulas eficaces para ponderar ambos factores. *Vid.* Lorán (1983) y Lorán y López Morales (1983), *apud* López Morales (1999: 15-19).

coeficiente actuaba con el valor establecido para λ , de manera que en las sucesivas posiciones (tercera, cuarta, etcétera), disminuía su valor a partir de su elevación al cuadrado hasta la n potencia⁴⁴.

La fórmula inicial, que partía de un marco estadístico basado en la teoría de las decisiones⁴⁵, fue modificada en diversas ocasiones para poder operar con listas abiertas y desiguales. Una vez mejorada, la fórmula consideraba que la disponibilidad de una palabra (p) era igual a la suma del número de veces que p aparece en primer lugar, más una serie de factores integrados por un coeficiente particular (λ , menor que la unidad), que va disminuyendo de forma exponencial multiplicado por la cantidad de veces que p aparece en segundo lugar, tercero, etcétera:

$$d(p) = \sum_{i=1}^n \lambda^{i-1} \frac{f_{pi}}{N_i}$$

En estos trabajos iniciales, el valor de λ fue .90. Si la unidad x aparecía 100 veces en el primer lugar de todas las listas, su valor era 100, pues en ese caso solo se contaba la frecuencia. Sin embargo, si aparecía 100 veces en segundo lugar, su índice era de .90, y así sucesivamente⁴⁶.

La propuesta, sin embargo, perdía poder discriminador a partir de la vigesimotercera posición, pues a partir de ese momento, la curva se mantenía prácticamente paralela al eje de abscisas debido al comportamiento intrínseco a la función exponencial, que disminuye rápidamente su valor tras esos primeros pasos. Tras diversas revisiones de la fórmula inicial⁴⁷, López Chávez y Strassburger (1987) proponen partir de una matriz de vectores en la que participan:

- a) la frecuencia absoluta que obtiene la unidad en cada posición de las listas;

⁴⁴ En López Morales (1991: 13-19) se explica con detalle el proceso de gradación de la disponibilidad léxica y se ofrecen ejemplos ilustrativos del modelo obtenidos de investigaciones previas.

⁴⁵ Lorán (1987), Butrón (1987, 1991).

⁴⁶ Para una explicación detallada de los trabajos encaminados a comprobar la efectividad de la fórmula propuesta, *vid.* Lorán (1987) y López Morales (1999).

⁴⁷ Butrón (1987, 1991).

- b) la frecuencia absoluta de la unidad; esto es, la suma de las diferentes frecuencias alcanzada en cada posición;
- c) el número de individuos considerados;
- d) el número de posiciones alcanzadas en el centro de interés analizado;
- e) las posiciones en que fue escrita la unidad.

La aplicación del factor e), elevado a un exponente complejo, es la responsable de la auténtica ponderación de la posición y lo que permite combinar convenientemente la frecuencia y la posición de cada unidad, pues ofrece una ponderación variable entre 1 y 0.1, con independencia del número de participantes, de la extensión de cada listado individual y de la frecuencia de aparición de cada elemento léxico.

La forma final de la fórmula propuesta queda expresada así:

$$d(p) = \sum_{i=1}^n \exp \left(C - \frac{i-1}{n!} \right) \frac{f_{pi}}{N_i},$$

donde n es la máxima posición alcanzada, f_{pi} es la frecuencia absoluta de la palabra en la posición i y N_i es el número de hablantes investigados.

1.2.2.1. Revisión del modelo clásico

La necesidad que tuvieron los precursores franceses de los estudios de disponibilidad léxica de ordenar los términos más disponibles y ofrecer listas basadas en cuantificaciones hizo que, como hemos visto, se propusieran fórmulas útiles para tales fines que poco a poco fueron mejorándose. No obstante, los estudios que se limitan a la utilización de fórmulas matemáticas para la obtención de listados de palabras resultan, desde un punto de vista estrictamente científico, algo simples. Como es lógico, estas limitaciones han sido observadas por muchos autores que han aprovechado las posibilidades que ofrecen los diversos programas diseñados para el cálculo de la disponibilidad (por ejemplo, *Lexidisp*) para ir un poco más lejos y realizar correlaciones entre variables léxicas y sociales⁴⁸. Sin embargo, las restricciones que imponen estos programas a propósito del número de

⁴⁸ Gómez Molina y Gómez Devis (2004); Hernández Muñoz (2006); Ávila Muñoz (2006), entre otros.