

RECENSIONES

RAMÓN ALMELA, PASCUAL CANTOS, AQUILINO SÁNCHEZ, RAMÓN SARMIENTO y MOISÉS ALMELA: *Frecuencias del español. Diccionario y estudios léxicos y morfológicos*, Madrid: Editorial Universitas, 2005, 591 pp., ISBN: 84-7991-171-9.

El retraso que las ciencias antropológicas han experimentado tradicionalmente respecto de las ciencias cosmológicas se observa también en el ámbito de lo que podríamos llamar los métodos cuantitativos. El déficit de este tratamiento es particularmente evidente en las ciencias del lenguaje. Durante siglos se ha venido construyendo un edificio (inacabado e inacabable) de afinadas y más o menos atinadas teorías acerca de los aspectos invariantes del sistema de la lengua y de algunas de sus realizaciones de discurso. Esto es y será siempre necesario. Sin embargo, estas consideraciones teóricas no suelen venir acompañadas del correspondiente estudio cuantitativo, que iluminen el uso de las palabras y hasta su destino en el sistema de la lengua. No debe olvidarse que la lengua es un organismo vivo, un organismo que cambia a medida que se emplea en la realidad concreta del hablar. Afortunadamente, estas deficiencias cuantitativas han empezado a corregirse en los últimos tiempos.

El libro que reseñamos, que ofrece la frecuencia con que los hispanohablantes usamos actualmente las palabras y los afijos de nuestra lengua, viene, pues, a llenar un hueco más que evidente en nuestra bibliografía lingüística. Hasta ahora, el español solamente contaba con el diccionario de frecuencias de Alphons Juillard y E. Chang Rodríguez. Redactado en inglés y publicado en 1964, fuera de España, sus resultados se basaron en un corpus de sólo 500.000 palabras, procedentes de fuentes no diversifica-

das. *Frecuencias del español*, por el contrario, está basado en un corpus de 20 millones de palabras, recogidas en sus contextos orales y escritos actuales de España e Hispanoamérica. Se trata del corpus *Cumbre*, que, por el volumen de textos y el número de palabras que contiene y por sus procedencias, cuidadosamente seleccionadas, puede ser considerado razonablemente el corpus de material léxico más representativo del español actual.

Nuestro libro consta de una sección explicativa y de una sección de anexos. En la sección explicativa, se incluyen una introducción, seis capítulos, un glosario y la bibliografía. En la sección de anexos, se presentan los datos de frecuencia extraídos del Corpus.

La «Introducción» presenta el diseño del corpus *Cumbre*, el método seguido en el trabajo, así como los argumentos que avalan una investigación de estas características. No les falta razón a los autores cuando afirman que en el estudio del lenguaje no se puede seguir el mismo método que se sigue en el estudio de la filosofía o en el análisis de la reacción subjetiva ante un fenómeno artístico. El lenguaje —dicen— «en un porcentaje altísimo, se fundamenta en elementos formales y claramente objetivables, cuales son, por ejemplo, los sonidos, las palabras o las formas en general y su ordenación en el discurso. En este sentido, no habría razón para no adoptar, en los estudios sobre el lenguaje, métodos de investigación menos especulativos y más dependientes de los datos que la realidad lingüística nos proporciona». La imagen intuitiva que tenemos de la estructura y de la frecuencia del léxico de nuestra lengua dista mucho a veces de responder a la realidad. Los datos revelan que la mayor parte de las palabras tienen una frecuen-





cia baja o muy baja, mientras que son relativamente pocas las que usan los hablantes con frecuencia muy alta.

Conviene distinguir entre el uso de las palabras como 'formas flexivas' y el uso de las palabras como 'lemas'. El «Capítulo 1» recoge el uso de las palabras consideradas como formas individuales, no integradas en la categoría del lema que les pudiera corresponder. Es interesante conocer este dato, porque nos permitirá comprobar que no todas las formas de un lema son objeto de la misma intensidad de uso. Cabe constatar, por ejemplo, que si bien el verbo *tener* es frecuentemente usado, la forma *tiene* es la más usada de todas sus flexiones. De igual manera, la forma más frecuente del verbo *ser* es la tercera persona singular del presente de indicativo (*es*), como lo es *está* en relación con el verbo *estar*. A un profesor de español como lengua extranjera le debería interesar este dato para convencerse —si no lo estuviera ya— de que vale la pena aprender pronto una forma irregular como *tiene* o *es*. Y a un lingüista le podría sugerir alguna idea útil sobre las razones del desdoblamiento vocálico de la *e* española: tal desdoblamiento aparece en las formas más frecuentes y usadas del verbo *tener* (*tiene*, *tienen*, *tienes*...). De no menor interés es constatar que las 100 primeras formas lingüísticas ofrecen un gran contraste de frecuencia: de más de 1.223.000 (*de*) a 14.965 (*poco*). Estas distancias en la frecuencia son notoriamente menores en el resto de la lista.

El «Capítulo 2» aborda el análisis de los lemas. La asignación de cada forma al lema que le corresponde requiere una revisión manual para ser fiable. De ahí que el listado de los lemas se haya elaborado a partir de un subcorpus de *Cumbre*, de 2 millones de palabras etiquetadas y revisadas. Dicho subcorpus se recopiló con criterios similares al corpus general, precisamente para garantizar una razonable representatividad del español actual. Del análisis de los resultados se deduce que el número de lemas en español es aproximadamente la mitad que el número de formas (de ahí que este listado de 5.000 lemas equivalga en realidad al listado anterior, de 10.000 formas), que el lema más frecuente es el artículo *el* (y no la preposición *de*, primera en la lista de formas), y que las diferencias de frecuen-

cia entre los veinte lemas más frecuentes son más acusadas que las que se dan entre los lemas que ocupan las posiciones sucesivas.

Con el fin de dar cuenta de la realidad a caballo entre la forma directamente observable y los aspectos más abstractos de la funcionalidad y del potencial comunicativo, el libro incluye, en el «Capítulo 3», un análisis de la frecuencia de grupos de formas categorizadas en función de criterios gramaticales y léxico-semánticos. La aplicación del primer criterio se concreta en la preparación de un listado con los 5.000 lemas más frecuentes del español, ordenados por frecuencia («Anexo III»). Los datos son extraídos del mismo subcorpus, de 2 millones de palabras. Junto a los datos descriptivos ya conocidos (*frecuencia Corpus* y *frecuencia relativa millón*), se analizan también la *banda de frecuencia* y la *categoría gramatical* (adjetivo, adverbio, conjunción, artículo, interjección, preposición, pronombre, sustantivo y verbo). En la medida en que estos 5.000 lemas puedan ser considerados como representativos del español actual, estos cómputos reflejan que la categoría gramatical más frecuente es el sustantivo, con casi el 54% sobre el total de los lemas, seguido a una distancia considerable por el verbo y el adjetivo, 22% y 18%, respectivamente. Se aprecia claramente que los lemas léxicos (sustantivos, verbos, adjetivos, adverbios e interjecciones) destacan sobre los lemas funcionales (conjunciones, artículos, preposiciones y pronombres): 96,92% frente a 3,08%.

El capítulo se completa con una abstracción de tipo léxico-semántico: los distintos usos concretos de cada uno de los lemas de un grupo escogido se agrupan en acepciones, y éstas se someten a un análisis cuantitativo. Además, para algunos lemas se incluye un desglose de la frecuencia de uso de cada acepción por formas del mismo lema. El estudio parte de los datos del corpus *Cumbre* y toma como referencia las entradas lexicográficas del *Gran Diccionario de Uso del Español Actual*, que está basado en el mismo corpus y fue dirigido por uno de los autores del libro. Las conclusiones apuntan a que el potencial semántico de una palabra no es indiferente a la frecuencia de uso de sus acepciones, y a que éstas no tienen por qué relacionarse por igual con todas las formas flexivas del lema.

El «Capítulo 4» ofrece el primer diccionario de frecuencias de los afijos del español. Tras la justificación teórica y la detallada explicación del método seguido, los autores muestran la frecuencia de los sufijos y los prefijos del español por orden alfabético, así como su frecuencia por orden decreciente de las formas y de los lemas que les sirven de base; por ejemplo: el sufijo *-ión* es el primero por cantidad de formas que lo portan y *-al*, el segundo; pero, por cantidad de lemas que los portan, *-ión* es el segundo y *-a*, el quinto; en cuanto a los prefijos, *in-* afecta a mayor número de formas que *des-*, pero *des-* afecta a más lemas que *in-*. Nos ofrece después una sencilla descripción lingüística de 10 sufijos elegidos entre los 20 más frecuentes y de los 10 prefijos más frecuentes de nuestra lengua. De cada uno de esos 20 afijos se han seleccionado los 30 lemas más frecuentes, con lo cual se alcanza la suma de 600 lexemas: *situación, director, social, importante, verdadero, disponer, información, encargar, transformar, explicar...* Tras haber examinado los contextos del corpus *Cumbre* correspondientes a esos 600 eductos, se extraen los sentidos (o rasgos designativos generales) asignados a cada uno de los 20 afijos elegidos —10 sufijos y 10 prefijos—; una muestra limitada de tales contextos acompaña a la explicación de tales sentidos.

El «Capítulo 5» resume los datos referidos a las 1.000 combinaciones de palabras más frecuentes del español (500 de dos y las otras 500 de tres). Los datos estadísticos de las 500 combinaciones de dos palabras más frecuentes del español suman una frecuencia total de casi 3.120.000. Importa destacar que todas estas combinaciones están catalogadas como muy frecuentes. Se trata, pues, de estructuras sintagmáticas muy utilizadas por los hablantes del español (p. ej.: *de la, en el, de los, en la, etc.*), muchas de ellas lexicalizadas (p. ej.: *creo que, hay que, después de, sin embargo, a ver, etc.*). En cuanto a las combinaciones de tres palabras, los datos descriptivos arrojan notables diferencias con respecto a sus homólogas de dos palabras. Se aprecia un descenso muy notable en las frecuencias con respecto a las combinaciones menos usadas de dos palabras. Las combinaciones de tres palabras más usadas por los hablantes tienen casi

la misma frecuencia que las menos frecuentes de dos palabras. Entre las combinaciones de tres palabras más frecuentes de nuestra lengua figuran las siguientes: preposición + artículo + sustantivo (*de la vida, en el caso, por otra parte, etc.*) (17%), artículo + sustantivo + preposición (*el caso de, una serie de, la mayoría de, etc.*) (11%) y preposición + artículo + pronombre (*en el que, a la que, etc.*) (7%).

El «Capítulo 6» presenta un modelo de análisis de las co-ocurrencias y los colocados de una palabra determinada. El objetivo de los autores es ofrecer un método de trabajo y sugerir nuevos y más eficientes procedimientos de análisis. Para ello se analizan y contrastan las co-ocurrencias y los colocados de los sustantivos *mujer* y *hombre*.

Tras el «glosario de términos estadísticos» y una «bibliografía» elemental, siguen más de 400 páginas con los anexos correspondientes. La reseña de los anexos es innecesaria; pero sí destacaremos algunos aspectos significativos. Los «anexos 1 y 2» incluyen datos específicos referidos a la frecuencia absoluta, frecuencia relativa por millón y banda de frecuencia de cada una de las formas y lemas. El «anexo 3» incluye además la categoría gramatical referida a cada lema. Los «anexos 4 y 5», centrados en las combinaciones de dos y tres palabras más frecuentes, contienen los mismos tipos de datos que los incluidos en los anexos 1 y 2, pero aplicados a las combinaciones de palabras. Finalmente, los distintos apartados del «anexo 6» ofrecen mapas de las co-ocurrencias y listados de colocaciones de las voces *mujer* y *hombre*.

La relativa novedad de este tipo de estudios explica que todavía no se hayan puesto de manifiesto los beneficios que los mismos pueden aportar en áreas como la lexicografía, la lexicología, la morfología en general, la sintaxis o, más específicamente, la enseñanza del español como lengua no materna. Disponer de una obra de estas características ayudará a calibrar, con datos objetivos, el peso de una palabra, de una combinación, de un afijo... Es posible constatar que algunos sentidos que aparecen en muchos diccionarios apenas tienen relevancia para los hablantes, o que determinados adjetivos prefieren la posición prenuclear o posnuclear dependiendo del tipo de sustantivo que los rija. Por ejemplo, la combina-

ción *persona alta* es mucho más frecuente que *alta persona*, pero *alta personalidad* lo es mucho más que *personalidad alta*. Incluso quienes buscan, por razones estilísticas, ayuda para no caer en la trivialidad del discurso repetido encontrarán aquí las palabras o combinaciones de palabras más infrecuentes del español.

Por todo ello, creo que hay que dar la bienvenida a este libro, que ofrece tantas buenas ra-

zones cuantitativas para conocer mejor nuestra lengua. Es cierto que lo que nos proporciona *Frecuencias del español* son solamente datos acerca de la norma (= uso consolidado) de nuestra lengua; pero no olvidemos que la norma es la cara visible del sistema.

Marcial MORERA
Universidad de La Laguna

