## II. LINGUISTICA

# LOS DICCIONARIOS DE FRECUENCIA EN ESPAÑOL (II)

*Raimundo Ezquerra*
*Universidad de París – X*

Desde que en marzo de 1974[1] publicamos nuestro primer trabajo de síntesis sobre las encuestas léxicas, hasta entonces publicadas, han aparecido otras tres nuevas en España:

– MARQUEZ VILLEGAS L.–*Vocabulario del español hablado.* Madrid, Sociedad General Española de Librería. 1975.

– GARCIA HOZ V.–*El Vocabulario General de Orientación Científica y sus estratos.* Madrid, C.S.I.C. 1976.

– DIAZ CASTAÑON C.–*Vocabulario Básico del Español y sus aplicaciones a la enseñanza.* Oviedo, I.C.E. de la Universidad de Oviedo. 1977.

Recientemente hemos estudiado las dos primeras obras citadas en sendos artículos ya publicados[2]; no así la tercera.

Nos pareció útil, al mismo tiempo que estudiábamos el *Vocabulario Básico,* hacer un balance general de nuestro trabajo sobre este tema y actualizar los resultados de conjunto elaborados en 1974.

Este artículo constará pues de dos partes bien diferenciadas; en la primera estudiaremos la reciente publicación del I.C.E. de Oviedo y en la segunda reuniremos todos los datos dispersos en los tres artículos anteriores añadiendo los aportados aquí.

## A.–EL VOCABULARIO BASICO DEL ESPAÑOL Y SUS APLICACIONES A LA ENSEÑANZA

### 1.–CARACTERISTICAS GENERALES

La primera observación que nos parece necesaria es que el título del trabajo de Carmen DIAZ CASTAÑON y su amplio equipo de colaboradores parece inexacto. En efecto, como más adelante demostraremos, la lista resultante de esta encuesta no es un vocabulario básico del español ni podía serlo, dadas las características técnicas y las premisas generales de la encuesta. No queremos decir que el resultado de esta obra carezca de utilidad; al contrario



BOLETÍN AEPE Nº 16. Raimundo EZQUERRA. LOS DICCIONARIOS DE FRECUENCIA EN ESPAÑOL (II)
Centro Virtual Cervantes

es un utilísimo instrumento para ulteriores estudios de sociolingüística y probablemente también para orientaciones pedagógicas, pero no es lo que se suele denominar un "vocabulario básico".

El *Vocabulario Básico* constituye, en realidad, una encuesta sobre el léxico, principalmente latente[3] y principalmente conceptual sustantivo, de los alumnos españoles de enseñanza media.

Una de las características más interesantes de la obra que estudiamos es el haber comunicado por separado las frecuencias y repartición de los vocablos resultantes para cada grupo de edades. Los grupos considerados son los de los alumnos de 8 y 9 años, de 10 y 11 años, de 11 y 12 años, de 12 a 14 años, de 15 y 16 años y de 17 años en adelante, con lo que podemos estudiar la correlación entre ciertos campos léxicos o semánticos y las diferentes edades. Además, otras variables como el sexo, el carácter urbano o rural de la zona, el carácter oficial o privado del centro de enseñanza, etc., aunque no publicadas fueron también registradas, con lo que, como más arriba indicamos, queda manifiesto el interés de esta encuesta desde un punto de vista sociolingüístico y, en ciertos aspectos, psicolingüístico.

La parte hasta ahora publicada del trabajo nos proporciona, en un primer volumen, la lista de todos los vocablos obtenidos por medio de "encuesta libre", más una lista, por orden de frecuencias decrecientes, de los vocablos más frecuentes.

El segundo tomo recoge una serie de listas elaboradas a partir de los "centros de interés".

Por último, señalaremos que hemos podido comprobar, con agrado, cómo ciertas recomendaciones y ciertas pautas preconizadas por nosotros en nuestro trabajo de 1974 han sido tenidas en cuenta por los autores del *Vocabulario Básico.* Este es el caso, por ejemplo, en la distinción metodológica entre "palabra" como unidad de texto y "vocablo" como unidad de léxico, que nosotros establecíamos, basándonos en la teoría de la estadística léxica de Ch. MULLER. También podemos comprobar como, en lo referente a la definición de la norma se han tenido en cuenta nuestras observaciones generales sobre el problema, recogiendo incluso en las primeras líneas del párrafo dedicado a este tema[4] exactamente las ideas expuestas por nosotros[5]. De manera general, muchos de los datos comunicados en nuestro primer trabajo de síntesis han servido de referencia al equipo del I.C.E. de Oviedo.

## 2. OBJETIVOS DEL VOCABULARIO BASICO

En la página 11 (párrafo 1.1) del primer tomo de la obra que comentamos, podemos leer: "Nuestro trabajo no pretende señalar un mínimo vocabulario básico para el aprendizaje de una lengua extranjera; es un intento de poner al alcance de los docentes una información sobre el léxico real que poseen nuestros alumnos...". Añadiríamos nosotros, como dijimos más arriba, el léxico principalmente latente, conceptual, principalmente sustantivo. No es obvio que se trate del léxico "real", ya que el tipo de encuesta efectuada favorece, por definición, la aparición del léxico latente que, siendo evidentemente "real" no es quizás exactamente el mismo que otro tipo de encuesta, más apta a hacer aparecer el léxico activo, hubiera puesto de manifiesto, y que no hubiera sido menos "real".

Salvo esta observación teórica y conceptual sobre la palabra "real", que no nos parece suficientemente explícita en este contexto, estamos de acuerdo con la cita precedente. En



BOLETÍN AEPE Nº 16. Raimundo EZQUERRA. LOS DICCIONARIOS DE FRECUENCIA EN ESPAÑOL (II)
Centro Virtual Cervantes

efecto, esto es lo que, a nuestro juicio, se puede considerar como objetivo de una encuesta de este tipo y, lo que, en todo caso, da como resultado. No entendemos pues por qué se dio al trabajo el título general de *Vocabulario Básico del Español y sus Aplicaciones a la Enseñanza* que induce a error.

La segunda parte del título: "y sus aplicaciones a la enseñanza", merece también algún comentario. En efecto, el párrafo anteriormente citado explica que "Nuestro deseo es que este vocabulario pueda servir y ser aprovechado por profesores, por autores de libros de texto y por cuantos deseen cooperar al enriquecimiento del vocabulario activo del alumno, obligándole a utilizar y a ser consciente de su vocabulario pasivo".

Si este deseo nos parece digno de interés, no creemos que la aplicación a la enseñanza de los resultados de la encuesta puedan ser tan directos como los autores parecen sugerirlos. El pedagogo puede tener dos actitudes frente a los datos suministrados por el *Vocabulario Fundamental.* La primera y más simple consistiría en basar sus textos en dicha lista léxica y hacer así pasar muchos vocablos del léxico latente al léxico activo. Pero esto no nos parece suficiente, pues desde un punto de vista conceptual nada o poco aportaríamos a los alumnos que ya poseen los conceptos asociados a dichos vocablos, más o menos bien, es cierto. Nosotros pensamos que tras esta primera fase, muy útil, la verdadera importancia del recuento está en indicarnos cuál es el mundo conceptual de los alumnos de cada grupo de edades, y en función de este conocimiento, decidir, si éste es suficiente, coherente y apropiado a la edad. Por lo tanto, la aplicación de los resultados de la encuesta puede ser, a veces, la utilización de vocablos que, precisamente, *no están* en ella.

La utilización del *Vocabulario Básico* en la pedagogía puede ser fundamental, pero, a nuestro parecer, indirecta, pasando antes por otros campos científicos como, por ejemplo, el de la sociolingüística o estudios generales como la formación en el niño del mundo de los conceptos. Discrepando, pues, en parte, con los autores, no señalaremos en los cuadros finales objetivos más bien pedagógicos.

## 3. EL CORPUS

### 3.1. Importancia cuantitativa

El corpus del *Vocabulario Básico* está constituido por 850.000 palabras, 170.000 de las cuales provienen de la "encuesta libre" y el resto de la "controlada", es decir, obtenidas por el método de los "centros de interés".

Los autores afirman[6]: "Hemos tenido muy en cuenta el corpus utilizado por los diferentes vocabularios de uso del español", pero sin explicar de qué manera.

En todo caso, esta extensión del corpus relaciona al *Vocabulario Básico* con los que nosotros hemos denominado en otros artículos de la "segunda generación", es decir, con un corpus no demasiado extenso. Sin embargo, el *Vocabulario,* al igual que *El Español Fundamental* no pertenece, desde un punto de vista matemático al grupo de las encuestas más elaboradas.



BOLETÍN AEPE Nº 16. Raimundo EZQUERRA. LOS DICCIONARIOS DE FRECUENCIA EN ESPAÑOL (II)
Centro Virtual Cervantes

### 3.2. Características del corpus

Aquí es donde se sitúa la más importante especificidad del diccionario que estudiamos y de ello dependen las más importantes características de sus resultados.

Estas 850.000 palabras fueron obtenidas de dos maneras:

a) Por "encuesta libre"; entendiendo bajo este concepto: "Escribir durante cinco minutos todos los términos que libre y espontáneamente acudan a la mente"[7].

b) Por "encuesta controlada"; "tomando como tema (Centro de Interés) un campo familiar al alumno, escribir los 20 primeros términos que acudan a la mente"[8].

Esta metodología ha sido empleada ya repetidas veces en la historia de las encuestas léxicas. Dentro del campo del español fue, por ejemplo, aplicada, al principio de los años cincuenta por RODRIGUEZ BOU[9] bajo el nombre de "ejercicios de asociación libre y controlada", oponiendo en bloque a las palabras así obtenidas al por él llamado "vocabulario oral", proviniente de palabras realmente pronunciadas por los alumnos en conversaciones libres, y apuntadas por sus ayudantes.

Más tarde, RIVENC y ROJO SASTRE, basándose en las técnicas del "Français Fondamental", utilizarán el método de los "Centros de Interés" para obtener 400.000 palabras que añadir a las 800.000 grabadas en la vida real. Estas 400.000 palabras constituirán el llamado "vocabulario disponible", caracterizado por mayor riqueza semántica que el "oral" e indispensable para equilibrar el corpus.

Los autores del "Francés Fundamental" explican muy bien el por qué de la necesidad de recurrir a estos métodos no extrictamente estadísticos[10].

En efecto, una estadística basada en la lengua realmente hablada da como resultados:[11]

1.º Ante todo, palabras gramaticales.

2.º Verbos.

3.º Adjetivos.

4.º Algunos nombres de carácter general (divisiones del tiempo, palabras como: cosa, hombre, persona, niño, etc.).

Esto es totalmente lógico, pues desde hace tiempo saben los lingüistas que la lengua hablada se compone en gran parte de palabras gramaticales, verbos, etc., ya que la omnipresencia del "referente" extralingüístico hace supérflua su denominación.

Si queremos obtener listas ricas en sustantivos, por ejemplo, tendremos pues que recurrir, o a textos escritos, o a fórmulas del tipo de los "centros de interés" o de las "asociaciones libres y controladas".

Como corolario, si utilizamos estos procedimientos no debemos esperar gran cosa fuera de la categoría de los sustantivos. Nos extraña pues la afirmación de los autores del *Vocabulario Básico:* "No ha habido problema con las llamadas palabras gramaticales porque no han sido elegidas por los alumnos"[12]. En efecto, en una encuesta basada únicamente en las asociaciones y en los centros de interés, como la presente, los sujetos *no pueden* "elegir" prácticamente nada que no sean sustantivos.



BOLETÍN AEPE Nº 16. Raimundo EZQUERRA. LOS DICCIONARIOS DE FRECUENCIA EN ESPAÑOL (II)
Centro Virtual Cervantes

Como segundo corolario, es obvio que el *Vocabulario Básico* así obtenido no puede, de ninguna manera, ser un "vocabulario básico", si por esto entendemos el conjunto de vocablos más frecuente en la lengua realmente utilizada por los sujetos.

Una encuesta de este tipo sólo nos puede proporcionar el vocabulario disponible, según su definición dada en 1956 por los autores del "Francés Fundamental".

Al estudiar los resultados de la encuesta comprobaremos que la realidad confirma plenamente esta visión teórica.

Como conclusión, hemos pues de apuntar, desde ahora, que el corpus del *Vocabulario Básico* es un corpus "disponible" y no contiene lo que se suele denominar un "corpus frecuente". La utilización de la noción de frecuencia para clasificar los vocablos obtenidos es, en efecto, independiente de lo primero. Por otra parte, tampoco podemos decir que el corpus del *Vocabulario Básico* tenga un origen escrito, ni mucho menos literario, por lo que lo clasificaremos como de origen "oral", con las reservas impuestas por los razonamientos precedentes.

Aunque más tarde estudiaremos la lista obtenida, podemos ya señalar que la forma en que se constituyó el corpus de la obra que estudiamos ha tenido grandes influencias en sus resultados. Valga como ejemplo citar que el vocablo "encuesta" se sitúa en el rango 372 de la lista de frecuencias decrecientes de la "encuesta libre". Apareció, dicho vocablo, 3 veces en el corpus de alumnos de 8 y 9 años, 13 veces en el de los alumnos de 10 y 11 años, 31 veces en el de los alumnos de 12 a 14 años, etc. Este solo ejemplo hace patente la influencia de las condiciones, un poco artificiales, en que se desarrolló la encuesta, sobre los resultados. No creemos que el vocablo "encuesta" se sitúa entre los 372 vocablos más utilizados por los alumnos de enseñanza media en la lengua "real".

## 4. LA TECNICA DEL RECUENTO

### 4.1. Criterios lingüísticos

La norma lingüística adoptada nos parece aceptable y tiene el mérito de prevenir con claridad los problemas que se hubieran podido plantear a los utilizadores del corpus. Citemos, a título de ejemplo, que los sinónimos han sido considerados como entradas diferentes, cosa indispensable en una encuesta léxica no únicamente formal, que se aceptaron los neologismos aunque no se encuentren en el Diccionario de la Real Academia Española, siempre y cuando se trate de palabras cuya derivación y composición se ajuste a las reglas generales del idioma, etc. Es también útil el haber contabilizado los vocablos en bable, pues esto puede servir para ulteriores estudios dialectológicos. Señalan los autores que indicaron siempre la categoría gramatical de un vocablo cuando ésta pudiera ofrecer dudas. Esto no es totalmente cierto, pues nos hemos encontrado algunas veces con vocablos que no podemos saber si son sustantivos o adjetivos sustantivos, como, por ejemplo, "tonto" (número 115) (¿se trata de un niño que es tonto o del tonto de la clase, del tonto del circo? ), "rosa" (número 131), etc.

Aparte de estos pequeños defectos, podemos considerar que la norma está claramente definida.



BOLETÍN AEPE Nº 16. Raimundo EZQUERRA. LOS DICCIONARIOS DE FRECUENCIA EN ESPAÑOL (II)
Centro Virtual Cervantes

### 4.2. Criterios matemáticos

En este terreno, nada nuevo aporta el *Vocabulario Básico.* Al contrario, su utilización simple de las nociones de "frecuencia" y de "repartición" lo sitúan más bien entre los primeros diccionarios de frecuencia del español.

La lista general de 1.352 vocablos denominada "Vocabulario Básico" utiliza la frecuencia decreciente como único índice de clasificación.

El umbral de admisión en esta lista[13] es de una frecuencia igual o superior a 20, lo que selecciona 1.352 vocablos entre los 9.782 obtenidos en la encuesta, a partir de las 850.000 palabras del corpus.

En cada caso se nos indica la repartición, dato de interés pero que hubiera sido más útil tratar de integrar en un índice más complejo de clasificación.

Los 3.150 sujetos de la encuesta fueron repartidos en 75 grupos, 15 en cada franja de edades, lo que hace que el factor R pueda variar entre 1 y 75 en la lista, aunque en realidad, para una frecuencia de 20[14] como mínimo, resultó una repartición de 6 como mínimo.

## 5. OBSERVACION DE LOS RESULTADOS

Sólo estudiaremos con algún detalle la lista por orden de frecuencia decreciente obtenida a partir del corpus de "asociaciones libres", pues las listas parciales, también por frecuencias decrecientes, a partir de los "centros de interés" no pueden sino confirmar y acentuar las grandes tendencias de la "asociación libre".

### 5.1. Influencia de las condiciones de la encuesta

Lo primero que llama la atención al examinar la lista es la considerable influencia que han tenido las condiciones y circunstancias de la encuesta en sus resultados.

En efecto, ya señalábamos más arriba el caso del vocablo "encuesta", sumamente ejemplar a nuestros ojos. Además, el entorno material de la encuesta (aula) ha condicionado, a nuestro parecer, en gran parte a ésta.

Entre los 25 primeros vocablos de la lista (por orden de frecuencia decreciente, recordémoslo), encontramos:

—mesa (número 1).
—libro (número 2).
—niño (número 3).
—silla (número 4).
—bolígrafo (número 5).
—lápiz (+ lapicero) (número 6).
—libreta (número 11).
—armario (número 12).
—encerado (número 15).
—ventana (número 16).
—estuche (número 17) (seguramente de lapiceros).



BOLETÍN AEPE Nº 16. Raimundo EZQUERRA. LOS DICCIONARIOS DE FRECUENCIA EN ESPAÑOL (II)
Centro Virtual Cervantes

—goma (número 18) (seguramente de borrar).
—puerta (número 22).
—gafas (número 23) (¿del profesor? ¿del compañero? ).
—papel (número 24).
—jersey (número 25).

Naturalmente, hay vocablos, como "silla", "armario", "ventana" y "puerta" que no son específicos de una situación escolar y que, de todas maneras hubieran estado bien clasificados en una encuesta a partir de asociaciones libres y controladas, pero la presencia de los demás de la lista proporciona sospechas razonables de que se trate de partes del aula en donde se encuentran los alumnos. Ante otros vocablos, como "jersey", no podemos dejar de pensar que esta prenda es la más utilizada por los niños de cualquier colegio de España.

En todo caso, se puede afirmar que más de la mitad de los vocablos catalogados como los 25 más frecuentes han sido sugeridos por el entorno inmediato de los sujetos.

Este índice, si no quita valor al conjunto de la lista y del léxico extraído del corpus, sí anula en gran parte la importancia de una posible clasificación de los vocablos en función de su frecuencia y repartición.

### 5.2. Aspecto "latente" del léxico obtenido

El campo léxico[15] de los animales es particularmente interesante en cuanto al carácter de "disponible" o "latente" del léxico obtenido. En efecto, junto a muy pocos animales domésticos o familiares de nuestra civilización, encontramos en la lista:

—tigre (número 108).
—elefante (número 133).
—pantera (número 169).
—oso (número 172).
—jirafa (número 309).
—leopardo (número 373).
—ballena (número 420).
—rinoceronte (número 456).
—cocodrilo (número 513).
—hipopótamo (número 572).
—avestruz (número 683).
—cebra (número 715).
—canguro (número 1.045).
—gorila (número 1.167).
—jaguar (número 1.204).

Todos estos animales pertenecen, obviamente, al mundo lúdico de los niños, al mundo fabuloso de las novelas, las películas de aventuras, etc., pero poco pueden formar parte de las conversaciones en la vida real de Asturias. Constituyen un buen ejemplo de vocabulario "latente" o "pasivo", es decir, conjunto de vocablos que el sujeto conoce y entiende pero utiliza poco. Este hecho ya había sido puesto de relieve por la encuesta del "Francés Fundamental" en su corpus proviniente de los "Centros de Interés"[16].



BOLETÍN AEPE Nº 16. Raimundo EZQUERRA. LOS DICCIONARIOS DE FRECUENCIA EN ESPAÑOL (II)
Centro Virtual Cervantes

### 5.3. Estructura del léxico obtenido

#### *5.3.1. Las "palabras gramaticales"*

Si examinamos el porcentaje de "palabras gramaticales obtenidas por la lista de *El Español Fundamental (Lista Alfabética de Formas Empleadas),* del *Frequency Dictionary of Spanish Words* y del *Vocabulario Básico* que ahora estudiamos, vemos que, en porcentaje[17] constituyen, con relación a las respectivas listas:

*Lista Alfabética:* 6,8 por ciento.
*Frequency Dictionary:* 4,9 por ciento.
*Vocabulario Fundamental:* menos del 2 por ciento.

Este dato es ya, en sí, significativo del aspecto "disponible" de nuestra lista, pero más todavía lo es que la primera palabra gramatical que aparece en ella (el pronombre "yo") ocupe el puesto número 517.

En una lista de frecuencia de cualquier lengua, siempre las primeras decenas de vocablos son palabras gramaticales más algún verbo. Por ejemplo, en el "Francés Fundamental" los 42 vocablos más frecuentes, y los primeros de la lista, son todos "palabras gramaticales" menos un adverbio (bien) y cinco verbos (être, avoir, faire, dire y aller).

#### *5.3.2. Los verbos*

Los cinco primeros verbos de la lista "básica" son:

—comer (número 142).
—estudiar (número 152).
—cantar (número 174).
—jugar (número 183).
—beber (número 218).

Como en el caso de las palabras gramaticales. Es de destacar que el primer verbo tenga sólo el número 142, que entre los primeros no estén presentes, por ejemplo, "haber", "ser" o "estar" y, por último, es interesante ver la importancia de los verbos "estudiar", "cantar" y "jugar", siendo esto coherente con los fines y premisas de la encuesta.

Para comparar esta lista a las dos que nos sirven de referencia (ver párrafo anterior), se plantea un problema: la *Lista Alfabética* contiene 1.553 vocablos, el *Frequency Dictionary,* por su particular presentación de los resultados, nos permitió hacer los cálculos pertinentes sobre esta misma cantidad (18), por la lista del *Vocabulario Básico* sólo contiene 1.352, es decir, 200 menos.

Deberemos, pues, efectuar una extrapolación de los resultados de los primeros 1.352 vocablos y adaptarlos a una lista hipotética de 1.552. Es esta una operación delicada y poco recomendable en estadística léxica, ya que la proporción entre las diferentes categorías gramaticales varía a lo largo de una lista de frecuencias (19) pero, en este caso, como la extrapolación sólo atañe a 200 hipotéticos vocablos suplementarios, el margen de error no es muy grande.

Si examinamos el número de verbos a lo largo de la lista del *Vocabulario Básico,* obtenemos:



BOLETÍN AEPE Nº 16. Raimundo EZQUERRA. LOS DICCIONARIOS DE FRECUENCIA EN ESPAÑOL (II)
Centro Virtual Cervantes

| N.º de vocablos | N.º de Verbos | Incremento |
|---|---|---|
| 200 | 4 | – |
| 400 | 18 | 14 |
| 600 | 27 | 9 |
| 800 | 33 | 6 |
| 1.000 | 47 | 14 |
| 1.200 | 64 | 17 |
| 1.352 | 78 | 14 |

Observamos pues que en la zona de frecuencias que nos interesa el incremento en el número de verbos tiene tendencia a aumentar, situándose entre 14 y 20. Podemos pues, razonablemente, atribuir a la zona situada entre el vocablo 1.400 y el 1.550, un incremento en el número de verbos del orden de 18, con un margen de error bastante mínimo. El número teórico de verbos en los 1.552 primeros vocablos de la lista del *Vocabulario Básico* sería pues de unos 78 + 18 = 96, lo que arrojaría en porcentajes los resultados siguientes:

– *Lista Alfabética:* 15,1 por ciento.
– *Frequency Dictionary:* 23,1 por ciento.
– *Vocabulario Básico:* 6,2 por ciento.

Una vez más queda demostrada la especificidad del *Vocabulario Básico* y su aspecto conceptual-sustantivo que ya anunciábamos al principio.

*5.3.3. Los sustantivos*

Los primeros 69 vocablos de la lista del *Vocabulario Básico* son todos sustantivos, y entre los 141 primeros los 5 adjetivos siguientes:

–azul (número 70).
–rojo (número 87).
–verde (número 89).
–amarillo (número 109).
–negro (número 127).

más, probablemente, "tonto" (número 115) y "rosa" (número 131).

La proporción de sustantivos en la lista es de 87,5 por ciento, que aún suponiendo una baja si la prolongáramos hasta el número 1.552, se situaría, por lo menos, hacia el 85 por ciento. Si comparamos esta cifra con las de las otras dos encuestas:

– *Lista Alfabética:* 49,2 por ciento.
– *Frequency Dictionary:* 46,3 por ciento.
– *Vocabulario Básico:* 85 por ciento.

vemos que, definitivamente, se confirma el aspecto nada "básico" de nuestro vocabulario.



BOLETÍN AEPE Nº 16. Raimundo EZQUERRA. LOS DICCIONARIOS DE FRECUENCIA EN ESPAÑOL (II)
Centro Virtual Cervantes

Para resumir, diremos en pocas líneas lo que ya decíamos en la primera página: el *Vocabulario Básico del Español y sus Aplicaciones a la Enseñanza* es una interesante encuesta sobre el léxico pasivo de los alumnos españoles de enseñanza media, con posibles interesantes aplicaciones en sociolingüística, psicolingüística e, indirectamente, en la pedagogía, pero no es un "vocabulario básico".

## B.—ACTUALIZACION DE LOS DATOS GENERALES SOBRE LOS DICCIONARIOS DE FRECUENCIA EN ESPAÑOL

Ampliando los cuadros generales de nuestro trabajo de 1974, podremos ahora señalar, en síntesis, las principales características de los siguientes diccionarios:

—BUCHANAN.—*A Graded Spanish Word Book.* Toronto, University Press. 1927.

—RODRIGUEZ BOU.—*Recuento de Vocabulario Español.* Universidad de Puerto Rico. 1952.

—GARCIA HOZ.—*Vocabulario común, vocabulario usual y vocabulario fundamental.* Madrid, CSIC. 1953.

—JULLIAND, CHANG RODRIGUEZ.—*Frequency Dictionary of Spanish Words.* La Haya, Mouton and Co. 1964.

—RIVENC, ROJO SASTRE.—*El Español Fundamental.*

—MARQUEZ VILLEGAS.—*Vocabulario del español hablado.* Madrid, SGEL. 1975.

—GARCIA HOZ.—*El Vocabulario General de Orientación Científica y sus Estratos.* Madrid, CSIC. 1976.

—DIAZ CASTAÑON.—*Vocabulario Básico del Español y sus Aplicaciones a la Enseñanza.* Universidad de Oviedo (ICE). 1977.

Sólo hemos eliminado de nuestros cuadros la ya estudiada en 1974 *Spanish Idiom List* de Keniston, por no tratarse de un diccionario, sino de una lista de frases e idiomatismos.



BOLETÍN AEPE Nº 16. Raimundo EZQUERRA. LOS DICCIONARIOS DE FRECUENCIA EN ESPAÑOL (II)
Centro Virtual Cervantes

Centro Virtual Cervantes

| | BUCHANAN (1927) | R. BOU (1952) | G. HOZ (1953) | JULLIAND (1964) | RIVENC (1973) | M. VILLEGAS (1975) | G. HOZ (1976) | D. CASTAÑON (1977) |
|---|---|---|---|---|---|---|---|---|
| Extensión del corpus | 1.200.000 | + de 7.000.000 | 400.000 | 500.000 | 1.200.000 | 800.000 | 2.302.000 | 850.000 |
| Fuentes Literarias | 90% | 31% | 25% | 80% | ○ | ○ | ○ | ○ |
| Fuentes escritas no literarias | 10% | 41% | 50% | 20% | ○ | ○ | 100% | ○ |
| Fuentes no escritas | ○ | 28% | 25% | ○ | 100% | 100% | ○ | 100% |
| Objetivos mas bien pedagógicos | + | + | + | – | + | + | + | – |
| Noción de Disponibilidad | – | – | + | – | + | + | – | – |
| Corpus "frecuente" | + | + | + | + | + | + | + | – |
| Corpus "disponible" | – | + | – | – | + | + | – | + |
| Homogeneidad geográfica | – | – | + | + | + | + | + | + |
| Homogeneidad temporal | – | + | + | + | + | + | + | + |
| Norma claramente definida | – | – | + | + | + | – | + | + |
| Recuento separado flexiones | – | + | – | + | + | – | – | – |
| Recuento de todos los vocablos | – | – | + | + | + | – | – | + |
| Indice de clasificación | I = f (F, R) | I = f (E) | I = f (F) | I = f (F, R) | I = f (E) | I = f (F) | I = f (F, R) | I = f (F) |

**NOTAS:**

[1] EZQUERRA, R.—*Los Diccionarios de Frecuencia en Español.* En "Boletín de la AEPE". número 10. Madrid, AEPE. 1974.

[2] EZQUERRA, R.—*Una nueva encuesta sobre el español hablado.* En "Revista Española de Lingüística", número 6,1. Madrid, Gredos. 1976 (Estudio de la encuesta de MARQUEZ VILLEGAS).

EZQUERRA, R.—*El Vocabulario General de Orientación Científica y sus Estratos.* En "Revista Española de Lingüística", número 7,2. Madrid, Gredos. 1977. (Estudio sobre la obra de GARCIA HOZ).

[3] Para la diferencia "léxico activo/léxico latente" ver (entre otras obras): J. CASARES.—*Nuevo concepto del Diccionario de la Lengua.* Madrid, R.A.E. 1921. (páginas 35, 36) y GOUGENHEIM y otros.—*l'Elaboration du Français Fondamental.* París, Didier. 1964.

[4] Ver las tres primeras líneas del párrafo 5.º (página 16) del primer tomo del *Vocabulario Básico.*

[5] EZQUERRA, R.—*Los Diccionarios de Frecuencia...* (páginas 12, 13 y 14).

[6] DIAZ CASTAÑON.—*Op. cit.*, tomo primero, página 16.

[7] DIAZ CASTAÑON.—*Op. cit.,* tomo primero, página 15.

[8] DIAZ CASTAÑON.—*Op. cit.,* tomo primero, página 15.

[9] RODRIGUEZ BOU.—*Recuento de Vocabulario Español.* Universidad de Puerto Rico. 1952, página 8.

[10] GOUGENHEIM, RIVENC, MICHEA, SAUVAGEOT.—*l'Elaboration du Français Fondamental.* París, Didier, 1964 (primera edición: 1956). Ver, en particular, la tercera parte: "La Disponibilité".

[11] Traducimos aquí las indicaciones dadas por GOUGENHEIM, etc., en la obra citada en la nota 10.

[12] DIAZ CASTAÑON.—*Op. cit.*, página 17.

[13] Nos referimos aquí a la lista obtenida a partir de las "asociaciones libres".

[14] Naturalmente, aquí se trata del concepto no matemático de "frecuencia", es decir, que F debe ser entendido como suma de ocurrencias de un vocablo. Para el concepto de frecuencia y su utilización en estadística léxica, ver nuestro artículo: *El Vocabulario General de Orientación Científica y sus Estratos,* en su párrafo primero.

[15] Utilizamos "campo léxico" como conjunto de vocablos de la lengua que designan los diferentes aspectos de una noción, una técnica, un objeto, etc., por oposición al "campo semántico" o conjunto de significados de una palabra.

[16] GOUGENHEIM, etc.— *Op. cit.*, página 161.

[17] Entendemos aquí por "palabras gramaticales" a las interjecciones, preposiciones, artículos, conjunciones y pronombres. Algunos autores incluyen también a los demostrativos y a algunos adverbios.

[18] Ver la página 19 de nuestro trabajo: *Los Diccionarios de Frecuencia en Español.*

[19] Ver a este propósito el final del párrafo primero de nuestro artículo: *El Vocabulario General de Orientación Científica y sus Estratos.*



BOLETÍN AEPE Nº 16. Raimundo EZQUERRA. LOS DICCIONARIOS DE FRECUENCIA EN ESPAÑOL (II)
Centro Virtual Cervantes