

Antonio María López González

# DISPONIBILIDAD LÉXICA

---

## TEORÍA, MÉTODO Y ANÁLISIS



WYDAWNICTWO  
UNIWERSYTETU  
ŁÓDZKIEGO

# DISPONIBILIDAD LÉXICA

---

## TEORÍA, MÉTODO Y ANÁLISIS



WYDAWNICTWO  
UNIWERSYTETU  
ŁÓDZKIEGO

[Kup książkę](#)

Antonio María López González

# DISPONIBILIDAD LÉXICA

---

## TEORÍA, MÉTODO Y ANÁLISIS



WYDAWNICTWO  
UNIWERSYTETU  
ŁÓDZKIEGO

ŁÓDŹ 2014

Antonio María López González – Uniwersytet Łódzki, Wydział Filologiczny  
Katedra Filologii Hiszpańskiej, 90-236 Łódź, ul. Pomorska 171/173

RECENZENT

*Jerzy Szalek*

REDAKTOR WYDAWNICTWA UŁ

*Bogusława Kwiatkowska*

SKŁAD KOMPUTEROWY

*Oficyna Wydawnicza Edytor.org*

*Lidia Ciecierska*

PROJEKT OKŁADKI

*Lukasz Orzechowski*

Zdjęcie wykorzystane na okładce: © Depositphotos.com/zagandesign

© Copyright by Uniwersytet Łódzki, Łódź 2014

Wydane przez Wydawnictwo Uniwersytetu Łódzkiego

Wydanie I. W.06638.14.0.M

ISBN 978-83-7969-437-2

e-ISBN 978-83-8088-088-7

Wydawnictwo Uniwersytetu Łódzkiego

90-131 Łódź, ul. Lindleya 8

www.wydawnictwo.uni.lodz.pl

e-mail: ksiegarnia@uni.lodz.pl

tel. (42) 665 58 63, faks (42) 665 58 62

# ÍNDICE

<b>Introducción</b> .....	9
i. <b>Ámbito de la investigación</b> .....	9
ii. <b>Esta investigación: Presentación y propósitos</b> .....	12
<b>I. Los estudios de disponibilidad léxica</b> .....	17
1.1. <b>La medición estadística del vocabulario</b> .....	17
1.1.1. <b>Los diccionarios normativos</b> .....	17
1.1.2. <b>La frecuencia léxica</b> .....	19
1.1.3. <b>El léxico básico</b> .....	22
1.1.4. <b>La disponibilidad léxica</b> .....	24
1.2. <b>El estudio de la disponibilidad léxica en español</b> .....	26
1.2.1. <b>Los estudios anteriores al Proyecto Panhispánico</b> .....	26
1.2.2. <b>El Proyecto Panhispánico de Disponibilidad Léxica</b> .....	30
<b>II. Bases teóricas y metodológicas de la disponibilidad léxica</b> .....	35
2.1. <b>Unidades de análisis del corpus léxico</b> .....	35
2.2. <b>El procesamiento del corpus léxico</b> .....	36
2.3. <b>Criterios de edición del Proyecto Panhispánico</b> .....	40
2.4. <b>El índice de disponibilidad léxica</b> .....	43
2.5. <b>El tamaño de la muestra</b> .....	45
<b>III. Los centros de interés</b> .....	49
3.1. <b>Concepto y selección</b> .....	49
3.2. <b>Clasificación de los centros de interés</b> .....	54
3.3. <b>Regularidades en los centros de interés</b> .....	57
3.3.1. <b>Promedio de respuestas</b> .....	58
3.3.2. <b>Número total de palabras</b> .....	65
3.3.3. <b>Número de vocablos</b> .....	68
3.3.4. <b>Índice de cohesión</b> .....	76
<b>IV. Relaciones entre las palabras en el léxico disponible</b> .....	83
4.1. <b>La fluencia semántica y la teoría de los prototipos</b> .....	84
4.1.1. <b>La elaboración de las listas</b> .....	84
4.1.2. <b>La categorización</b> .....	86
4.2. <b>La asociación de palabras y las redes semánticas</b> .....	90
4.2.1. <b>La asociación de palabras en las listas de léxico disponible</b> .....	90
4.2.2. <b>La representación gráfica de la red semántica</b> .....	94
4.2.3. <b>Las relaciones significativas</b> .....	97

<b>V. Los condicionantes extralingüísticos de la disponibilidad léxica</b> .....	103
5.1. El sexo .....	104
5.2. El nivel sociocultural .....	108
5.3. El factor espacial: urbano/rural .....	112
5.4. El tipo de centro escolar .....	115
5.5. La edad .....	119
5.6. Los condicionantes extralingüísticos en las comunidades bilingües .....	124
5.6.1. ‘Lengua materna’ .....	125
5.6.2. ‘Lengua de uso familiar’/‘Lengua habitual’ .....	129
5.6.3. ‘Modelo de enseñanza’ (monolingüe/bilingüe) .....	133
5.6.4. Las transferencias léxicas en las comunidades bilingües .....	136
<b>VI. El análisis cualitativo del léxico disponible</b> .....	143
6.1. El análisis cualitativo: fines y posibilidades de estudio .....	143
6.2. La composición de los listados de léxico disponible .....	145
6.3. Clasificación del léxico disponible .....	156
6.3.1. Análisis según el índice de disponibilidad léxica .....	157
6.3.2. Análisis descriptivo lexicológico .....	161
6.3.3. Análisis descriptivo según niveles de lengua (registros) .....	166
6.4. Los extranjerismos en el léxico disponible del español .....	169
6.5. El léxico peculiar: los dialectalismos .....	176
<b>VII. La comparación de los léxicos disponibles</b> .....	183
7.1. La norma léxica del español peninsular .....	183
7.1.1. La delimitación de la norma léxica según el léxico disponible .....	183
7.1.2. Las comparaciones interdialectales y el léxico compartido .....	185
7.1.3. Cotejo de cinco léxicos disponibles peninsulares .....	186
7.2. Análisis descriptivo de los centros de interés y su norma léxica .....	190
7.2.1. ‘Partes del cuerpo humano’ .....	190
7.2.2. ‘La ropa’ .....	194
7.2.3. ‘Partes de la casa (sin muebles)’ .....	197
7.2.4. ‘Los muebles de la casa’ .....	200
7.2.5. ‘Comidas y bebidas’ .....	203
7.2.6. ‘Objetos sobre la mesa para la comida’ .....	207
7.2.7. ‘La cocina y sus utensilios’ .....	210
7.2.8. ‘La escuela (muebles y materiales)’ .....	214
7.2.9. ‘Iluminación, calefacción y medios de airear un recinto’ .....	217
7.2.10. ‘La ciudad’ .....	221
7.2.11. ‘El campo’ .....	224
7.2.12. ‘Medios de transporte’ .....	227
7.2.13. ‘Trabajos del campo y del jardín’ .....	231
7.2.14. ‘Los animales’ .....	234
7.2.15. ‘Juegos y distracciones’ .....	237
7.2.16. ‘Profesiones y oficios’ .....	242
<b>Conclusiones</b> .....	247
<b>Bibliografía</b> .....	259
<b>Dostępność leksykalna. Teoria, metoda i analiza. Streszczenie</b> .....	279
<b>Lexical availability: theory, method and analysis. Summary</b> .....	283
<b>Od Redakcji</b> .....	287

## AGRADECIMIENTOS

Un trabajo como el que aquí se presenta no sería posible sin la ayuda y el apoyo de muchas personas, a las que quiero dedicar estas líneas, dándoles nombre propio.

Así, quiero dar las gracias a Francisco García Marcos, mi maestro en la sociolingüística, por iniciarme en el camino de la disponibilidad léxica. A él y a M.<sup>a</sup> Victoria Mateo García, mi agradecimiento por compartir tantas ideas y materiales. Igualmente quiero dar las gracias a José A. Samper Padilla, verdadero biógrafo de la disponibilidad léxica en español, por responder siempre a mis consultas y facilitarme sus publicaciones. Gracias también a José A. Bartol Hernández y Natividad Hernández Muñoz, por permitirme el acceso al banco de datos y a las herramientas informáticas de *Dispolex.com*. Asimismo, en el campo informático, quiero agradecer a Marta Sánchez-Saus Laserna y Paula Urzua sus explicaciones sobre el funcionamiento del programa *Dispografo*.

Quiero dar las gracias, en fin, a los estudiosos de la disponibilidad léxica que han compartido conmigo sus materiales y que han respondido a mis consultas. Por este motivo, además de a los arriba mencionados, les estoy agradecido: Antonio M. Ávila Muñoz, Carmen Fernández Juncal, M.<sup>a</sup> Victoria Galloso Camacho, Antonio García Megía, Adolfo González Martínez, Belén López Meirama, Francisco Moreno Fernández, Josefina Prado Aragonés, Ventura Salazar García, Marta Samper Hernández, Javier de Santiago Guervós, Cristina Tabernero Sala, Antonia Torres González, Ester Trigo Ibáñez.

Quiero tener un recuerdo aquí para dos personas, que ya no están con nosotros, y que han influido mucho con sus investigaciones en el devenir de la disponibilidad léxica y, por consiguiente, en esta monografía. Se trata de Alberto Carcedo González, lectura obligada para mis trabajos sobre la disponibilidad léxica en español como lengua extranjera, y de Max. S. Echeverría, autor del programa *Dispografo*, a quien tuve la enorme suerte de conocer como alumno del taller de Didáctica que impartió en el XIII Congreso de ALFAL.

Gracias, finalmente, a Wiaczeslaw Nowikow, director de la Cátedra de Filología Española de la Universidad de Lódz, por su ánimo e infinita paciencia para conmigo.



Y siempre, un lugar especial en mi corazón, para Janina, mi mujer, y para mis hijos Ana y Antonio, por todas las horas en común que este trabajo les ha robado, y los que, sin embargo, siempre, siempre, me han acompañado en los momentos de fatiga. Os quiero.

## INTRODUCCIÓN

### i. **Ámbito de la investigación**

El estudio de la disponibilidad léxica ha demostrado ser un campo fecundo en las investigaciones lingüísticas hispánicas desde finales de los años 90, tanto desde el punto de vista cuantitativo –número de trabajos– como cualitativo –reflexión teórica y desarrollo metodológico. De ello han levantado acta distintos estudios que, sucesivamente y a partir de una exhaustiva revisión bibliográfica, han dado cuenta del progreso de estas investigaciones.

Así, si en 2003 Samper Padilla, Bellón Fernández y Samper Hernández (2003: 27) se referían a que “varios grupos de investigación de ambos lados del Atlántico [trabajaban] en este aspecto de la léxico-estadística”, lo que había supuesto un cambio radical con respecto a las pocas investigaciones realizadas en ese campo hasta aquel año, sólo tres años más tarde, los mismos Samper Padilla y Samper Hernández (2006: 5) certificaban que “el estudio de la disponibilidad léxica [había] experimentado un desarrollo muy destacado en el mundo hispánico”. Desde entonces, la investigación en este campo no ha hecho más que crecer, lo que hace que ya en 2012 Paredes García (2012: 78–79) pueda presumir del “estado efervescente en que se encuentra la investigación de la disponibilidad léxica” y de “la eclosión de la disponibilidad léxica en España y América”, habida cuenta del amplio censo de monografías, tesis doctorales, memorias de licenciatura o máster, artículos en revistas u obras colectivas, etc., siempre en constante crecimiento.

El extraordinario desarrollo de este campo de investigación en el mundo hispánico se debe sobre todo al *Proyecto Panhispánico de Estudio sobre la Disponibilidad Léxica*, un macroproyecto que coordina y dirige Humberto López Morales, con equipos inscritos en las diferentes zonas dialectales del mundo hispánico. No obstante, “hay que destacar especialmente por sus aportes teóricos y/o metodológicos las investigaciones llevadas a cabo en Puerto Rico, la República Dominicana,

na, Chile y México”, así como España, donde “el territorio estudiado cubre más de cincuenta provincias españolas [...] lo que posibilitará en un futuro próximo la publicación del *Diccionario del léxico disponible de España*” (Paredes García, 2012: 79). El fin último de este macroproyecto es la elaboración de un diccionario del léxico disponible del español.

A la coordinación de los estudiosos de la disponibilidad léxica en el ámbito hispanohablante ha contribuido también desde octubre de 2000 la página electrónica [www.dispalex.com](http://www.dispalex.com), creada y mantenida por José Antonio Bartol Hernández y Natividad Hernández Muñoz desde la Universidad de Salamanca. Esta página sirve como medio de contacto e información a todas las personas que investigan o se interesan por la disponibilidad léxica, proporcionando detalles sobre la forma más habitual de proceder en este campo, y acceso gratuito a las herramientas necesarias para realizar los cálculos más habituales en disponibilidad léxica y a un banco de datos de las investigaciones realizadas.

Hablar de *disponibilidad léxica* significa referirse al “caudal léxico *utilizable* en una situación comunicativa dada” (López Morales, 1996: 245), a aquellas palabras que acuden antes a la mente del hablante como reacción a un estímulo temático concreto (López Morales, 1986, 1999a). Estas palabras constituyen el *vocabulario disponible* de ese hablante o de esa comunidad de hablantes. Se entiende, entonces, “que existe en el lexicón mental una serie de términos que no se actualizan a menos que sea necesario para comunicar una información muy específica” (López Morales, 2009: 26).

Los estudios de disponibilidad léxica se iniciaron en Francia a mediados del siglo XX, cuando un grupo de lingüistas recibió el encargo de elaborar *Le Français Élémentaire*, un instrumento de planificación didáctica para la enseñanza del francés como L2, en las excolonias de l’Union Française, para los inmigrantes llegados a Francia, y para los extranjeros en general (Gougenheim *et al.*, 1956: 8–9). Para seleccionar los vocablos que debían integrar el léxico “elemental”, recurrieron a los listados léxicos de frecuencias, pero se dieron cuenta de que en esos recuentos faltaban términos cotidianos y familiares, “palabras que se referían a realidades concretas y que no aparecían en los listados basados en la frecuencia porque su uso estaba muy condicionado por el tema de la conversación o del escrito” (Bartol Hernández, 2006: 379).

Enseguida, los investigadores franceses se percataron de que el léxico fundamental de una lengua incluye dos subconjuntos léxicos fácilmente distinguibles (Michéa, 1953):

a) el *léxico básico*, frecuente y atemático, principalmente palabras gramaticales (adjetivos, verbos y sustantivos de carácter general), formado por palabras que aparecen continuamente en cualquier conversación o escrito, independientemente del tema de que se trate (ej. – *a, los, no, mucho, hay, dar, persona, poner*, etc.);

b) el *léxico disponible*, temático, de contenido semántico muy concreto, integrado por palabras que, aun siendo de uso común, sólo se emplean ligadas a un tema (ej. – *sartén* y *tenedor* ligadas a la ‘cocina’, o *carta* y *sello* relacionadas al ‘correo’).

Para recabar ese léxico disponible, los estudiosos franceses prepararon en 1953 unas pruebas asociativas a las que respondieron 416 niñas y 488 niños de 9 a 12 años en cuatro departamentos franceses. Con este estudio, y los realizados posteriormente por Mackey (1971a, 1971b) en áreas francoparlantes de Canadá, se cimentó la base metodológica de la disponibilidad léxica: la recogida de materiales mediante encuestas de listas abiertas en las que durante dos minutos los informantes anotan todas las palabras que les vienen a la mente en relación con una serie de estímulos temáticos.

Como se aprecia, la disponibilidad léxica nació –y se mantiene– ligada a una clara finalidad pedagógica: la necesidad de seleccionar los vocablos que deben ser enseñados en la clase de lengua (extranjera o materna). Al mismo tiempo, se caracteriza por indagar en el mundo conceptual-real significativo de una determinada comunidad (Gómez Devís, 2004: 8). La utilidad de estas investigaciones es, por tanto, innegable para la lingüística aplicada dada la importancia excepcional que revisten las aplicaciones pedagógicas. Junto a estas, los estudios sobre disponibilidad léxica realizan distintas tareas que, en su día, Azurmendi Ayerbe (1983: 187–195) clasificó en los siguientes tres grupos:

- El establecimiento de léxicos básicos, para los que la disponibilidad léxica refleja mejor el uso oral y es más representativa de las realidades sociales.
- La reflexión teórica sobre las regularidades en el comportamiento de la disponibilidad y la validez del método de asociación controlada, sobre las relaciones entre el comportamiento de la disponibilidad y el tamaño de la muestra, así como sobre el tamaño de respuestas óptimas.
- La comparación inter e intralingüística, sea buscando la comparación y la distancia interlingüística, sea estudiando la comparación intralingüística al analizar las variedades geolingüísticas de la misma lengua.

Además, se han multiplicado los trabajos de índole dialectológica, sociolingüística, psicolingüística y etnolingüística (cf. Samper Padilla *et al.*, 2003: 105 y ss.). Al respecto, Ávila Muñoz y Villena Ponsoda (2010: 42–44) enumeran las siguientes aplicaciones:

1) La sociolingüística “somete la disponibilidad a análisis de covariación con los factores sociales que le interesa manejar y determina la caracterización léxica de los sociolectos que integran la comunidad en cuestión” (López Morales, 1999a: 25). Trabajando principalmente a base de cuantificación y no (siempre) de variación cualitativa, diversos estudios<sup>1</sup> han demostrado que, “por encima de cualquier

---

<sup>1</sup> Entre otros, destacan: 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 (2003a).

otro condicionante sociocultural, 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” (Ávila Muñoz y Villena Ponsoda, 2010: 43).

2) La psicolingüística se interesa por la producción de la lista de léxico durante la prueba de disponibilidad léxica. Esta prueba conlleva una tarea cognitiva compleja en la que intervienen determinados procesos psicológicos que, en gran medida, condicionan la mencionada lista, y que han sido investigados, entre otros, por Hernández Muñoz (2006) y Ávila Muñoz y Villena Ponsoda (2010).

3) La etnolingüística estudia las diferencias culturales que se observan entre las comunidades de habla estudiadas a través de los listados de vocablos disponibles, en los que el léxico disponible de cada sociedad aparece como un reflejo de su cultura. Las relaciones entre vocabulario y cultura fueron señaladas ya por Michéa (1950) e investigadas por Mackey (1971a).

4) La dialectología comparada realiza comparaciones diatópicas de los materiales proporcionados por la disponibilidad léxica. El punto de inflexión en estos estudios comparados lo marcó el desarrollo de los índices matemáticos de López Chávez (1992, 1993), 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 dialectos”<sup>2</sup> (Ávila Muñoz y Villena Ponsoda, 2010: 44).

5) La lexicología analiza la composición léxica de las listas, especificando el léxico patrimonial, los grupos léxicos, los acrónimos, las marcas, los extranjerismos, etc. En especial, se han desarrollado estudios sobre la penetración y el alcance de los extranjerismos<sup>3</sup> o la presencia de faltas de ortografía en las listas, como en el trabajo de Galoso Camacho (2003a: 185–286), en el que esta autora advierte del “panorama desolador” en materia ortográfica y hace una llamada de atención a todos los “organismos competentes” para empezar a tomar medidas sin demora.

## ii. Esta investigación: Presentación y propósitos

El volumen que tiene el lector entre sus manos tiene como objetivo analizar de una manera crítica las bases teóricas y metodológicas en las que se apoyan los estudios de léxico disponible. A partir de una amplia revisión bibliográfica, se pasa revista a las aportaciones más destacadas del Proyecto Panhispánico, presentando los diferentes ámbitos lingüísticos en los que se desarrollan estos estudios.

<sup>2</sup> Vid. López Morales (1999: 20–21).

<sup>3</sup> Entre otros, destacan los trabajos de Paredes García (2001b), Carrera y Bradley (2004), Bartol Hernández (2005) y Castañer Martín (2008).

En este sentido, pretende completar y actualizar los trabajos anteriores de Samper Padilla *et al.* (2003), y de Samper Padilla y Samper Hernández (2006), en los que se trazaba un amplio panorama de los logros de la disponibilidad léxica en lengua española<sup>4</sup>.

Como hemos anticipado, la disponibilidad léxica nació con fines pedagógicos en Francia (Gougenheim *et al.*, 1956). Estos mismos objetivos inspiraron a Humberto López Morales (1973) a llevar a cabo la primera investigación de disponibilidad léxica en el mundo hispánico, en la que analiza la adquisición del vocabulario y las variables que en él inciden, a partir de una muestra de escolares de la zona metropolitana de San Juan de Puerto Rico. Naturalmente, a esta finalidad didáctica hacemos alusión en este libro.

Sin embargo, con ser la perspectiva didáctica de importancia máxima para el desarrollo de estos estudios, en esta ocasión nos ocupamos solo de manera secundaria de este tipo de aplicaciones, ya que el objetivo de esta monografía es el análisis de la fundamentación teórica y metodológica de estos trabajos. Se trata, pues, de una investigación de corte epistemológico, en el sentido en que este estudio resulta en una aportación sobre el origen, la estructura, los métodos y la validez misma de la investigación sobre el léxico disponible. Para ello, este trabajo ha buceado en los estudios introductorios de los trabajos que aquí se mencionan, comparando los planteamientos teóricos de los que parten, la metodología empleada que de ellos se deriva, y los resultados que con su aplicación se obtienen.

Estructuralmente, este libro se compone de siete capítulos, aparte de la Introducción y el capítulo de Conclusiones. El capítulo I (*Los estudios de disponibilidad léxica*), está dedicado, por una parte, a la evolución histórica de los procedimientos principales que miden el vocabulario existente en una comunidad de habla. Entre estos procedimientos, el léxico disponible estudia un léxico de contenido semántico muy concreto, existente en el lexicón mental, pero de carácter potencial. Por otra parte, se da testimonio de cómo el español ha extraído el mejor provecho de la investigación cuantitativa del léxico por medio de la disponibilidad léxica. Después del trabajo inicial de 1973, Humberto López Morales siguió investigando en torno a diversos aspectos descriptivos y metodológicos de la disponibilidad léxica, investigaciones que alcanzaron su punto culminante con la publicación del *Léxico disponible de Puerto Rico* (López Morales, 1999a). Entre una y otra fecha, los estudios sobre el léxico disponible se extendieron por distintos países de Hispanoamérica (Puerto Rico, República Dominicana, Costa Rica, México, Chile), llegando a España en 1992 con los primeros estudios de Benítez Pérez (1992) en Madrid, y Samper Padilla y Hernández Cabrera (1992) en Las Palmas de Gran Canaria. Desde entonces, los estudios de léxico disponible han proliferado por toda la geografía española, sobre todo, gracias a la existencia de

---

<sup>4</sup> Con anterioridad, López Morales (1995, 1999a) y Carcedo González (1998b) ya habían descrito el devenir histórico de la disponibilidad léxica.

un ambicioso proyecto panhispánico, el Proyecto de Estudio de la Disponibilidad Léxica en Español, dirigido por el profesor López Morales.

Los capítulos II (*Bases teóricas y metodológicas de la disponibilidad léxica*) y III (*Los centros de interés*) abordan la renovación epistemológica y metodológica que han experimentado los estudios de léxico disponible en el marco del Proyecto Panhispánico, dedicados a indagar y definir la norma comunitaria adulta de las distintas comunidades de lengua española. No obstante, la discusión metodológica no ha estado exenta de controversias, como la que afectó al diseño de la fórmula matemática utilizada en los estudios estadísticos de disponibilidad léxica (Lorán y López Morales, 1983; López Chávez y Strassburger, 1987), o la que se ha manifestado a través de las múltiples propuestas sobre la cantidad y naturaleza de los campos léxicos investigados<sup>5</sup>. Incluso los criterios de edición del Proyecto Panhispánico (Samper Padilla, 1998a: 311–313), que otorgan homogeneidad al proyecto, se han visto sometidos a frecuentes matizaciones y adaptaciones, como veremos más adelante.

Una parte inherente y constitutiva de los estudios de léxico disponible es el uso de herramientas y procedimientos de análisis estadístico, que incluyen tanto la estadística descriptiva como el uso de fórmulas especializadas (índice de disponibilidad, índice de cohesión, análisis multivariante, etc.). Como afirma Mateo García (1998: 14), “mueve a cierta sorpresa un análisis que coloca entre números, índices y magnitudes sociolingüísticas una realidad tan familiar y cotidiana como la abordada por la disponibilidad léxica”. En cualquier caso, este tratamiento científico sirve para descubrir las características estadísticas más o menos constantes entre los estudios realizados en diferentes comunidades que afectan tanto a la producción léxica de los centros de interés como al comportamiento de las variables sociolingüísticas consideradas. Estos aspectos son analizados en el mencionado capítulo III y en el V (*Los condicionantes extralingüísticos de la disponibilidad léxica*), a partir de la comparación de los resultados de las investigaciones españolas del Proyecto Panhispánico.

El capítulo IV (*Relaciones entre las palabras en el léxico disponible*) se inspira en los estudios de Hernández Muñoz (2006), y Ávila Muñoz y Villena Ponsoda (2010), e indaga en las líneas de investigación abiertas por estos estudiosos, encaminadas a descubrir la estructura del lexicón mental partiendo de los datos proporcionados por los estudios de disponibilidad, y mediante la aplicación de los modelos psicolingüísticos más relevantes en la producción del léxico. El capítulo muestra que el análisis de las asociaciones que se producen entre las palabras, sea

---

<sup>5</sup> Al respecto, entre otras, podemos adelantar la propuesta de Ayora Esteban (2006), quien manifiesta la necesidad, por un lado, de recabar información contextual y, por otro, de enunciar centros de interés que ayuden al investigador a compilar listas, no solo de sustantivos, sino también de otras categorías lingüísticas como adjetivos o verbos. A estas características corresponderían centros como ‘Los colores’, ‘Defectos y cualidades físicos y morales’ y ‘Acciones’, que la citada autora incorpora a su investigación.

en el interior de una lista léxica producida por un informante, sea dentro del conjunto léxico perteneciente a un colectivo, es una potente herramienta para deducir cuál es la realidad psicológica del léxico disponible. Estudios como el de Hernández Muñoz (2006), que consideran el léxico disponible como un fenómeno cognitivo complejo constituyen la aportación más seria a la consecución de la teoría integrada de la disponibilidad léxica preconizada por López Morales (1999a: 26), en la que se conjuguen factores como la memoria semántica, la familiaridad, etc.

El capítulo V (*Los condicionantes extralingüísticos de la disponibilidad léxica*) analiza la incidencia de los factores sociales en la producción léxica. Una de las características principales del Proyecto Panhispánico de Disponibilidad Léxica es la adopción de una perspectiva sociolingüística de corte variacionista, que intenta mostrar la estratificación sociolectal de la disponibilidad léxica de la comunidad estudiada. Según las normas comunes del Proyecto Panhispánico, las variables sociales obligatorias en cada investigación son ‘sexo’, (hombre/mujer), ‘centro escolar’ (público/privado), ‘zona geográfica’ (urbano/rural) y ‘nivel sociocultural’ (alto/medio/bajo). Este capítulo se corresponde con el análisis cuantitativo que de estas variables se hace en las diferentes investigaciones. En el presente estudio, mediante la comparación de los datos procedentes de estas investigaciones, se señalan las tendencias que con respecto a cada variable marcan los distintos estudios. Se analizan, además, otras variables, como ‘la edad’ y los condicionantes extralingüísticos que se tienen en cuenta en las situaciones de contacto de lenguas.

En el capítulo VI (*El análisis cualitativo del léxico disponible*) estudiamos una serie de factores lingüísticos que determinan la composición de los listados de léxico disponible que conforman los diccionarios que resultan de las distintas investigaciones. Como afirma García Marcos (1997: 25), además de los factores sociales, es necesario tomar en cuenta en los estudios de léxico disponible “los hechos estrictamente lingüísticos que, no sólo son —o pueden ser— responsables de la configuración que adquiera un determinado nivel de léxico disponible, sino que están en la base misma de la dinámica lingüística y, sociolingüística, por supuesto, de una determinada comunidad de habla”. La naturaleza de los hechos lingüísticos que analizamos permite diferentes clasificaciones de los vocablos que integran las listas, y que reflejan la variación (diatópica, diastrática, diafásica), la importación léxica, la formación y creación de palabras, etc.

Por último, antes de las conclusiones finales, el capítulo VII (*La comparación de los léxicos disponibles*) analiza la metodología empleada en las comparaciones interdialectales. El capítulo presenta, asimismo, el cotejo de cinco léxicos disponibles de España —fin último del Proyecto Panhispánico de Disponibilidad Léxica—, en lo que representa un ensayo para establecer la norma léxica del español de España. Este ejercicio nos sirve para mostrar lo que de común y diverso tienen estos léxicos disponibles, y para ilustrar lo que será la fase final del mencionado macroproyecto: la elaboración de diccionarios de léxico disponible que incluyan los resultados de las investigaciones precedentes.





# I. LOS ESTUDIOS DE DISPONIBILIDAD LÉXICA

## 1.1. La medición estadística del vocabulario

La aparición y desarrollo de los estudios de disponibilidad léxica se relacionan con la necesidad de optimizar el proceso de adquisición de la lengua materna, de una parte, y el aprendizaje como segunda lengua por hablantes extranjeros, de otra. Esta necesidad, en lo referente a la programación de la enseñanza del léxico, se ha traducido en la evolución y desarrollo de los procedimientos encaminados a hacer efectiva una adecuada selección de los vocablos que se deben enseñar tanto en las clases de lengua materna como en las de lengua extranjera.

¿Qué palabras hay que enseñar? ¿Cuáles no? ¿Por qué unas sí y otras no? A estas preguntas ha tratado de responder la léxico-estadística, y en la evolución de los cuatro procedimientos principales para la medición estadística del vocabulario existente en una comunidad de habla –los diccionarios normativos, la frecuencia léxica, el léxico básico y la disponibilidad léxica– (López Morales, 1986: 53–66) se pueden seguir los intentos de plasmar de forma precisa el uso real que los hablantes hacen de la lengua.

### 1.1.1. Los diccionarios normativos

De los diferentes tipos de diccionarios existentes (bilingües, de sinónimos, etimológicos, de colocaciones, de rimas, de dudas, etc.), son los diccionarios generales o normativos los que se ocupan de recoger de manera exhaustiva el caudal léxico de una lengua. En el caso del español, es el *Diccionario de la Real Academia Española*, conocido por sus siglas *DRAE*, la referencia obligatoria del patrimonio léxico para todos los hablantes de español. Su carácter de “depósito” léxico de la lengua española hace que la vigésima segunda, de 2003, cuente con unas 88.500 entradas o lemas. Con respecto a la vigésima primera

edición, de 1992, de estos lemas, unos 11.500 son novedades léxicas, entre ellas *zapear*, *liposucción* o *videojuego*, y desaparecen, porque realmente han muerto, otras seis mil palabras. Se añaden, al mismo tiempo, para un mejor reflejo de la realidad lingüística universal, numerosas marcas de americanismos. Se sobreentiende, por tanto, que en los once años que van de una edición a la otra, las nuevas palabras y acepciones ahora incorporadas ya eran antes usadas, al tiempo que muchas de las que se han eliminado habían caído en el olvido desde hacía varios años.

Ni qué decir tiene que “ningún hablante conoce todos los vocablos de un diccionario, ni siquiera todas las acepciones de los considerados como más frecuentes o con mayor rendimiento funcional” (Najt y Reizábal, 1989: 220). Al respecto, tomando al azar una página del *DRAE*, Humberto López Morales (1986: 53) ha clasificado el vocabulario que integra un *diccionario normativo*, según el grado de conocimiento y uso que un hablante hace de las palabras, en las siguientes clases:

a) palabras que nos son totalmente desconocidas, ya que ni las hemos leído ni escuchado en ninguna parte, como *empulgadura*, *empuntar*, *empurrarse*, *emuntorio*, etc.;

b) palabras que nos parecen conocidas, pero que en realidad no sabemos bien lo que significan, como *emulgente*, *emulsivo*, *emulsor*, *enacerar*, etc.;

c) palabras que sí conocemos y entendemos, pero que no utilizamos –léxico pasivo– al hablar o escribir, como *emulación*, *emulador*, *enajenable*, *enanismo*, etc.;

d) palabras que en realidad conocemos, no solo porque las oímos o leemos con frecuencia, sino porque nosotros mismos las usamos –léxico activo–, como *empuñadura*, *enamorado*, *enamoradoizo*, *enamorar*, *enano*, etc.

A continuación, el mismo autor hace un listado de los vocablos que se apartan del léxico general al que, en su afán de exhaustividad, los diccionarios dan cabida a. Entre estos vocablos se hallan:

– arcaísmos, palabras anticuadas que se usaron en otra época pero que ya han caído en el olvido, como *empulgar* (‘armar la ballesta’), *enaciyar* (tratar las lanas con aceite, acije o aceite de vitriolo’);

– regionalismos, palabras que solo se conocen en alguna zona o país del mundo hispánico, como *empuntar* (‘encarrilar, encaminar, dirigir’, en Colombia, Ecuador y El Salvador), *engrudo* (‘cola de pegar’, en Canarias);

– tecnicismos, palabras científicas manejadas solo o principalmente por especialistas, como *emunción* (‘evacuación de los humores y materias superfluas y nocivas’), *enálage* (‘figura que consiste en mudar las partes de la oración o sus accidentes; como cuando se pone un tiempo del verbo por otro, etc.’);

– cultismos, palabras empleadas muy ocasionalmente por hablantes de gran cultura y de amplio dominio léxico, como *enajenar* (‘pasar o transmitir a otro el

dominio de una cosa o algún otro derecho sobre ella’), *enalbar* (‘caldear y encender el hierro en la fragua tanto, que parezca blanco de puro resplandeciente’).

El listado anterior muestra que los diccionarios normativos constituyen un buen depósito de todo –o casi todo– el caudal léxico de una lengua. Sin embargo, –concluye López Morales (1986: 54) no son adecuados pedagógicamente<sup>1</sup> para delimitar y enseñar las palabras “que tienen realmente vida común entre los hablantes de una comunidad”.

### 1.1.2. La frecuencia léxica

La exhaustividad de los diccionarios normativos, que recogen todo el caudal léxico de una lengua se vio depurada por las listas de frecuencia, compuestas por los vocablos que se repiten con más frecuencia en los textos de una comunidad. La *frecuencia léxica*, se basa en el índice ‘frecuencia’, que indica el número de apariciones de los vocablos (o palabras diferentes) en los textos computados. Estos vocablos se ordenan según rangos, es decir, estableciendo una jerarquía de acuerdo a la cantidad de veces que se repiten en un texto. Si se toman muchos textos, orales y escritos, representativos de la lengua de un país o comunidad, se llegaría a tener una lista de las palabras *realmente* usadas allí, jerarquizada por rangos de aparición, en definitiva, una lista de frecuencia.

“Las palabras más frecuentes son, principalmente, las que sirven para mantener la estructura del mensaje, mientras que las voces realmente portadoras de información se repiten relativamente poco en el discurso” (Alvar Ezquerro, 2005: 22). Tanto es así que, según Alvar Ezquerro (2003: 100) “en español, las 100 palabras más frecuentes representan más del 30% del material léxico que constituye cualquier texto; con las 1000 más frecuentes se llega a más del 50% de todo texto; con las primeras 5000 se alcanza más del 90%”. Las palabras más frecuentes son de tipo gramatical (preposiciones, determinantes, pronombres, conjunciones, adverbios), además de los nombres de los objetos inmediatos y los verbos y adjetivos elementales. Ello se puede apreciar en la lista de las 160 palabras más frecuentes (véase Cuadro 1.1) del *Diccionario de frecuencias de las unidades lingüísticas del castellano* de Alameda y Cuertos (1995), que maneja un corpus de textos de 2.000.000 de palabras del periodo 1980–1995:

---

<sup>1</sup> Un buen diccionario para la enseñanza de L2 debe contener, según Alvar Ezquerro (2003: 104–105), información sobre: a) división silábica, b) pronunciación, c) léxico usual en la lengua escrita y hablada, d) definiciones claras, breves y no sinónimas, e) abundantes ejemplos de uso, f) uso gramatical y estilístico, g) familias de palabras, h) ilustraciones contextualizadoras, i) pesos, medidas, nombres geográficos y otros elementos culturales.

**Cuadro 1.1. Las 160 palabras más frecuentes del español, según Alameda y Cuetos (1995)**

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