

Por Raimundo EZQUERRA  
Universidad de Besançon.

## LOS DICCIONARIOS DE FRECUENCIA EN ESPAÑOL

**1.- PROPOSITOS DE ESTE TRABAJO.** La confección de recuentos o estudios estadísticos sobre el vocabulario utilizado por los hispanohablantes y su posterior utilización con fines lingüísticos, en sincronía, diacronía o lingüística aplicada a la enseñanza del español como lengua extranjera, es un problema cada vez más actual. En efecto, nadie hoy puede ya pensar seriamente en confeccionar un método de español para extranjeros sin consulta previa de algún recuento de vocabulario. También parece cada vez más evidente a los investigadores la utilidad de tales recuentos para conocer las características de la lengua en un momento dado de su historia.

Dentro de esta doble perspectiva, teórica y práctica, nos ha parecido útil el volver la vista atrás y tratar de hacer una pequeña recapitulación comparativa de los trabajos efectuados hasta hoy en este campo de la Lingüística. Trataremos en nuestro trabajo de ver cómo las nociones teóricas previas a la confección de un recuento, tales como frecuencia, repartición, clasificación de los vocablos, extensión del "corpus", etc., han evolucionado en su contenido o en su aplicación a través de las diferentes obras conocidas por nosotros. También, como conclusión, trataremos de comparar los resultados de estos diferentes trabajos, es decir, de comparar las características de las listas obtenidas. Esto último nos permitirá llegar a ciertas conclusiones sobre la utilización de las listas.

Antes de continuar, y para que no haya confusiones de aquí en adelante, nos parece necesario hacer un pequeño paréntesis teórico y fijar el significado que daremos en nuestro trabajo a las dos nociones, vecinas pero no idénticas, de "vocablo" y "palabra". De acuerdo con Ch. Muller, (1) de la Universidad de Estrasburgo, diremos que el vocablo es una unidad de léxico y la palabra una unidad de texto. Si, por ejemplo, un autor utiliza para su recuento un "corpus" de 10 páginas de novela en la que hay 500 palabras, y entre estas encuentra 70 diferentes, la extensión del corpus es de  $N = 500$  unidades o palabras, y la extensión de la lista resultante es de  $70 V = 70$  unidades, o vocablos. Muller utiliza su definición sólo en el caso de las palabras gráficas (2) nosotros la aplicaremos también a los corpus orales, obtenidos por grabación. Naturalmente, los problemas planteados por la noción de palabra, no están, ni mucho menos, resueltos, pero aunque es éste un problema que nos preocupa, su desarrollo rebasaría los límites de este artículo. Aceptaremos pues, provisionalmente, la noción de palabra gráfica como unidad válida para efectuar un recuento, ya sea a partir de textos escritos, ya sea a partir de la transcripción, fonética, taquigráfica u ortográfica, de textos orales. Nuestra noción de palabra

gráfica es pues más amplia que la de Muller que se ocupa, en la obra citada, exclusivamente de obras literarias impresas.

**2.— MATERIAL UTILIZADO.** No hablaremos aquí de las obras de referencia, artículos o trabajos de lingüística general, que serán citados, a medida que sea necesario, a lo largo del artículo. Sólo señalaremos a continuación, por orden cronológico, los diccionarios de frecuencia y recuentos estudiados en este trabajo. Estos son:

- BUCHANAN.— *A Graded Spanish Word Book*. Toronto, University Press. 1927.
- KENISTON.— *Spanish Idiom List*. N. York, The Macmillan Company. 1929. (3)
- RODRIGUEZ BOU.— *Recuento de Vocabulario Español*. Universidad de Puerto Rico. 1952.
- GARCIA HOZ.— *Vocabulario común, vocabulario usual y vocabulario fundamental*. Madrid, C.S.I.C. 1953.
- JULLIAND, CHANG-RODRIGUEZ.— *Frequency Dictionary of Spanish Words*. La Haya, Mouton and Co. 1964.
- RIVENC, ROJO-SASTRE.— *El Español Fundamental* (4).

Aunque existen otras obras sobre el vocabulario español y su frecuencia (5), estas son las principales, y las que han servido y sirven, hasta hoy, para la elaboración de métodos de Español y como base de referencia para estudios estilísticos y filológicos.

Por último, antes de comenzar el estudio propiamente dicho de los diferentes vocabularios, agradezco públicamente su ayuda a cuantos, con sus indicaciones, me han permitido dar con la pista de ciertos volúmenes de difícil acceso. En Francia, A. Sauvageot y R. Michéa, además de P. Rivenc, coautores de *L'Elaboration du Français Fundamental*, y en España, M.<sup>a</sup> Rosa Suárez Inclán, de A.E.P.E., y Pedro García de O.F.I.N.E.S.

**3.— FINALIDADES.** La finalidad con que los autores de los recuentos estudiados han efectuado estos trabajos, es variable. A menudo la preocupación por las aplicaciones pedagógicas ha dominado la investigación: "... la finalidad de esta investigación no es puramente filológica ni lingüística, sino que se concibe el vocabulario como un medio de conocimiento psicológico y como contenido de enseñanza..." (6). A veces se trata de la enseñanza del Español como lengua extranjera, este es el caso, por ejemplo en el *Español Fundamental*, y otras veces se trata tanto de su enseñanza como lengua extranjera, como de su enseñanza como lengua materna, este es el caso del *Recuento de Vocabulario Español*.

En otros casos, las preocupaciones didácticas no son primordiales para los autores: "Improvement of elementary language teaching is only a secondary byproduct of our investigation". (7).

En realidad, ningún autor excluye el doble aspecto posible en la utilización de los resultados obtenidos. Por ello nos sentimos autorizados a examinar estos vocabularios desde un punto de vista didáctico. Se puede decir, sin embargo, que, menos el *Frequency Dictionary*, todos los demás vocabularios obedecen a preocupaciones más bien pedagógicas.

**4.— EL CORPUS.** Las características y la importancia cuantitativa del corpus empleado por los diferentes autores es muy variable. Esta variación obedece a criterios explícitos o no explícitos, o a obligaciones materiales. Vamos a tratar de examinar todo esto a continuación.

**4.1.— Importancia cuantitativa del corpus.** El número de palabras necesarias, según los autores, para hacer un buen recuento de vocabulario español, ha sufrido variaciones considerables según la época de que se trate. En efecto, parece como si en una primera generación de recuentos se hubiera tendido a establecer un corpus tan extenso como lo permitieran los medios materiales con que se contaba, para, en una segunda generación, en la cual se situarían los recuentos más recientes, disminuir considerablemente esta extensión del corpus, por lo menos desde un punto de vista estrictamente cuantitativo. Verdad es que en estos últimos casos, la menor extensión del corpus se ha acompañado de un mayor rigor matemático a la hora de definir los criterios de clasificación de vocablos.

Por ejemplo, Buchanan trabajó a partir de un corpus de 1.200.000 palabras, del que dedujo una lista de 18.331 vocablos. (8). Rodríguez Bou efectuó su recuento a partir de un corpus de 7.066.637 palabras. Es éste el corpus más extenso utilizado hasta hoy para un recuento de vocabulario español. En esta época parecía evidente que cuanto más extenso fuera el corpus, más garantías ofrecería el recuento correspondiente: "... el mejor estudio de frecuencia de uso del vocabulario español publicado hasta el presente es el del profesor Buchanan. Sin embargo, la limitación de su recuento a 1.200.000 palabras, (.....), introduce limitaciones que nos propusimos superar..." (9). García Hoz imprime una dirección diferente a los recuentos de vocabulario pues, a partir del *Vocabulario usual*... en donde encontramos por primera vez este nuevo aspecto de enfocar el problema de que habláramos más arriba, un corpus restringido, 400.000 palabras en este caso, y criterios y técnicas matemáticas mucho más evolucionadas que en los casos anteriores. García Hoz justifica la extensión de su corpus por las afirmaciones de Watts (10). Este autor da como cifra media para la extensión del vocabulario de un adulto la de 15.000 vocablos. García Hoz había encontrado 11.000 vocablos diferentes en las primeras 240.000 palabras, por lo que pensó que un corpus de 400.000 sería suficiente para llegar a la cantidad citada por Watts (11). "Por otra parte, el recuento de 400.000 palabras, (...), dio como resultado la aparición de 12.913 palabras, cifra que, aumentada con los vocablos específicos de la profesión, lugar, etc., que cada hombre posee, viene a coincidir con la aproximativa señalada por Watts, nueva razón para considerar suficiente el recuento..." (12). A esto se añaden otras razones, más de fondo. Para García Hoz, en efecto, un corpus de varios millones de palabras supondría forzosamente una lista con vocablos no usuales. Naturalmente, esto se prestaría a discusión, ya que se puede muy bien concebir una lista que, efectivamente, comprenda vocablos no muy usuales, pero el verdadero problema no es éste. Nada habría que objetar si los vocablos no usuales incluidos en la lista fueran dados como tales, con un rango que correspondiera a su frecuencia real de uso y de utilización en la lengua. El verdadero problema está, a nuestro parecer, en los criterios matemáticos y las técnicas utilizadas en el recuento, más que en la extensión del corpus. Dicho de otra forma, sobre la cuestión de fondo estamos de acuerdo con García Hoz, pero su argumento, citado en último lugar, no nos parece muy adecuado.

En cuanto al *Frequency Dictionary of Spanish Words*, fue el resultado del recuento de "... a corpus of about 25.000 sentences totalling about 500.000 words." (13). Este número de palabras constituyentes del corpus sitúa al *Frequency Dictionary*

de lleno en la segunda generación de recuentos, con corpus reducido, aunque los autores escriben en la introducción que el corpus utilizado sitúa a su trabajo entre los dos tipos citados. (14). Más lejos, en su introducción, los autores del *Frequency Dictionary* dan, de su decisión de trabajar a partir de un corpus de 500.000 palabras, razones que, a nuestro parecer, sitúan el problema en su verdadero plano: "The smallest adequate size for a sample underlying a basic vocabulary conceived in these terms is determined by the point at which additional data cease to affect the formal properties of the classes..." (15). Naturalmente, se trata de las "statistical classes" entre las que se pueden distribuir los vocablos de la lista obtenida por el recuento. (16) Además, los autores justifican su decisión por el hecho de que tal cantidad de palabras es ampliamente suficiente para todos los fines no pedagógicos del recuento. Recordemos que el establecimiento de una lista básica aprovechable para fines pedagógicos no es más que una preocupación secundaria para los autores. Añadamos que este criterio tiene la ventaja sobre todos los que hemos visto hasta ahora, de tener una objetividad científica que no está siempre presente en los trabajos de este tipo y, que caracteriza, como veremos más adelante a esta obra. No queremos decir con esto que la "objetividad científica" no esté sujeta a discusión como todo lo demás, pero, por lo menos, la discusión será facilitada y clarificada si los argumentos de cada uno son claros, precisos y buscan la objetividad.

Por último, nos queda por ver cuál es la posición que, a este respecto toma *El Español Fundamental*. La lista final de este trabajo fue establecida a partir de un corpus de 1.200.000 palabras, de las cuales 800.000 fueron obtenidas por grabación de conversaciones. Esta cifra parece colocar a *El español Fundamental* entre los recuentos que hemos llamado de la "primera generación". Esto es, aparentemente, contradictorio con el hecho de que, cronológicamente, *El Español Fundamental* sea el más reciente de los recuentos efectuados en Español. En efecto, *El Español Fundamental* posee características que lo sitúan como el más moderno de los trabajos sobre vocabulario fundamental español, como, por ejemplo, la importancia que en el corpus tiene la lengua hablada, al mismo tiempo que desde otros puntos de vista, por ejemplo los criterios de selección, podemos decir que utiliza criterios menos elaborados que otros diccionarios aparecidos anteriormente. Más adelante examinaremos estos problemas más detalladamente.

**4.2.— Características del corpus.** Cuando Buchanan efectuó su recuento, el hecho de que una lengua es algo que evoluciona con gran velocidad, sobre todo desde el punto de vista léxico, no debía de estar tan presente en los pensamientos de los filólogos como lo está hoy en día. En efecto, las fuentes utilizadas por Buchanan para establecer su corpus van del siglo 17 hasta el siglo 20. Este hecho hace que, independientemente de otros criterios, los resultados de Buchanan sean difícilmente comparables con los de otros trabajos como el *Frequency Dictionary* o el *Español Fundamental* que se caracterizan, entre otras cosas, por una gran homogeneidad temporal. Buchanan dividió su corpus de 1.200.000 palabras en 40 "unidades" de 30.000 palabras cada una. La división del corpus en estas "unidades", es una de las piedras angulares del razonamiento matemático de Buchanan: "Categories or units of approximately equal length (30.000 running words), were used in order to provide for each word a credit number..." (17). Estas 40 unidades se reparten de la forma siguiente: Teatro: 7 unidades; Novela: 8 unidades; Obras en verso: 3 unidades; Folklore: 3 unidades; viajes, descripciones geográficas, etc.: 8 unidades; Técnicas y Ciencias: 7 unidades; Publicaciones periódicas: 4 unidades. Otra característica de este corpus es la heterogeneidad geográfica. Algunas de estas unidades, cuyo contenido exacto sería demasiado largo de explicitar en este artículo, contienen textos es-

critos por autores hispanoamericanos. Este último hecho, a diferencia de la heterogeneidad temporal, ya señalada, nos parece más bien positivo, sobre todo si trata de obtener, como en el caso presente, una lista fundamental de vocabulario para confeccionar métodos para la enseñanza del Español a extranjeros. Buchanan quiso determinar un verdadero vocabulario fundamental o de base; "A word count determines the common denominator, or the essential basis of language..." (18). Pero para él, y en esto difiere de obras más modernas, un vocabulario de base debe de abarcar niveles de lengua bastante diferentes del de la lengua corriente, de todos los días, como el nivel poético, y gran parte del novelístico o teatral. (19). No hay que olvidar, para comprender mejor estos criterios, que sólo recientemente, desde hace algunos años, se empezó a afirmar que la enseñanza de un primer nivel, de base, de una lengua extranjera, se debía de hacer basándose en la lengua corriente, de todos los días, tal como la habla el hombre medio en la calle, con hombres de otras profesiones diferentes de la suya propia. Hasta hace muy poco, no se concebía la enseñanza de una lengua extranjera sin hacerlo a través de la literatura de esta lengua. Tampoco se concebía la enseñanza de una lengua sin dar, paralelamente, la máxima cantidad posible de datos sobre la civilización representada por esta lengua. Hoy, nuestras ideas, las de los docentes de lengua, literatura, lingüística, civilización, etc., son más claras sobre este problema. Para terminar, diremos pues que la composición del corpus de Buchanan, tiene una ventaja notable, la heterogeneidad geográfica, y dos desventajas principales, la heterogeneidad temporal y la importancia relativa dada a las fuentes de tipo literario. Ventajas y desventajas, se entienden en la perspectiva de elaborar un vocabulario de base para la enseñanza del Español como lengua extranjera.

La composición del corpus del *Recuento de Vocabulario Español* significa un cambio de rumbo fundamental en la concepción que, desde aquel momento, los autores tienen de un vocabulario fundamental, usual, o frecuente.

En efecto, el corpus de Rodríguez Bou tiene las dos características esenciales que todos los demás recuentos tendrán desde entonces; utilización de fuentes no exclusivamente literarias, y utilización de fragmentos de lengua no escrita. La primera característica, se puede decir que apuntaba ya en el recuento de Buchanan, en la medida en que éste utilizaba la prensa periódica como fuente de palabras para el corpus. Pero en el *Recuento de Vocabulario Español*, estas fuentes, escritas, pero no literarias toman una gran importancia, en efecto, podemos clasificar en este tipo de fuentes a las composiciones escritas de alumnos de Puerto Rico, con 803.622 palabras a los periódicos, con 1.050.000 palabras, a los programas de radio (20), con 465.600 palabras, y a la recopilación de Rodríguez y Casanova, con 586.141 palabras provenientes de composiciones escritas por alumnos portorriqueños. En total, más o menos un tercio de la extensión total del corpus.

En cuanto a las fuentes no escritas, podemos clasificar en esta categoría, en primer lugar a lo que el mismo Rodríguez Bou llama "vocabulario oral" (21), con 1.073.245 palabras oídas y apuntadas por sus colaboradores, y a las 926.404 palabras provenientes de los ejercicios de "asociación libre y controlada". Esta última categoría nos parece interesante porque se trata de técnicas que han sido más tarde vueltas a emplear y desarrolladas por los autores de *El Español Fundamental* para determinar el vocabulario disponible. La técnica empleada por Rodríguez Bou consiste en la presentación al alumno, puesto que se trata de alumnos de enseñanza media, de listas de "palabras-estímulo", 90 palabras, diciendo que escriba todas las palabras que se le ocurran en función de cada pala-

bra estímulo. Esto en cuanto a la “asociación controlada”, la “asociación libre” consiste en pedir a los alumnos que escriban las palabras que se les ocurre, sin ningún estímulo, durante cinco minutos. La técnica aplicada por Rivenc y Rojo-Sastre, ya aplicada por los autores del *Français Fondamental*, es ligeramente diferente, está basada en los “centros de interés”. Más adelante la comentaremos.

El resto del corpus del *Recuento de Vocabulario Español* está constituido por palabras obtenidas de fuentes más tradicionales. En primer lugar, el corpus de Buchanan, vuelto a utilizar por Rodríguez Bou; libros de texto, publicaciones de índole religiosa así como “textos de lectura suplementaria”, es decir, lo que se podría denominar categoría tradicional o de carácter más bien literario.

Como se ve por lo que precede, no aplicamos la distinción de Rodríguez Bou: “Vocabulario oral”, “vocabulario de reconocimiento” y “vocabulario basado en el juicio de distintos autores”, porque no nos parece bastante explícita, y, además, nos parece mezclar en una misma rúbrica fuentes de índole diferente. Sobre todo, nuestra clasificación tiene como uno de sus fines principales el poder ser aplicable a todos los recuentos, lo que no ocurriría si aplicaríamos una cualquiera de las clasificaciones de los autores. Este es un problema general. En un trabajo de este tipo, uno de los problemas fundamentales está constituido por la definición de criterios de descripción que sean, al mismo tiempo, criterios de comparación. No ignoramos, por otro lado, que esta servidumbre le quita muchas veces precisión y profundidad a la descripción.

El recuento de García Hoz marca un pequeño retroceso respecto al de Rodríguez Bou en lo que respecta a las fuentes no escritas del corpus. En efecto, un cuarto del corpus, es decir 100.000 palabras, está constituido por 620 cartas familiares, provenientes de las clases más modestas de la sociedad, trabajadores agrícolas o de la industria, para contrarrestar la influencia de otras partes del corpus provenientes de niveles lingüísticos más ricos. Mediante estas cartas, García Hoz piensa poder reflejar en su corpus a la lengua de expresión oral y espontánea. Nosotros pensamos que, en efecto, así es en gran parte, pero que, de todas maneras, nada sustituye a la encuesta “in vivo” sobre la expresión oral. La lengua de las cartas, aunque sean familiares, no puede coincidir exactamente con la de expresión oral espontánea. Naturalmente, aquí, como en el caso de Buchanan, sería injusto olvidar las enormes dificultades que hubiera causado el tratar de hacer una encuesta directa sobre el vocabulario efectivamente hablado en España. El mismo autor es el primero en reconocer el papel jugado por las servidumbres materiales: “En cuanto al procedimiento para determinar qué palabras constituyen el vocabulario usual, el mejor sería el directo, que consiste en ir registrando las palabras que dice en la vida corriente el hombre corriente, así como las que lee o escribe, comprendiéndolas. Pero este procedimiento hay que desecharlo como físicamente imposible...”. (22). Recordemos, para dar razón a García Hoz, que, varios años después, los autores del *Français Fondamental*, aprovechando las posibilidades de los primeros grabadores de discos autónomos, pudieron efectuar esta encuesta sobre el Francés, pero gracias a enormes molestias y sacrificios. Sólo desde hace pocos años, con la aparición de los magnetófonos autónomos ligeros, se pueden hacer encuestas sobre la lengua hablada, sin demasiados inconvenientes.

El resto del corpus de García Hoz está constituido por fuentes escritas de tipo no literario, como las 100.000 palabras de los periódicos o de las publicaciones periódicas estatales, eclesiásticas y sindicales (otras 100.000), (23) y por fuentes literarias: los libros más vendidos entre 1943 y 1949.

Un aspecto muy interesante del recuento de García Hoz es que, por primera vez, vemos aparecer en él la noción de “*léxico activo*” y “*léxico latente*”, noción que será más tarde vuelta a emplear por otros autores como veremos más adelante. (24). El léxico activo es “el conjunto de palabras que un hombre emplea corrientemente en su conversación o escritura espontánea”, y el léxico latente está constituido por “palabras que sin ser usadas de modo espontáneo, son, sin embargo, comprendidas cuando se oyen o se leen.” (25). También afirma el autor: “Léxico activo y léxico latente forman parte del que puede llamarse *vocabulario usual* en contraposición al *vocabulario total*.” (26).

En resumen, el corpus de García Hoz es homogéneo temporal y geográficamente, y parece bien adaptado a fines pedagógicos, pero siempre dentro de la perspectiva tradicional, según la cual el vocabulario que aparece en las obras literarias debe de jugar su papel en el aprendizaje de una lengua extranjera. Este es un punto común con el recuento de Rodríguez Bou.

El *Frequency Dictionary* no es una encuesta cuyo objeto esencial sea la enseñanza del Español como lengua extranjera. Las características de su corpus deben pues analizarse, desde un punto de vista diferente del que hemos tenido hasta ahora. El corpus que ahora estudiamos está exclusivamente constituido por palabras provenientes de fuentes escritas, algunas de tipo no exclusivamente literario (prensa periódica). Como en el *Vocabulario usual...*, en el *Frequency Dictionary* tenemos un corpus constituido por unidades de 100.000 palabras cada una, pero aquí las unidades son cinco mientras que en la obra anterior éstas eran cuatro. La homogeneidad de este corpus ha sido muy cuidada: “To insure spatial homogeneity, only Spanish peninsular sources were admitted..” y “To insure temporal or chronological homogeneity, only works written between the two world wars were admitted...” (27). Esta homogeneidad nos parece indispensable para estudios de tipo filológico, lo que constituye uno de los fines fundamentales del *Dictionary*, pero, desde el punto de vista de la didáctica del español, esta unidad geográfica no nos parece provechosa, sobre todo si se trata de la enseñanza del español como lengua extranjera. Hasta la fecha, todavía no se ha hecho una buena encuesta con fines pedagógicos que tienda a establecer una especie de vocabulario fundamental del mundo hispánico. No dudamos que tal trabajo provocaría enormes dificultades, tanto prácticas como, sin duda, teóricas, pero pensamos que ha llegado el momento de hacerlo.

En resumen, el corpus del *Frequency Dictionary* está constituido por cuatro quintas partes de palabras procedentes de lengua escrita, de tipo literario (Teatro, Novela, Ensayo y Técnicas diversas), y por una quinta parte de palabras provenientes de fuentes escritas pero no puramente literarias.

Por último, nos queda por analizar el corpus del *Español Fundamental*. Como los resultados completos y el informe científico correspondiente no han sido publicados todavía, debemos nuestra información, además de a la amabilidad del Sr. Rivenc, a la que ya hicimos alusión, a dos artículos y conferencias del Sr. Rojo Sastre (28) y a la información contenida en las publicaciones correspondientes al método audio-visual *Vida y Diálogos de España* (29). Además, dada la coincidencia en muchos puntos entre la metodología de los autores de *El Español Fundamental* y la de los autores de *Le Français Fondamental*, ciertas informaciones pueden ser obtenidas por la lectura de *L'Elaboration du Français Fondamental* (30).

Los autores de *El Español Fundamental* establecieron primero un corpus de 800.000 palabras resultantes de la transcripción de 1.600 grabaciones de conversaciones en varias pro-

vincias españolas. Señalemos de paso, que la homogeneidad territorial existe en este recuento como en el precedente, así como, por otro lado, una homogeneidad temporal que se puede calificar de absoluta. A partir de este corpus "oral", pudo establecerse una lista de vocablos "frecuentes", pero como esta lista no es muy rica en vocablos con fuerte contenido semántico, nombres concretos por ejemplo, los autores la completaron con cierto número de vocablos "disponibles" obtenidos a partir de otras 400.000 palabras provinientes de una encuesta basada en el método de los "centros de interés". Se trata, como ya lo hemos apuntado más arriba, del mismo método que el empleado por Rodríguez Bou y sus colaboradores, pero aquí la palabra-estímulo es reemplazada por el centro de interés. La selección de los vocablos provinientes del corpus "disponible" se hizo teniendo en cuenta la distribución, el rango de aparición y la frecuencia.

La diferencia fundamental entre *El Español Fundamental* y el *Recuento de Vocabulario Español* es que, en el *Recuento*, las palabras obtenidas por el método de asociación, equivalentes a las disponibles de *El Español Fundamental*, pasan a formar parte del corpus, en bruto, por decirlo así, mientras que en *El Español Fundamental* son los vocablos, ya clasificados, de este corpus, los que pasan a engrosar la lista de vocablos frecuentes, ya obtenidos independientemente de ellos.

El corpus de *El Español Fundamental* es pues el que más corresponde a los criterios modernos sobre la lengua que se debe de enseñar en un primer nivel de Español como lengua extranjera. La metodología del *Français Fondamental* ha sido aplicada aquí con más rigor, y más extensamente, que en aquella obra. Veremos más tarde cómo estas características del corpus y de metodología influyen en la lista de base obtenida.

**4.3.— Cuadro Recapitativo.** Los principales datos sacados a la luz por lo que precede serán más fácilmente comparados y consultados en el cuadro siguiente:



	<u>Buchanan</u> (1927)	<u>R. Bou</u> (1952)	<u>G. Hoz</u> (1953)	<u>Julliard, Chang</u> (1964)	<u>Rivenc, Rojo</u> (1973)
Extensión del corpus	1.200.000	+ de 7.000.000	400.000	500.000	1.200.000
Fuentes literarias	90 %	31 %	25 %	80 %	0 %
Fuentes escritas, no literarias	10 %	41 %	50 %	20 %	0 %
Fuentes no escritas	0 %	28 %	25 % (31)	0 %	100 %
Objetivos más bien pedagógicos	+	+	+	-	+
Noción de disponibilidad	-	-	+	-	+
Corpus frecuente	+	+	+	+	+
Corpus disponible	-	+	-	-	+
Homogeneidad geográfica	-	-	+	+	+
Homogeneidad temporal	-	+	+	+	+

**5.— LA TÉCNICA DEL RECuento.** De la misma forma que la composición del corpus es variable según los métodos empleados por sus autores, o según sus opiniones lingüísticas, los métodos y técnicas matemáticas para, a partir del corpus, obtener una lista de vocablos, son muy variables. Trataremos ahora de analizar, someramente, estas técnicas, para luego ver si las características conjugadas del corpus y de los criterios matemáticos o lingüísticos empleados para su explotación, se manifiestan en los resultados.

**5.1.— Criterios lingüísticos.** No es necesario insistir sobre los problemas que la definición de una *norma* plantean a todo aquél que se lance a hacer algún estudio estadístico de tipo lingüístico. (32). Tampoco tenemos la pretensión de resolver aquí problemas tan complejos como el del concepto de “palabra”. Sólo queremos indicar algunos de los criterios lingüísticos fundamentales que los diferentes autores han utilizado al efectuar sus encuestas y la explotación de éstas.

Buchanan cuenta en una sola unidad las diferentes flexiones de las palabras. Esto puede ser aceptable, pero lo que lo es menos es que, por ejemplo los adverbios en —mente sean contados junto con los adjetivos de que derivan. Los diminutivos, aumentativos, superlativos, peyorativos, “and other suffix forms”, en general, no son contados como formas independientes. Los adjetivos y participios en -ado, -ido, son contados, en general, con el infinitivo correspondiente, etc.

Como vemos, una de las características de esta norma es su falta de fijeza. El hecho de contar tal o tal forma independientemente o no, según los casos, tiene plena justificación desde un punto de vista lingüístico, pero no puede defenderse desde un punto de vista estadístico. En efecto, aunque la norma perfecta no existe, y toda norma no sea sino un compromiso entre las necesidades de la estadística y las realidades lingüísticas, para que una norma sea válida y aplicable debe, por lo menos, ser clara y explícita, de forma que cualquier caso pueda entrar en sus reglas. Es preferible utilizar normas simples y claras aunque a veces esto nos obligue a inexactitudes lingüísticas. Si existen algunas inexactitudes, o, más probablemente, algunos casos no evidentes, esto no tiene importancia si lo sabemos, si lo hemos enunciado claramente. El utilizador de nuestras estadísticas contará entonces con ello para sus deducciones.

La técnica de recuento de Buchanan tiene otra particularidad que encontraremos en otras obras, entre las menos recientes, y que hoy en día nos parece de difícil justificación. “In order that the count might be made economically and yet extend to at least a million running words, it was decided at the outset to eliminate a considerable number of such common words as it is certain would be within the first five hundred”. Así llegamos a que un centenar de palabras, 32 nombres, 26 verbos y cierto número de palabras gramaticales por razones de facilidad material, no forman parte del recuento pero figuran en la lista fundamental. Esta técnica se puede tratar de justificar en el caso en que el objetivo del recuento sea exclusivamente pedagógico, aunque incluso en este caso nos parezca tener más inconvenientes que ventajas, pero si quisiéramos utilizar los resultados del recuento para algún estudio lingüístico sobre el español de tal o tal época o lugar, este procedimiento hace inútiles a estos resultados. En efecto, son, a menudo, precisamente esos vocablos, los más frecuentes, los que más información nos dan sobre ciertas características de una lengua.

Apuntaremos, antes de examinar otro recuento, que, en la lista de Buchanan, las expresiones idiomáticas son contadas junto con determinadas palabras. Por ejemplo, “a me-

dias" es contada con "medio" y "a oscuras" con "oscuro". También aquí nos podemos preguntar cuál es la utilidad de tales procedimientos.

A partir de este párrafo haremos alusión a la obra de Keniston, pues las técnicas empleadas en ella pueden ayudarnos a trazar la evolución de los conceptos en el campo de los diccionarios de frecuencia de la misma forma que si se tratara de un diccionario de vocablos.

Este autor, a partir de un corpus (33) de 1.000.000 de palabras, obtiene una lista de expresiones idiomáticas del español. La primera dificultad con que se encuentra es, naturalmente, la de la definición de "expresión idiomática": "For the purposes of the present study, then, an "idiom" is defined as "an expression peculiar to Spanish or, an expression which differs in form from its English equivalent" (34). La definición no es muy precisa, pero pensamos que difícilmente puede serlo en este caso. El problema es todavía más complejo que en el caso de las palabras. Por otra parte, desde un punto de vista pedagógico, una lista de expresiones idiomáticas, como la de Keniston, es de una gran utilidad, por ejemplo para escribir diálogos para un método.

Rodríguez Bou cuenta separadamente las formas de inflexión de cada vocablo. Así, para 20.542 vocablos contados, cuenta 62.888 formas de inflexión. Rechazando las técnicas de Buchanan y Thorndike, acepta la de Horn: "Horn, sin embargo, en la lista (. . .) tuvo en cuenta las unidades de inflexión. Este plan nos pareció aconsejable y así contamos separadamente todas las variaciones en la forma de las palabras. Todas las formas de los nombres: masculino y femenino, singular y plural, común y propio, aumentativos y diminutivos; todas las formas verbales, adjetivales, adverbiales; todas las abreviaturas, etcétera, se contaron por separado." (35). Esto nos parece una gran virtud de este *Recuento de Vocabulario Español*. Sin embargo, más lejos dice el autor: "La lista final, sin embargo, no es de carácter semántico. Con frecuencia el recuento de las palabras no se hace a base de los distintos significados de la misma palabra." (36). Es éste un problema importante porque, por un lado, no se puede ignorar que no se puede hablar de verdadero recuento mientras las diferentes acepciones de los vocablos polisemánticos sean registrados en una sola entrada, y por otro lado, tampoco ignoramos que, en la medida en que se utilizan máquinas mecánicas o electrónicas para hacer los recuentos, se trata de un problema de difícil solución. Quizás haya que encaminarse hacia un tipo de recuento en que ninguna palabra ni vocablo sea aislada de su contexto. Esta solución ha sido intentada recientemente, en la elaboración de *El Español Fundamental*, (por lo menos en la *Lista Alfabética* que acompaña al método *Vida y Diálogos de España*). Otro tipo de solución, que consiste en dar sistemáticamente, con cada vocablo, la categoría gramatical a la que pertenece, ha sido intentado en el *Frequency Dictionary of Spanish Words*.

Una característica original del *Recuento de Vocabulario Español*, con relación a las obras anteriores, el haber acogido en el recuento ciertos neologismos, todavía no registrados en aquel momento en los diccionarios, pero de gran frecuencia: "Se aceptaron en las listas porque se han formado siguiendo las leyes de composición y derivación de palabras que rigen en el idioma y porque, además, son de uso corriente en la lengua oral —y a veces en la escrita— de un buen número de personas cultas." (37). Este punto de vista nos parece positivo y casi, diríamos, necesario, cuando se trata de hacer un recuento con fines pedagógicos, o con determinados fines lingüísticos.

Un gran defecto de este *Recuento* es el haber excluido de los cálculos, como Buchanan, a los 105 vocablos más frecuentes a partir del primer millón de palabras contado. Las

razones que da el autor son las mismas que las de Buchanan: "para ahorrar tiempo y dinero".

García Hoz, consciente de la dificultad de definir una unidad de recuento, emplea la noción de "palabra", por referencia a las normas de la R.A.E. en su *Diccionario Vulgar*.

También, como Buchanan, registra juntas las formas de inflexión: "Se ha registrado como un solo vocablo la misma palabra en sus variantes debidas a los accidentes gramaticales: los nombres, adjetivos y pronombres no personales en la forma masculina del singular y los verbos en infinitivo (. . .). Una sola dificultad se planteó: la de los participios pasivos." (38). García Hoz resuelve la dificultad en función del contexto, pero con la salvedad siguiente: ". . . he de advertir que para que un participio pudiera ser registrado como adjetivo necesariamente ha de ser consignado como tal en el Diccionario de la Real Academia". Lo que es congruente con la norma general aceptada por este autor.

Digamos que la norma del *Vocabulario usual, común. . .*, tiene el mérito de ser clara y, al parecer, aplicable sin ambigüedades ni casos imprevistos. Como hemos dicho más arriba, esto es, a juicio nuestro, una gran ventaja, independientemente del hecho de que una norma como ésta pueda ser discutida o revisada por los lingüistas.

El *Frequency Dictionary* utiliza esta misma técnica de recuento, todas las formas regulares de inflexión son registradas bajo una única entrada, para la clasificación de los vocablos, pero, al mismo tiempo, la lista obtenida nos proporciona las frecuencias y reparticiones particulares para cada forma de inflexión. Por ejemplo, en el caso de los verbos, bajo el infinitivo con la frecuencia, repartición y "usage" acumulados, podemos ver todas las formas registradas de este verbo, con indicaciones que permiten, en los casos dudosos, saber de qué persona o de qué tiempo se trata, sin posibilidad de error. Esta es también la metodología empleada por los autores de *El Español Fundamental* (39).

Hay también que señalar el hecho de que en ambos trabajos todas las formas y vocablos del corpus han sido registrados en el recuento. Como en el caso del *Vocabulario usual común y fundamental*, el error de Buchanan y de Rodríguez Bou ha sido evitado.

Todo esto hace que, a juicio nuestro, estas dos obras, el *Frequency Dictionary* y el *Español Fundamental* sean las que ofrezcan más posibilidades de utilización y de explotación de las listas obtenidas.

**5.2.—Criterios matemáticos.** Es quizás aquí en donde más salta a la vista la evolución en los conceptos y en las ideas lingüísticas en estos últimos cincuenta años, sin olvidar la evolución en las posibilidades materiales de trabajo, como por ejemplo las máquinas de calcular. Como es sabido, entre la evolución de los conceptos y la de las herramientas se ha establecido una serie de influencias recíprocas de cuyos buenos resultados tenemos la prueba en los diccionarios de frecuencia del Español.

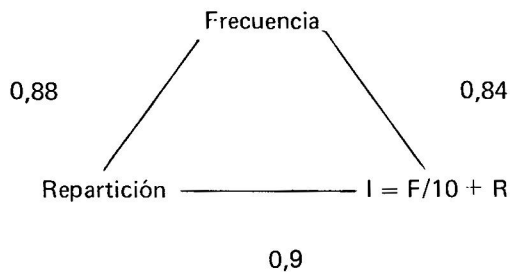
Ya Buchanan resumía en pocas líneas el problema, cuyo punto de partida sigue hoy siendo el mismo. "There are three methods by which the order of merit can be determined in a quantitative analysis: (1) by frequency, (2) by range, and (3) by a combination of the two." (40). En lo que sigue veremos como las tres posibilidades han sido explotadas con resultados diversos. La que más interpretaciones ha originado es, naturalmente, la tercera.

El mismo Buchanan escoge la tercera solución para clasificar los vocablos obtenidos: “. . . for each word a credit number representing a combination of range and frequency of occurrence. (. . . . .) The problem of finding a satisfactory method of combining the two factors was submitted to Professors Henmon, Sandiford and Thorndike, who, not without misgivings, decided upon the simple expedient of dividing the number representing frequency by ten and adding one for each category in which the word occurs.” (41).

Buchanan da el ejemplo siguiente: un vocablo que tiene una frecuencia de 751 (aparece 751 veces en el corpus), y una repartición de 40 (existe en cuarenta unidades, es decir, en todas en las que está dividido este corpus), tiene un índice de clasificación (“credit number”) de  $751/10 + 40 = 75,1 + 40 = 115,1$ .

Hoy, tal sistema de atribución de un índice a cada vocablo nos parece un tanto simplista y subjetivo, pero en aquella época esto podía ya ser considerado como un método extremadamente objetivo con relación a la tónica media de los estudios de letras, menos, quizás, en el campo de la Filología.

Buchanan también calcula en su estudio los índices de correlación, para los primeros 4.000 vocablos de su lista, entre los diferentes modos de clasificación. Estos índices pueden representarse de la forma siguiente:



La más alta correlación existe, como vemos, entre el índice de Buchanan y la repartición. Estos resultados han influenciado a otros autores posteriores.

En el caso de Keniston, esto no es evidente, pero es probable. Este autor, en efecto, parece dudar en clasificar las expresiones de su lista con relación a la frecuencia o a la repartición: “The question of whether range or frequency provides a better order of merit has been much discussed. If the 100 units had all been checked with perfect accuracy it is probable that the correlation between the two orders would have been so high as to make the choice indifferent.” (42).

Finalmente, toma partido por la clasificación con relación a la repartición decreciente. Si varias expresiones poseen la misma repartición, los clasifica por orden de frecuencia decreciente. Toda expresión cuya repartición sea inferior a tres es excluida de la lista, pero la frecuencia no tiene valor mínimo eliminatorio, a menos que supongamos que este valor mínimo sea también tres, una aparición del vocablo en cada una de las tres unidades en donde debe de estar representado.

Vemos, pues, que la *Spanish Idiom List* utiliza un método de clasificación de sus

unidades, todavía más simplista que el de Buchanan, ya que sólo uno de los datos obtenidos es prácticamente tenido en cuenta.

El criterio de Rodríguez Bou es también de este tipo, y explícitamente influenciado por los trabajos de Buchanan: "Aunque muchos recuentos de palabras dan crédito a la variedad de las fuentes, Buchanan demostró en su estudio que la correlación entre la variedad o extensión de las fuentes y la frecuencia es de 0,88. En vista de esta correlación positiva y alta, se decidió que, para fines prácticos, la frecuencia era suficiente en este recuento." (43). No creemos que tal índice de correlación sea suficiente para decidir que la clasificación por orden de frecuencia sea suficiente.

García Hoz marca, en este plano, un cambio de dirección fundamental, de la misma forma que en el plano de la composición del corpus lo marcaba Rodríguez Bou. En efecto, en el *Vocabulario usual, común y fundamental* encontramos tres listas de vocablos. La primera, que el autor denomina "vocabulario usual" que comprende 12.402 vocablos, todos los que aparecieron en el corpus, clasificados por orden alfabético, cada uno con su frecuencia total y con sus frecuencias parciales en cada uno de las cuatro partes del corpus. Otra segunda lista es la llamada "vocabulario común", en el que sólo encontramos los vocablos que aparecen en las cuatro partes del corpus de forma simultánea, es decir, que tienen una repartición  $R = 4$ . "... aparece inmediatamente una doble clase de palabras desde el punto de vista de su uso. Hay unas palabras que aparecen en los cuatro tipos de vocabulario (. . . . .), mientras otras aparecen solamente en algunos (. . . . .). Este conjunto de palabras que figuran en todos los tipos de vocabulario constituyen el que llamo *vocabulario común*". (44).

Naturalmente, surgió el problema de que ciertos vocablos poseían una repartición inferior a cuatro, pero una frecuencia bastante alta. García Hoz, consciente de la contradicción, pero no queriendo romper la regla que se impuso (en lo que le damos la razón), comunicó estos vocablos en una lista particular.

Obtiene así, un vocabulario común de 1.971 palabras ordenadas por orden alfabético, con una frecuencia  $F \geq 4$  y con una repartición  $R = 4$ .

Aquí vemos cómo los problemas que hoy se plantean los autores de encuestas lexicales aparecen por primera vez. García Hoz plantea el problema de los vocablos con  $R$  inferior a 4 pero con frecuencia importante, con relación a los vocablos de repartición igual a 4 pero con frecuencia baja. García Hoz ve el problema pero no da el paso siguiente que consistiría en buscar un índice que combine  $R$  y  $F$ , con suficientes garantías matemáticas, que permitiera clasificar los vocablos. Este paso será dado diez años más tarde por los autores del *Frequency Dictionary*.

Finalmente, encontramos el "vocabulario fundamental". En este vocabulario están recogidos los vocablos cuyas apariciones están distribuidas de forma regular entre las cuatro partes del corpus. Por ejemplo, un vocablo con frecuencia  $F = 400$  que aparece 100 veces en cada parte del corpus, forma parte del vocabulario fundamental. Este vocabulario comprende 208 vocablos.

Otra vez, aquí encontramos las preocupaciones de los autores del *Frequency Dictionary*, puestas en evidencia pero no resueltas de forma satisfactoria.

Julliard y Chang-Rodríguez dividieron su corpus en cinco partes iguales en importancia cuantitativa. El factor R puede, pues, tomar todos los valores enteros entre 1 y 5 ambos inclusive. Para cada vocablo, estos autores calculan primero la frecuencia media teórica (suma de las frecuencias parciales dividida por cinco), luego la variancia (45), y a partir de la variancia, un índice D de dispersión (46) que puede tomar todos los valores posibles entre 0 y 1. D será igual a 0 en los casos en que todas las apariciones de un vocablo ocurran en una sola de las cinco partes del corpus, y D será igual a 1 en los casos en que las apariciones del vocablo estén repartidas regularmente entre estas cinco partes.

He aquí pues, una forma de medir objetivamente y de hacer entrar en los cálculos, la distribución, por medio de este índice de dispersión. El defecto del método, si defecto hubiera, sería más bien debido a esta división del corpus en cinco partes, quizás no suficiente para que cada una de ellas sea realmente homogénea y significativa.

Luego, los autores obtienen un coeficiente U o "índice de uso", mediante el producto  $U = F \times D$ . De esta forma, se puede decir que U es el valor de la frecuencia, pero corregido, o ponderado, en función de la repartición del vocablo en el corpus. En efecto, U será igual a F en los casos en que D sea igual a 1, es decir, en los casos de repartición completamente regular, y U será igual a 0 en los casos en que D sea igual a 0. Es decir, que un vocablo con una frecuencia muy alta, si todas sus apariciones se sitúan en una sola de las cinco partes del corpus, será eliminado de la lista del *Frequency Dictionary*. Por lo tanto, estamos ante un criterio matemático que permite clasificar a los vocablos con bastantes garantías de objetividad.

*El Español Fundamental* adopta criterios mucho menos elaborados que son, en lo esencial, los ya adoptados en *Le Français Fondamental*. En efecto, en *El Español Fundamental* los vocablos son clasificados según las frecuencias decrecientes. En los casos en que hay varios vocablos con la misma frecuencia, el criterio de clasificación es el de la repartición decreciente. Una diferencia fundamental entre el significado del factor R en esta obra y en el *Frequency Dictionary* o el *Vocabulario usual, común y fundamental*, es que, en estos últimos R puede tomar valores entre 4 0 y 4 ó 5, mientras que en el trabajo que estudiamos R puede variar entre 0 y 1.600. (1.600 unidades de 500 palabras cada una).

El interés fundamental de este trabajo reside, pues, más en el carácter "oral" del corpus, completado por un corpus de palabras "disponibles", que en los criterios de selección utilizados. Será interesante la comparación de los resultados obtenidos por estos dos trabajos, en los que corpus y criterios matemáticos de selección son tan diferentes.

### 5.3.—Cuadro recapitulativo.

	<i>Buchanan</i> (1927)	<i>Keniston</i> (1929)	<i>R. Bou</i> (1952)	<i>G. Hoz</i> (1953)	<i>Julliard, Chang</i> (1964)	<i>Rivenc, Rojo</i> (1973)
Norma claramente definida	—	—	—	+	+	+
Recuento separado de las flexiones	—		+	—	+	+
Recuento de todos los vocablos	—		—	+	+	+
Índice de clasificación	I = f (F, R)	I = f (R)	I = f (F)	I = f (F)	I = f (F, R)	I = f (F)

**6.—COMPARACION DE LOS RESULTADOS.** Según lo que precede, tres son los diccionarios de frecuencia que nos parecen reunir las condiciones necesarias para que una comparación de sus contenidos sea posible. Se trata de los tres más recientes cronológicamente; el *Vocabulario usual, común y fundamental*, el *Frequency Dictionary of Spanish Words*, y *El Español Fundamental*. En efecto, de los tres restantes, uno, el de Keniston no es un diccionario de léxico frecuente sino de expresiones idiomáticas, y los otros dos, aún reconociendo sus méritos, no nos parecen responder a los mismos conceptos que los tres últimos. Sería, por otra parte, injusto, olvidar que gracias a los primeros trabajos se abrió el camino a los más recientes.

La comparación del contenido de las listas del *Vocabulario usual, común y fundamental* y del *Frequency Dictionary* no plantea graves problemas. Hay, sin embargo, un problema en lo que respecta a las listas de *El Español Fundamental*. En efecto, como sabemos, los resultados de esta encuesta serán publicados próximamente, pero no podemos todavía consultarlos. Pero, como ya lo hemos señalado, podemos conocer parte de estos resultados por medios indirectos. Hace ya algún tiempo, apareció el método *Vida y Diálogos de España*, que constituye la aplicación directa a la pedagogía de los resultados de *El Español Fundamental*. Uno de los fascículos que componen la obra es la *Lista Alfabética de Formas Empleadas*, en la que encontramos la lista de todas las formas empleadas en el método, incluidos los nombres propios. Hay, pues, en esta *Lista Alfabética*, vocablos provinientes de la encuesta sobre la frecuencia en la lengua hablada así como vocablos provinientes de la encuesta sobre la disponibilidad. Por otro lado, la *Lista Alfabética* no contiene todo el vocabulario registrado en la encuesta, sólo una parte, la más urgente, la más necesaria, a juicio de los autores, para un primer nivel de aprendizaje del Español. En resumen, tenemos en esta *Lista Alfabética* las consecuencias en el plano de la didáctica de un determinado concepto de la encuesta sobre el léxico empleado por los españoles. Por otro lado, con los dos vocabularios restantes tenemos el material necesario para, en cada caso, poder sacar nosotros mismos estas consecuencias, estableciendo una lista fundamental. Partiremos, pues, de la hipótesis siguiente: Tenemos por un lado una *Lista Alfabética* en la que encontramos los 1.544 vocablos más urgentes para la



enseñanza del Español, según los autores de *El Español Fundamental*. Tenemos, por otro lado, el *Vocabulario común* de García Hoz, con 1.971 vocablos, de los que podemos escoger los 1.544 de más alta frecuencia, que podemos considerar como los más adaptados a un primer nivel de la enseñanza del Español. Tenemos también los 5.024 vocablos del *Frequency Dictionary*, de los que podemos escoger los 1.544 con más alto índice de uso. Podremos, de esta forma, comparar, por lo menos parcialmente, los resultados de estas tres formas, tan diferentes de concebir una encuesta sobre el léxico.

Naturalmente, somos conscientes del carácter teórico de esta comparación y de sus límites. En efecto, nada nos asegura que un autor de métodos de Español cogería los 1.500 vocablos de uno u otro diccionario, sin añadir algunos y sin quitar otros, para elaborar un método. Esto sería una visión idealista del problema que habría que evitar, pero, no pudiendo hacer otra cosa, pensamos que este análisis dará indicaciones lo bastante significativas como para tener una idea de las consecuencias pedagógicas de estos trabajos.

No haremos una comparación triangular, sino, más bien, una comparación con dos vertientes, de las que el *Frequency Dictionary* será la arista. Esto, sobre todo, por razones cronológicas, y también por razones teóricas. En efecto, hay demasiada diferencia entre la edad del corpus del *Vocabulario común* y la del *Español Fundamental* para que los resultados de la comparación no se vieran influenciados por ella. Esto nos escondería las verdaderas diferencias y convergencias. Por otro lado, hay también demasiada diferencia de concepciones teóricas entre las dos obras. También hay enormes diferencias de concepción entre el *Frequency Dictionary* y *El Español Fundamental*, pero aquí el problema es distinto, la elaboración es más cercana en el tiempo, y, sobre todo, nos interesa ver, precisamente, si un corpus oral y una encuesta sobre la disponibilidad pueden hacer llegar a resultados muy diferentes de una encuesta con un corpus mucho más tradicional, pero con criterios matemáticos mucho más elaborados.

Como el *Vocabulario común* da la lista de los vocablos por orden alfabético, hemos tenido que proceder por tanteo. La eliminación de los vocablos con una frecuencia igual o inferior a 10, arrojó una cifra de 1.808 vocablos, todavía demasiado importante con relación a los 1.544 que sirven de base de comparación en la otra vertiente. La eliminación de los vocablos con una frecuencia igual o inferior a 20, arrojó una cifra de vocablos restantes de 1.388 demasiado baja. Por fin, la eliminación de la lista de los vocablos con frecuencia inferior o igual a 15 dio la cifra de 1.553 vocablos, cifra ya del orden de la de 1.544, obligada por las características de la *Lista Alfabética*, que no nos comunica frecuencias.

El vocablo 1.553 de la lista del *Frequency Dictionary* es "hipótesis" (n.), último cuyo valor de índice de uso es de 18,00. El vocablo siguiente posee un índice de 17,97.

Un primer punteo sobre los artículos comunes en los dos diccionarios da los resultados siguientes: (Porcentaje de artículos comunes) (48).

*Vocabulario común* 78% *Frequency Dictionary*

El porcentaje de vocablos comunes a las dos listas es, pues, bastante importante.

Veamos, con dos letras al azar, en qué consisten estas diferencias de 22 %.

Entre las letras F y G, p.e. los vocablos del *Frequency Dictionary* que no se encuentran entre los 1.553 primeros del *Vocabulario Común* son: falso, fantasía, filósofo, fino, firme, físico, fortuna, franco, fresca, frontera, fuego, fundamental, futuro, gastar, generación, gesto, golpe, gris, gritar, gustar.

Los vocablos de la lista de García Hoz que no se encuentran en la de Julliard son: facilitar, factor, fallecer, fama, familiar, favorable, febrero, felicitación, felicitar, finalidad, finalizar, finalmente, finca, firma, firmar, formación, formal, formidable, fotografía, francamente, frecuente, fundador, fútbol, gana, ganado, gasto, generoso, gestión, gestionar, girar, gobernador, grano, gratitud, grato.

Una primera observación es que la diferencia de contenido entre ambos diccionarios estriba en los diferentes vocablos con fuerte contenido semántico; nombres, verbos, y algunos adverbios y adjetivos. Este es un hecho que se repite en todas las letras, y que, además, es previsible. En efecto, las "palabras gramaticales" se encuentran siempre en las frecuencias y, estadísticamente, se demuestra que las diferencias entre las listas se sitúan en las frecuencias medianas y bajas. (49). En este punto, los resultados coinciden con lo que intuitivamente se hubiera podido esperar. Desgraciadamente, dado que el *Vocabulario común* no especifica la categoría gramatical a que pertenecen los vocablos, y no da tampoco contexto que pudiera darnos la solución, no podemos ir mucho más lejos en el plano lingüístico. Tampoco el *Vocabulario común* separa las diferentes acepciones de un vocablo, lo que sí hace el *Frequency Dictionary*. Esto nos impide analizar las diferencias o convergencias en las estructuras lingüísticas de las dos listas. Este trabajo, quizá más interesante que el meramente estadístico, será realizado en la comparación entre el *Frequency Dictionary* y la *Lista Alfabética*, cuyas características lo hacen posible.

A continuación damos la lista de los vocablos del *Frequency* que no se encuentran entre los 1.533 primeros del *Vocabulario común* con indicación de la frecuencia en esta última lista. En el caso en que un vocablo no figure en ella, se le otorgaría la frecuencia 0. (Entre paréntesis, va la dispersión señalada en el *Frequency Dictionary*).

falso: 15 (0,83)	fuego: 0 (0,80)
fantasía: 15 (0,86)	fundamental: 0 (0,64)
filósofo: 0 (0,55)	futuro: 0 (0,58)
firme: 0 (0,83)	gastar: 0 (0,76)
físico: 15 (0,41)	generación: 0 (0,66)
fortuna: 0 (0,71)	gesto: 0 (0,46)
franco: 0 (0,77)	golpe: 0 (0,68)
fresca: 0 (?) (50)	gris: 14 (0,49)
frontera: 0 (0,82)	gritar: 0 (0,56)
	gustar: 0 (0,39)

Vemos por estos datos numéricos que hay dos tipos de vocablos bien diferenciados que no se encuentran entre los 1.553 primeros del *Vocabulario*. El primer tipo está constituido por los que forman parte del centenar siguiente. Su frecuencia es el límite de la frecuencia de eliminación en el *Vocabulario Común* (nuestra eliminación, naturalmente). Podemos decir que la diferencia proveniente de estos vocablos no es significativa. En efecto, falso, fantasía, físico y gris, 21 % de los vocablos considerados, se puede considerar como si prácticamen-

te fueran comunes a las dos listas. Donde se aprecia mejor los resultados de la diferente técnica de clasificación de los vocablos, es en la otra categoría, la de las unidades que no figuran en el *Vocabulario Común*. Entre estas últimas, algunas han sido eliminadas del *Vocabulario Común* a causa de una repartición irregular, como por ejemplo *gustar*, con un índice de dispersión de 0,39. Otras, al contrario, con una repartición sin duda correcta, han sido eliminadas por tener frecuencias bajas. Este es el caso, por ejemplo, de *frontera*, con una dispersión de 0,82. Como lo demuestra la lista del *Frequency Dictionary*, son todas ellas unidades que por una u otra cosa, frecuencia o repartición, presentan interés bastante como para formar parte del primer millar y medio de una lista fundamental. Gracias al cálculo del índice de uso, esto se realiza en el *Frequency Dictionary*, y, por una limitación demasiado rigurosa de la repartición, no forman parte del *Vocabulario Común*.

Si comparamos los vocablos de la segunda lista obtenida, es decir, los que figuran entre los 1.553 primeros del *Vocabulario Común*, y no figuran en la lista correspondiente del *Frequency Dictionary*, observamos que la gran mayoría de ellos tienen a la vez baja frecuencia y baja dispersión.

Todo esto no hace más que corroborar algo que ya podíamos suponer, con sólo analizar los criterios matemáticos de clasificación, en sí mismos, es decir, que los criterios del *Frequency Dictionary* permiten una clasificación más realista y más fina de los vocablos, y una más perfecta explotación de un corpus, que los criterios del *Vocabulario Fundamental*.

Si ahora comparamos el contenido del *Frequency Dictionary* con el del *Español Fundamental* observamos que las diferencias originadas por la composición del corpus en cada caso saltan a la vista.

Por ejemplo, hay 91 locuciones en la *Lista Alfabética* para 0 en el *Frequency*, aunque esto se deba también a la norma del recuento, y 60 pronombres en la *Lista* para 40 en el *Frequency*. También es notable la diferencia en el número de interjecciones que figuran en ambas listas, 13 en la *Lista* y 0 en el *Frequency*. Estos datos muestran que la *Lista Alfabética* constituye un léxico más fácilmente y más directamente utilizable para la elaboración de métodos de enseñanza del Español que el *Frequency*.

Desde el punto de vista de la estructura del léxico en las dos listas, es interesante comparar los porcentajes de las diferentes categorías gramaticales en ambas listas: (51).

	<b>Lista Alfabética</b>	<b>Frequency Dictionary</b>
Sustantivos	49,2 %	46,3 %
Adjetivos	17,1 %	19,4 %
Verbos	15,1 %	23,1 %
Adverbios	8,2 %	5,7 %
Locuciones	3,1 %	0,0 %
Interjecciones	1,1 %	0,0 %
Preposiciones	1,1 %	0,9 %
Artículos	0,5 %	0,3 %

Conjunciones	0,8 %	1,1 %
Pronombres	3,3 %	2,6 %

Las estructuras de los dos léxicos son, pues, muy vecinas. En efecto, aparte del problema de las locuciones o de las interjecciones, en las demás categorías las diferencias no son significativas. Es esta una constatación interesante si pensamos en las fuentes tan diferentes en uno y otro caso. La diferencia oral/escrito queda reflejada en algunas categorías, como acabamos de ver, pero el hecho de que el *Español Fundamental* obtenga un tercio de su corpus de las encuestas sobre la disponibilidad, no parece quedar reflejado en este nivel estructural. Los porcentajes totales de palabras con fuerte contenido semántico y “palabras gramaticales”, en uno y otro caso, no son bastante diferentes para que se pueda decir que reflejan esta diferencia de fuentes. Por ejemplo, entre sustantivos y verbos obtenemos 72% del contenido de la obra de Jullian y Chang. Incluso, la pequeña diferencia que se observa entre estas cifras va en sentido contrario a lo que podríamos esperar.

Si comparamos el contenido de las dos listas en unidades de estas dos categorías, importantes y significativas, vemos que sólo 40,4% de los sustantivos y verbos contenidos en la *Lista Alfabética* se encuentran también en el *Frequency*. Ahora bien, si vemos separadamente verbos y sustantivos, observamos que 70% de los verbos de la *Lista* están entre los 1.544 primeros vocablos del *Frequency Dictionary*, mientras que para los sustantivos, obtenemos la baja cifra de 36,2%. Si buscáramos estos verbos y sustantivos entre los 5.000 vocablos y algo, repertoriados en el *Frequency*, encontraríamos 90% de los verbos y sólo 71% de los sustantivos. Queda pues demostrado, que las diferencias observadas, en lo que concierne los sustantivos, que provienen solamente de la diferencia de criterios matemáticos de clasificación, sino, en gran parte, de las diferencias en la composición del corpus, sin que podamos precisar todavía si lo más importante es que el corpus sea oral o, si lo más importante, en el origen de esta diferencia, es que un tercio de él esté constituido por palabras “disponibles”. Para resumir, digamos que aquí aparece la diferencia esperada, que no aparecía en el plano de la estructura. De todas formas, a priori, pensamos que la diferencia en el contenido en sustantivos, y no en el contenido verbal, señala la importancia del papel jugado por el corpus disponible, ya que en las encuestas sobre la disponibilidad son sobre todo sustantivos las unidades que se obtienen. Para comprobar estas indicaciones y dar una respuesta definitiva, habría que conocer el origen exacto de todos y cada uno de los sustantivos, lo que no podemos saber antes de la publicación íntegra de los resultados del *Español Fundamental*.

Es interesante enumerar algunos de los vocablos que encontramos en la *Lista Alfabética* y no en el *Frequency Dictionary*. Por ejemplo, el campo léxico (52) de la alimentación, rico y extenso en la *Lista*, es bastante pobre en el *Frequency Dictionary*. En efecto, en la *Lista* encontramos los vocablos siguientes inexistentes en el *Frequency*: bebidas, bocadillo, bombón, bota (de vino), cacharros (de cocina), calamares, caramelo, carnicería, cerveza, cocido, cocinar, cocinera, comilona, conservas, cordero, cubiertos, cuchara, chorizo, churros, ensalada, filete, fresa, frutería, helado, horchata, invitado, jamón, lechuga, mantecado, manteles, mantequilla, mariscos, melocotón, melón, paella, paellera, panadería, pavo, peras, pescadería, pimientos, refresco, remolacha, supermercado, tarta, tenedor, tomate, turrón, uvas, vaca, vainilla.

Es decir, 51 vocablos bastante importantes en la vida de todos los días.

También encontramos en la Lista, cierto número de vocablos como: aduana, aeropuerto, aspirador, azafata, disco, electricista, electrodomésticos, foto, fútbol, garaje, gasolina, gasolinera, kilo, lavadora, litro, moto, nevera, película, piscina, supermercado, taxi, televisión, tocadiscos, tractor, tráfico, vacaciones.

Estos vocablos son lo que se podría llamar “vocablos actualizados”. En efecto, se trata de vocablos que, en su mayor parte, no habrían formado parte de una lista fundamental de 1.500 vocablos hace, por ejemplo, 20 años. Entonces existían, no se trata de neologismos, pero la realidad cotidiana hacía que no fueran seguramente empleados con la frecuencia con que lo son ahora.

Esto pone en evidencia un factor, sin duda importante, de los que diferencian entre ambas listas. Se trata del **envejecimiento del corpus**. En efecto, un problema, y no sin importancia, es la necesidad de renovar el corpus al cabo de períodos de tiempo no muy largos. Naturalmente, esto no resta interés a las encuestas precedentes, pero, si lo que queremos es establecer un vocabulario de base, cuanto más extenso sea este vocabulario, más frecuentemente habrá necesidad de renovarlo. Los mismos autores de recuentos y encuestas son los primeros en sentir esta necesidad: “El mundo del niño, al par que el del adulto, cambia. Los inventos, las artes, las ciencias, la industria y el comercio cada día acuñan nuevas palabras, muchas de las cuales pasan a formar parte del vocabulario del niño. Los recuentos de vocabulario deben revisarse y suplementarse a intervalos de tiempo regulares y no muy largos. . .” (53).

El problema que subsiste, es el de saber de qué forma, concretamente, habría que efectuar esta puesta al día de un vocabulario. No pensamos nosotros que la constitución de un nuevo corpus, completo, sea necesaria. Existen posibilidades matemáticas para, a partir de una pequeña encuesta suplementaria, atribuir un “peso” determinado a cada nuevo vocablo obtenido y, en consecuencia, modificar la lista fundamental.

Por otro lado, si consultamos la lista de vocablos del *Frequency Dictionary* (los 1.544 primeros) que no se encuentran en la *Lista Alfabética*, vemos unidades como: academia, actor, arte, artículo, autor, biblioteca, capítulo, crítica, crítico, cultura, curso, etc. También aquí aparece la diferencia de constitución del corpus.

Una vez más llegamos a la constatación de que la oposición lengua hablada/lengua escrita se manifiesta en el contenido de las listas, pero las consecuencias de la parte “disponible” del corpus en el léxico de la *Lista* son mucho más difíciles de poner en evidencia.

En todo caso, según los resultados de la comparación de las categorías gramaticales entre una y otra lista, contenido en palabras “concretas” y contenido en “palabras gramaticales”, no parece que sea indispensable utilizar la técnica de los Centros de Interés para obtener una estructura aceptable del corpus resultante. El problema fundamental, como decíamos más arriba, estriba en los métodos y criterios matemáticos de explotación del corpus y de clasificación de vocablos.

Esto no impide que, desde el punto de vista de una eventual explotación pedagógica, los resultados de *El Español Fundamental* parezcan más apropiados y más fáciles de explotar.

De todas formas, y a manera de conclusión, diremos que un Vocabulario Fundamental, o una encuesta sobre la lengua hablada o escrita, no son más que herramientas, de una extraordinaria utilidad, pero que hay que saber utilizar con mucho sentido común, y teniendo muy bien en cuenta los fines que nos proponemos alcanzar, pues como decía Irving Lorge: "Cualquier regla para seleccionar y controlar el vocabulario no debe interpretarse en sentido estrictamente mandatorio. Los primeros escritores que utilizaron los recuentos de vocabulario se preciaban de no haber usado ni una sola palabra fuera de ciertas frecuencias estipuladas, por ejemplo, las 1.500 más frecuentes. Esto es una virtud dudosa (. . .). Los recuentos de léxico, sin embargo, no son absolutos ni dicen la última palabra. Su utilidad depende de la habilidad y experiencia de quien los usa". (54).

## NOTAS

- (1) Para estos problemas es fundamental la obra: Ch. MULLER.—*Initiation à la Statistique Lexicale*. París, Larousse. 1968.
- (2) *Ibid.*— p. 133. “. . . nous réserverons le terme de mot aux unités élémentaires, bien distinguées par la typographie et l'écriture, qui constituent le texte, donc aux occurrences d'un vocable, quel qu'il soit.”
- (3) La obra de Keniston a que hacemos alusión no es un recuento de vocabulario, sino un recuento de expresiones idiomáticas, destinado a completar los resultados ya obtenidos en aquella época, sobre el vocabulario frecuente en Español. No nos servirá por lo tanto para comparar los resultados con los de las demás obras estudiadas, pero sí para poder apreciar mejor la evolución de las técnicas y de los conceptos aplicados.
- (4) *El Español Fundamental* no se ha publicado todavía. Su publicación está prevista durante el año 1973. Si podemos, sin embargo, citar este importante trabajo y, en parte, compararlo con los demás realizados, esto se debe, sobre todo, a la amabilidad de nuestro colega Paul Rivenc, profesor de la Universidad de Toulouse y director de su Centro de Lingüística Aplicada, que nos comunicó ciertos datos científicos sobre la elaboración de esta obra. También la consulta de *L'Elaboration du Français Fundamental* y del método *Vida y Diálogos de España* ha hecho posible, por razones que explicaremos más adelante, la inclusión de *El Español Fundamental* en nuestro estudio.
- (5) Se encontrará una buena relación de los trabajos realizados sobre el vocabulario español y su frecuencia, en la página 5 y ss. de RODRIGUEZ BOU.—*op. cit.*
- (6) GARCIA HOZ.—*op. cit.* p. 6.
- (7) JULLIAND, CHANG-RODRIGUEZ.—*op. cit.* p. XXII.
- (8) “Total number of different words found in 1.200.000 running words is 18.331”. BUCHANAN.— *op. cit.* p. 9.
- (9) RODRIGUEZ BOU.—*op. cit.* p. 643.
- (10) A.F. WATTS.—*The language and mental development of children*. London, D.C. Heath. 1946.
- (11) En realidad, hoy en día se piensa que el vocabulario “de base” de un adulto medio (vocabulario fuera de su profesión o especialidad), es bastante menor que la cifra señalada por Watts. Se pueden consultar, a este respecto, los trabajos de Mialaret y Guiraud en Francia, y la obra, ya señalada, *L'Elaboration du Français Fundamental*. El impulso definitivo a estos estudios ha sido dado últimamente por las necesidades urgentes de la Lingüística aplicada a la enseñanza de las lenguas extranjeras. Por ejemplo, hoy se piensa que un léxico de unos 1.500 vocablos es suficiente para un sólido primer nivel en una lengua extranjera. Estos 1.500 vocablos, bien escogidos, pueden asegurar una comprensión superior al 90% de cualquier texto no especializado. Aunque con ello no queramos decir que el vocabulario de un adulto en su propia lengua sea de este orden, tampoco debe de llegar a cifras del orden de la señalada por Watts. Sobre la aplicación actual de estas nociones a la enseñanza del Español como lengua extranjera, se puede consultar nuestro trabajo: *Etude comparée de cinq méthodes modernes pour l'Enseignement de l'Espagnol*, chap. 3, “Le Lexique”. Existen ejemplares (policopiados) en varias bibliotecas universitarias francesas (entre ellas, Burdeos, Estrasburgo y Besançon), y en la Biblioteca Hispánica del I.C.H., de Madrid.
- (12) GARCIA HOZ.—*op. cit.* p. 25.
- (13) JULLIAND, CHANG-RODRIGUEZ.—*op. cit.* p. XIII.
- (14) JULLIAND, CHANG-RODRIGUEZ.—*op. cit.* p. XXII. “After examining the alternatives, we decided to establish a universe of 500.000 words, which is somewhat larger than the 400.000 word sample underlying García Hoz, *Vocabulario*, and theoretically smaller than the 1.200.000 words underlying Buchanan's *Word Book* and the 7.200.000 words counted in Rodríguez-Bou's *Recuento*.”

- (15) *Ibid.* p. XXV.
- (16) *Ibid.* p. XXIII: "A basic vocabulary can be conceived continuously or discontinuously, as a continuum of decreasing units or as a discontinuum set of statistical classes. . .".
- (17) BUCHANAN.—*op. cit.* p. 8.
- (18) *Ibid.*—p. 7.
- (19) Con esto no queremos decir que Buchanan no haya dado importancia a la lengua "de todos los días". En efecto, en su prólogo, página 7, dice: "The speech of every day life is not represented directly, but appears in modified form in novels, plays and newspapers". Buchanan no excluye, pues, la lengua corriente, pero considera que es un aspecto más de la lengua, junto con otros niveles. Sería también injusto olvidar cuáles son, en la época de Buchanan, los medios materiales de estudiar la lengua hablada, de todos los días. Estos son, en comparación con los actuales, prácticamente inexistentes.
- (20) Pudiera parecer paradójico clasificar a los programas de radio como lengua escrita, pero en realidad no lo es ya que, en este caso, se trata de guiones preparados con antelación, charlas, etc. que, por lo tanto, no tienen las características de la expresión oral espontánea, aunque, desde un punto de vista formal, se utilice un "canal" propio de la expresión oral.
- (21) RODRIGUEZ BOU.—*op. cit.* p. 8.
- (22) GARCIA HOZ.—*op. cit.* p. 16.
- (23) Está perfectamente justificado el clasificar a las fuentes estatales. . . etc. en la misma categoría, no literaria, que los periódicos, pues la correlación entre ambos vocabularios es, según el mismo García Hoz, de 0,925, la más alta de todas las establecidas por él entre las cuatro partes del corpus. (*Ibid.* p. 455). El análisis de todas estas correlaciones demuestra también que el vocabulario de las cartas pertenece a un registro muy diferente de los de las fuentes escritas clásicas. Sus correlaciones con las otras tres categorías son, en efecto, las más bajas. Estas correlaciones, pensamos, justifican nuestro análisis del corpus estudiado.
- (24) García Hoz se basa para estas afirmaciones, como él mismo señala, en: J. CASARES.—*Nuevo concepto del Diccionario de la Lengua*. (Discurso de recepción en la R.A.E.). Madrid, R.A.E. 1921. pp. 35, 36.
- (25) GARCIA HOZ.—*op. cit.* p. 16.
- (26) *Ibid.*—p. 16. El subrayado es del autor de la citación.
- (27) JULLIAND, CHANG-RODRIGUEZ.—*op. cit.* p. XIV.
- (28) ROJO SASTRE.—*Los vocabularios de base. Sobre la elaboración y aplicaciones de El Español Fundamental*. Ponencia presentada por el autor en el II Congreso Internacional para la Enseñanza del Español. Madrid. 1970. El texto, policopiado, fue distribuido por la Secretaría del Congreso. I.C.H. Madrid.
- (29) RIVENC, ROJO-SASTRE.—*Vida y Diálogos de España*. París, Didier. 1970.
- (30) GOUGENHEIM, MICHEA, RIVENC, SAUVAGEOT.—*L'Elaboration du Français Fondamental*, París, Didier, 1964.
- (31) Este porcentaje se indica con toda reserva. Ver, a este respecto, el problema de las "cartas familiares", del *Vocabulario usual*. . ., en el párrafo 4.2.
- (32) Ver el libro, ya citado, *Initiation a la Statistique Linguistique*, sobre todo, los capítulos 2, 3, 4 y 5 de la segunda parte.
- (33) El corpus de Keniston es, como el de Buchanan, muy tradicional. De las 100 unidades de 10.000 pa-



labras cada una, sólo 12 pertenecen a las publicaciones periódicas y el resto fuentes de tipo únicamente literario o técnico.

- (34) KENISTON.—*op. cit.* p. 2.
- (35) RODRIGUEZ BOU.—*op. cit.* p. 644.
- (36) *Ibid.*—p. 644.
- (37) RODRIGUEZ BOU.—*op. cit.* p. 645.
- (38) GARCIA HOZ.—*op. cit.* p. 38.
- (39) Esto es lo que se deduce del examen de la *Lista Alfabética* del método *Vida y Diálogos de España*.
- (40) BUCHANAN.—*op. cit.* p. 10.
- (41) BUCHANAN.—*op. cit.* p. 8.
- (42) KENISTON.—*op. cit.* p. 10.
- (43) RODRIGUEZ BOU.—*op. cit.* p. 649.
- (44) GARCIA HOZ.—*op. cit.* p. 385. El subrayado es del autor de la citación.
- (45) La variación viene dada por la fórmula  $\sigma^2 = \sum (F_i - \bar{F})^2 / 5$
- (46)  $D = 1 - \frac{\sigma / f}{2}$
- (47) Un interesante artículo, cuya lectura recomendamos a los lectores interesados por los criterios matemáticos del *Frequency Dictionary* es: Ch. MULLER.—*Fréquence, dispersion et usage*. in "Cahiers de Lexicologie". vol. 7. pp. 41 y ss. París, Didier-Larousse. 1965.
- (48) La diferencia de norma lexicográfica entre los tres diccionarios implica acomodaciones que no tiene interés explicar aquí y que hacen que, por ejemplo, haya que contar más de 1.533 artículos del *Frequency Dictionary* para obtener el equivalente de 1.533 vocablos de García Hoz.
- (49) A este propósito, ver el artículo: H. ANDRIES.—*Quelques considérations sur l'Enseignement du Français, deuxième langue*. in "Etudes de Linguistique Appliquée". vol. n.º 3. París, Didier, 1964.
- (50) En este caso, no podemos dar la dispersión, ya que en el *Frequency Dictionary* figuran dos "fresca", un nombre y un adjetivo. No podemos saber si en el *Vocabulario Común* es uno u otro, o los dos.
- (51) Aquí se trata de los resultados obtenidos por nosotros, válidos para los primeros 1.544 vocablos del *Frequency Dictionary*, por orden de "índice de uso". Para cálculos estadísticos completos sobre este diccionario, consúltese el interesante artículo: H.N. URRUTIBEHEITY.—*The statistical properties of the Spanish lexicon*, in "Cahiers de Lexicologie". vol. n.º 20. París, Didier/Larousse. 1972.
- (52) Utilizamos las nociones de "campo léxico" y de "campo semántico" adoptadas en la obra: GENOUVRIER, PEYTARD.—*Linguistique et Enseignement du Français*. París, Larousse. 1967. p. 206.
- (53) RODRIGUEZ BOU.—*op. cit.* p. 14.
- (54) I. LORGE, en el prefacio del diccionario de RODRIGUEZ BOU.—*op. cit.* pp. XV y XVI.

