

Listas de frecuencias de palabras: una revisión de la literatura en español y de sus posibles usos en investigación

Nelson Rodríguez Trujillo *

Resumen

En este artículo se hace una revisión de las listas de frecuencias de palabras existentes en español y se plantean algunas posibilidades de su utilización en el contexto de la investigación educacional, psicológica y lingüística. Una lista de frecuencia de palabras es básicamente un listado de todas las palabras diferentes encontradas en una muestra de material escrito u oral con sus respectivas frecuencias de ocurrencia.

Las listas de frecuencia de palabras en español se revisaron en función de algunas condiciones “ideales” que debe poseer este tipo de listado. Se concluye que los mejores estudios en español son el “Recuento de Vocabulario Español”, realizado por la Universidad de Puerto Rico, y el “Vocabulario usual, común y fundamental” de García Hoz. Se considera que esos dos trabajos pueden ser utilizados en investigaciones aplicadas a Venezuela, mientras se desarrollan listas basadas en el español venezolano. Actualmente, gracias a las computadoras digitales, la elaboración de listados de frecuencia de palabras se ha hecho más económico y factible que en el pasado, cuando la mayor parte del trabajo debía realizarse manualmente.

1. Introducción

Una observación obvia, pero al mismo tiempo de gran importancia, es que toda lengua emplea palabras como componentes básicos para la construcción de estructuras más complejas tales como oraciones y párrafos. También puede observarse que algunas palabras ocurren más frecuentemente que otras. Esto sugiere que la frecuencia de ocurrencia de una palabra en la lengua puede ser usada como indicador de su importancia. Ese principio es utilizado precisamente en las listas de frecuencia de palabras.

Una lista de frecuencia de palabras consiste básicamente en una **lista ordenada, ya sea alfabéticamente o de acuerdo con la frecuencia de ocurrencia de las palabras, o sobre la base de cualquier otro principio, que registra la palabra con sus variantes** (conjugaciones, afijaciones o declinaciones) y la frecuencia de ocurrencia en la muestra de materiales utilizados para desarrollar la lista.

Dado que esas listas registran la importancia relativa de las palabras, pueden ser empleadas en la selección del vocabulario necesario para dominar una lengua extranjera; en el desarrollo de materiales educativos, tales como libros de lectura; y en general, en todas aquellas actividades que exijan un conocimiento sistemático del vocabulario.

* Psicólogo, especialista en lectura. Dirige el Departamento de Investigaciones del Banco del Libro, en Venezuela.

El desarrollo de estas listas parece iniciarse muy temprano; data quizá de los tiempos helénicos (Monteith, 1976; Bongers, 1941).

Hoy en día, se encuentran en la mayoría de los idiomas occidentales. Aparentemente, el inglés es el idioma en que se ha hecho el mayor número de estudios y de aplicaciones prácticas (Fries y Travers, 1940). Con respecto al español, el autor pudo localizar nueve trabajos realizados en diferentes países en un período que se extiende de 1920 a 1953.

Es curioso, sin embargo, que no se localizaron referencias de aplicaciones prácticas de los listados en español, a pesar de los múltiples usos que se les ha dado en otros idiomas. Esto puede deberse a que nuestros profesionales, especialmente en el campo educativo, lingüístico y psicológico, desconozcan este tipo de trabajo. El propósito de este artículo es presentar una reseña de las listas de frecuencias de palabras localizadas en español y discutir su posible uso en diversos campos del quehacer científico, como una contribución a la divulgación de su utilidad.

II. Elaboración de listas de frecuencia de palabras

En la elaboración de listas de frecuencia de palabras se sigue el siguiente proceso:

1. Se selecciona el medio sobre el cual se desea determinar la frecuencia con que ocurren las palabras, es decir, se define el Universo. El Universo puede ser la comunicación oral o escrita, y dentro de cada uno de ellos se puede trabajar con distintos medios de difusión. Por ejemplo, para la comunicación escrita se pueden emplear libros de textos, periódicos, cartas personales, etcétera.
2. Se define lo que se va a entender como una palabra. Este paso es importante, porque aun cuando tipográficamente puede definirse como palabra todo grupo de letras contenido entre dos espacios en blanco, conceptualmente puede haber discusión sobre si consideramos, por ejemplo, a los elementos "niños" y "niño" como una o dos palabras. Este problema se presenta cada vez que encontramos un morfema y sus afijaciones.
3. Se selecciona una muestra de materiales del Universo. La representatividad de la muestra debe ser cuidadosamente garantizada.
4. Se realiza el conteo de las palabras. Cada vez que se encuentra una palabra nueva, debe abrirse una entrada. Cuando se encuentra una palabra ya registrada, se le da un crédito por cada ocurrencia. Hasta hace poco, el conteo de palabras se realizaba manualmente, lo cual era, no solamente un trabajo tedioso y sujeto a innumerables fuentes de error, sino también muy costoso. La aparición de las computadoras digitales facilita el trabajo rutinario de conteo de palabras. Ahora puede darse mayor énfasis a la tarea de planificación de los estudios, y a la interpretación de los resultados. Más aún, la alta velocidad, fácil rutinización y bajo costo del trabajo de computación, permite hacer listados de palabras basados en muestras de materiales de mayor tamaño y con mayor cobertura, así como la realización más frecuente de

estudios de actualización de las listas, y el desarrollo de listas de palabras en campos especializados.

III. Características de las listas de frecuencia de palabras

En la elaboración de una lista de frecuencia de palabras existen muchas posibilidades de muestreo de los materiales a utilizar para su desarrollo, de procesamiento y de presentación de los resultados. Idealmente las listas deben poseer ciertas características para facilitar su uso y maximizar su utilidad.

1. Las listas de frecuencia de palabras deben indicar la ocurrencia real de las palabras registradas y no recurrir a otros indicadores. Esta característica es importante, debido a la práctica usual en algunas listas de emplear letras como indicadores de la importancia. Por ejemplo, en la lista de Thorndike de 20.000 palabras (Thorndike, 1932), una palabra "AA" es más importante que una palabra "A" o "B". Esto limita la utilidad de las listas, al no permitir comparaciones entre diferentes estudios.
2. No deben excluir las palabras de mayor uso o las de ciertas categorías gramaticales. Es común en muchas listas excluir las palabras "funcionales" (artículo, preposiciones, conjunciones y verbos copulativos), que son precisamente las que ocurren con mayor frecuencia, e incluir solamente los verbos, sustantivos, adjetivos y algunos adverbios. Esta práctica mutila las listas innecesariamente y distorsiona la imagen que se obtiene de la frecuencia de ocurrencia del vocabulario.
3. Deben elaborarse sobre la base de muestras grandes de lenguaje, ya que a medida que aumenta el tamaño de las muestras, las frecuencias tienden a estabilizarse. El tamaño de la muestra puede definirse de dos maneras distintas. Por un lado, por el número de materiales diferentes que se incluyen. Por el otro, por la longitud de la muestra de material, es decir el número de palabras que incluye. Es ideal cuando se incluyen muchas muestras de materiales diferentes y de muchas palabras.
4. Deben incluir un índice de amplitud, es decir el número de fuentes diferentes en que se encuentra la palabra. No es lo mismo que una palabra aparezca diez veces en un texto, a que aparezca diez veces en diez textos distintos.
5. El ordenamiento de la lista debe ser lo más flexible posible, y la presentación, no solamente por orden alfabético, sino también por orden de frecuencia. Este doble ordenamiento permite su uso para determinar cuáles palabras son las más importantes, y también para determinar la importancia relativa de una palabra en particular. Para el primer uso es necesario un ordenamiento por la frecuencia de ocurrencia; para el segundo, un ordenamiento alfabético.
6. De ser posible, la lista debe incluir no solamente la palabra y su frecuencia de ocurrencia, sino también su clase gramatical y el contexto en que ocurrió. Esto permite determinar los significados en el caso de homónimos. A esta modalidad se le da el nombre de listado semántico.
7. Las muestras de lenguaje empleadas para desarrollar la lista deben estar relacionadas con el uso al cual se destina. Por ejemplo, si el uso es general, como cuando se desea conocer cuáles son las palabras utilizadas más frecuentemente en el vocabulario escrito, las muestras

deben provenir de muchas fuentes diferentes; si se intenta describir el vocabulario escrito de los niños de primaria, las fuentes serán naturalmente de ese origen.

Las características descritas anteriormente son condiciones ideales y no requisitos prescriptos. Debe pensarse que, mientras más estrictas sean las condiciones que se exijan a las listas, más elevados serán los costos de la investigación para producirlos. Más aún algunas características son ideales para un uso específico, pero no para otro.

A continuación se presenta la discusión de las listas de palabras encontradas en español, haciendo uso de esas siete características.

IV. Revisión de la literatura en español

La referencia más antigua encontrada es el trabajo de H. Keniston, **A Basic List of Spanish Words and Idioms** (Lista básica de palabras y expresiones idiomáticas españolas), publicada en 1920 en la Universidad de Chicago. Esta lista fue elaborada con la intención de ayudar a profesores de español a seleccionar el vocabulario más útil en la enseñanza de literatura española y en cursos de conversación. Se basó en una muestra de obras dramáticas y novelas, e incluye un total de 1322 palabras diferentes ordenadas en ocho listas. En la primera se incluyen las palabras comunes a un 80 por ciento de los textos revisados; se va disminuyendo de una lista a otra en el porcentaje de las fuentes comunes, hasta que en la última se incluyen las palabras comunes al 33 por ciento de los textos. Desafortunadamente, Keniston no señala cuáles obras empleó ni el tamaño de la muestra.

Tampoco incluye las frecuencias absolutas de cada palabra; su trabajo refleja solamente la amplitud. Una ventaja es la clasificación de las palabras por clases gramaticales, característica poco común en trabajos de este tipo.

La segunda referencia data de 1924, cuando Jamieson publicó **A Standardized Vocabulary for elementary Spanish** (Un vocabulario normalizado para el español elemental). Este trabajo está dirigido también a ayudar a los profesores de español y es un intento de estandarizar el vocabulario básico a enseñarle a los estudiantes. La autora encontró que 4488 palabras diferentes incluidas en diez libros básicos de enseñanza del español, solamente 249 eran comunes a todos. Jamieson preparó una lista final de 1093 palabras comunes a cinco o más textos, considerando que ése era un núcleo central que debía ser aprendido y ejercitado por los estudiantes como base para la comunicación en español. El artículo de Jamieson indica los libros utilizados como muestra y la amplitud de fuentes en que aparece cada palabra, no da información sobre el tamaño de la muestra ni la frecuencia absoluta de cada palabra. Más aún, su muestra está restringida a textos de gramática española en la enseñanza de la lengua.

En 1925, Cartwright publicó su trabajo **A Study of the Vocabularies of Eleven Spanish Grammars and Fifteen Spanish Reading Texts** (Estudio del vocabulario de once gramáticas españolas y quince libros de lectura). Tal como su nombre lo indica, el estudio se basó en once libros de

gramática y quince de lectura española. El autor encontró 12.916 palabras diferentes y seleccionó 1684 que aparecían en diez o más libros, dando de esta manera una indicación de la amplitud de fuentes en que aparecen las palabras. Excluyó de su lista cifras; nombres de estaciones climáticas; días y meses del año; pronombres interrogativos, relativos, posesivos y personales; adjetivos demostrativos y artículos. Cartwright no especifica el propósito de esta lista, pero por la naturaleza de la muestra, parece obvio su propósito didáctico. La clasificación de las palabras da una idea de su importancia, ya que indica cuando su aparición es común a 10, 11 y así sucesivamente hasta llegar a 26 libros. Sin embargo, dada la fragmentación de la lista, la falta de información en cuanto a la frecuencia absoluta de cada palabra, y a la exclusión de cierto tipo de palabras, este trabajo es de poca utilidad.

En 1927, Meras y Roth publicaron su **Pequeño Vocabulario** (en español en la versión original), el cual lleva como subtítulo: “Una lista de 2000 palabras españolas ordenadas en grupos lógicos para la construcción de oraciones en los primeros dos años” (de aprendizaje del idioma). Este trabajo no es una lista de frecuencia de palabras en el sentido estricto, porque las palabras fueron seleccionadas según el juicio subjetivo de sus autores.

Sin embargo, esta lista refleja la importancia de algunas palabras en español.

Céspedes publicó en 1929 lo que parece ser el primer estudio de frecuencia de palabras en español realizado en un país de habla hispana. Lleva por título **Investigaciones acerca de las palabras usadas en castellano en Panamá**. Céspedes revisó medio millón de palabras tomadas de libros y periódicos. Encontró 9255 palabras diferentes, las cuales ordenó de acuerdo con su frecuencia absoluta. Esto fue un trabajo admirable, pues lo realizó prácticamente solo y en un tiempo escasamente superior a un año. La única falla que podría señalarse, falla fácilmente explicable dadas las condiciones adversas en que trabajó, es el no haber reportado la amplitud de fuentes en que aparece cada palabra.

En 1929, Buchanan publicó su libro **A Graded Spanish Word Book** (Un libro graduado de palabras españolas) con la intención de que fuera utilizado como material básico para el desarrollo de escalas (tests) de vocabulario. Como propósito secundario, se esperaba que sirviera también para usos pedagógicos, literarios y psicológicos. Se revisaron más de 1.200.000 palabras, lo que arrojó un total de 18.331 palabras diferentes. La muestra fue tomada de cuarenta tipos distintos de materiales, incluyendo un 75 por ciento de literatura y el 25 por ciento restante de periódicos, literatura técnica y folklore. Para facilitar el recuento, que fue realizado manualmente, se eliminaron las palabras funcionales. El ordenamiento de la lista viene dado de acuerdo con frecuencias absolutas y por orden alfabético, además, se incluye un índice de amplitud y un índice de mérito, el cual fue desarrollado al dividir la frecuencia absoluta por diez y sumarle la amplitud.

Sin duda éste es uno de los trabajos más completos que existen en español y puede compararse fácilmente con la lista de frecuencia de palabras en inglés de Thorndike (1921), que era el estudio más exhaustivo para la

época en que se realizó y que influyó profundamente el trabajo de Buchanan. En 1952, la Universidad de Puerto Rico publicó **Recuento del vocabulario español**. Este trabajo, realizado con la intención de sustituir el uso de la lista de Buchanan, contiene no solamente una muestra de publicaciones contemporáneas, sino también muestras de material escrito y oral de niños y adultos, programas de radio y la misma lista de Buchanan. El estudio se basó en más de siete millones de palabras y se obtuvo un total de 20.542 palabras diferentes. La presentación es de gran utilidad ya que incluye una lista por orden alfabético, y otra por orden de frecuencia. En volumen aparte se incluyen todas las palabras con sus frecuencias absolutas en cada una de las diferentes fuentes de donde surgieron; hay una columna para la frecuencia de palabras en programas de radio, otra para la comunicación oral y escrita de niños, otra para los adultos, y así sucesivamente, incluyendo también un gran total.

Irving Lorge, asesor de este proyecto, afirma en el prólogo que esta lista... “va más allá de otros recuentos publicados previamente en idioma alguno”.

Finalmente, en 1953, Víctor García Hoz publicó en España **Vocabulario usual, común y fundamental**, el cual constituye un esfuerzo por determinar el vocabulario mínimo para expresarse en español. El autor utilizó una muestra de medio millón de palabras de cuatro medios de comunicación escrita: cartas personales, documentos públicos, periódicos y libros. Encontró 12.913 palabras diferentes. Su interés es netamente pedagógico y psicológico, ya que trata de determinar cuál es el “núcleo” de la lengua. Para ello usa métodos de análisis factorial, un procedimiento sin precedentes en estudios de esta naturaleza, para determinar que además de un factor de vocabulario general a los cuatro medios de comunicación escogidos en su muestra, existen cuatro factores específicos, uno correspondiente a cada medio de difusión. El trabajo es presentado en forma de tres listas: un vocabulario usual, con aproximadamente 13.000 palabras; un vocabulario común, con 2000 palabras; y un vocabulario fundamental, con cerca de 200 palabras. Cada lista tiene la frecuencia absoluta de ocurrencia, pero no la amplitud. La ventaja de este estudio es la muestra bien definida de materiales incluidos, pero tiene el inconveniente de hacer solamente una presentación por orden alfabético.

De este recuento de la literatura española, puede concluirse que existen por lo menos dos listas realizadas con gran rigor científico en español; ellas son: **Recuento de vocabulario español** de la Universidad de Puerto Rico y **Vocabulario usual, común y fundamental** de García Hoz. Estos trabajos pueden ser empleados como base para estudios más profundos del vocabulario español.

Una de las características de las listas de frecuencia de palabras es su caducidad. El dinamismo de la lengua hace que palabras usuales en un momento dado, sean arcaísmos o desaparezcan en pocos años; también existen variaciones significativas, en términos de usos, de una sociedad a otra. Más aún, las frecuencias de utilización de las palabras fluctúan mucho de un estudio a otro, salvo para las 500 palabras de mayor ocurrencia que generalmente mantienen su posición. Dadas esas características de las listas,

su caducidad y la inestabilidad de las frecuencias, es recomendable la realización de un estudio del vocabulario usado en Venezuela. Este estudio es posible hoy en día dada la velocidad y fácil rutinización de las computadoras digitales.

V. Posibles usos de las listas de frecuencia de palabras

Una lista de frecuencia de palabras es esencialmente un intento por describir en alguna medida la composición y la estructura de esa entidad abstracta e infinita que es el vocabulario, utilizando una muestra del mismo. Esta tarea es, naturalmente, difícil. Si bien la mayoría de las personas son capaces de determinar cuáles palabras de un texto o una comunicación oral son comunes y cuáles son de ocurrencia infrecuente, existe un gran número de términos y sus correspondientes conceptos, cuya importancia dentro del contexto total del vocabulario es difícil de discernir. En este sentido, pueden ser de ayuda las listas. El problema está en que aún los estudios de este tipo más sofisticados, llevan en sí elementos de arbitrariedad.

Cuando se piensa sobre el vocabulario debe recordarse que los morfemas no son más que las formas externas de los conceptos, y que éstos a su vez, son entidades mentales que se desarrollan a través de procesos psicológicos complejos. Esto lleva a que los listados de frecuencia de palabras sean instrumentos demasiado primitivos para describir a ciencia cierta lo que ocurre en el proceso de comunicación, y menos aún los procesos mentales que acompañan al intercambio de información.

Considérense solamente los siguientes aspectos para entender cuán arbitrarios pueden ser los listados de frecuencia de palabras como descriptores del vocabulario. Por un lado, el vocabulario puede ser considerado como infinito ya que constantemente se expande en algunas áreas mientras se contrae en otras; un listado de palabras refleja, naturalmente, sólo lo que está contenido en la muestra y es por ello un instrumento sujeto a caducidad. Por otro lado, hay más de una manera de definir los elementos de vocabulario, ya que cualquier morfema con sus afijaciones puede implicar variaciones enormes en el significado de una palabra. Finalmente, el vocabulario está íntimamente ligado con su uso en situaciones específicas y dentro de un contexto gramatical.

Al tratar de describir ese dinámico proceso que es la comunicación a través de listados, es obvio que se tienen que emplear supuestos que a veces son de difícil aceptación. Por ejemplo, la frecuencia de ocurrencia de un término determinado puede utilizarse como un indicador de su importancia dentro de la lengua. Sin embargo, muchas palabras frecuentes son semánticamente “vacías” como son las palabras funcionales. Otras en cambio, a pesar de su escasa ocurrencia, resumen en sí conceptos de gran abstracción y valor cognoscitivo, emocional o social. A pesar de estos problemas, los listados de frecuencia de palabras constituyen instrumentos que nos permiten conocer algo más acerca de los términos en que se realiza la comunicación.

Tres áreas sirven para ilustrar la utilidad de las listas: educación, psicología y lingüística. En cada una de ellas, se hará una breve discusión los posibles usos.

1. Usos educacionales

Las primeras listas de frecuencia de palabras fueron desarrolladas para el uso en la enseñanza de lenguas nativas y extranjeras (Bongers, 1974). Su utilidad en este contexto consistía en que permitían seleccionar las palabras que debían enseñarse primero y aquellas que podían enseñarse posteriormente. Se seleccionaban para enseñar primero aquellas palabras que ocurrían más frecuentemente en el tipo de material que confrontaría el estudiante en sus lecturas y conversaciones. Basado en esto se desarrollaron programas de enseñanza de idiomas, que definen un vocabulario mínimo y un conjunto de reglas gramaticales básicas. Ejemplo de estos programas es el Inglés Básico que emplea solamente 800 palabras pero que permite expresarse correctamente.

Esta selección de términos más usuales ha ganado importancia desde los trabajos de Gagné sobre **Jerarquías de aprendizaje** (1970) y de Bloom sobre **Dominios de aprendizaje** (1972), en los cuales se enfatiza la importancia de “escalonar” la instrucción de lo más simple a lo más complejo hasta alcanzar un criterio mínimo de rendimiento. En este sentido, es importante comprender la distinción entre los conceptos de contenido normativo y contenido prescriptivo. El primero se refiere a lo que una persona conoce en relación con el grupo con el cual se le compara. El segundo se refiere a lo que una persona debe poder hacer como resultado de un proceso de enseñanza. En Venezuela el sistema de evaluación y de promoción escolar pretende basarse en un contenido prescriptivo definido en los objetivos específicos de los programas de estudios. Sin embargo, los programas no definen claramente cuál es el contenido mínimo que deben dominar los alumnos al terminar cada año escolar.

Tampoco definen si existe una jerarquía entre los conceptos que se deben aprender. Esto lleva a que el maestro utilice el programa como una guía que se sigue literalmente, y que quede a su arbitrio definir como mejor puede cuándo un alumno debe ser promovido al grado inmediato superior.

Los listados podrían ser útiles en la definición de una jerarquía de dificultad de los contenidos y de los criterios mínimos de rendimiento. Esto puede hacerse no sólo en la enseñanza de la lectura, sino también en materias especializadas como las ciencias sociales y naturales, y las matemáticas. En la enseñanza del vocabulario en cualquiera de esas materias, se requeriría establecer cuáles son los conceptos usuales en cada campo, establecer cuáles conceptos presuponen el conocimiento de otros, y en qué forma se interrelacionan. Esto definiría una jerarquía de aprendizaje de los conceptos que podrían emplearse para establecer un orden de enseñanza y niveles mínimos de rendimiento. El conocer cuánto saben los alumnos al terminar un determinado grado, y cuánto deben aprender en el grado superior implica, naturalmente, dos tipos de universos a definir: uno es el vocabulario normativo, es decir, lo que conocen los niños a su arribo a un grado

determinado; el otro, el vocabulario prescriptivo, usado en la descripción de los conceptos a aprender. Se puede describir el universo de los conceptos que el alumno debe saber antes de exponerlo al universo de los conceptos que aprenderá.

En este contexto, listados especializados en cada campo ayudarían a definir los conceptos a enseñar, ya que al fin y al cabo, las palabras empleadas en los libros de texto y en clase son precisamente los estímulos que evocarán en los alumnos los conceptos que ellos representan.

2. Usos en el contexto psicológico

Un uso importante en el Contexto psicológico de los listados es el desarrollo de tests de vocabulario. Thorndike (1923) utilizó la versión de 1921 del **Teacher's Word Book** para la selección de las palabras a incluir en el test. Burt (1921) basó su **Prueba graduada de vocabulario** en un listado de 2.000 palabras diferentes utilizadas por niños en composiciones escritas. Ambos autores hicieron dos suposiciones: 1) consideraron a los listados como descripciones del vocabulario empleado en la comunicación; 2) utilizaron la frecuencia relativa de las palabras en los listados como índice de su dificultad, ya que a mayor frecuencia, se supone que hay mayor familiaridad con los conceptos que las palabras representan. En sus pruebas, las preguntas fueron ordenadas de acuerdo con la posible dificultad que tendrían empleando como indicador la frecuencia.

Nótese que si bien se hacen supuestos bastante fuertes en la elaboración de las pruebas, este procedimiento es menos arbitrario que la práctica, común en el medio hispanoamericano, de adaptar pruebas de vocabulario tales como las de Wechsler (Wais, Wisc y Wippsi) a partir de simples traducciones de los términos empleados en las pruebas de inglés.

Cuando se elaboran pruebas de vocabulario, sería ideal tener estudios normativos del vocabulario a diferentes niveles de edad, de las cuales se podrían "muestrear" las palabras para elaborar las preguntas de los tests. Las normas por edad, a pesar de que no pueden considerarse descripciones formales por presentar una amplia variación de un estudio a otro (Fries y Travers, 1946; p. 42) dan una idea del desarrollo psico-lingüístico de los niños, y pueden servir como descriptores de su vocabulario y familiaridad con los conceptos.

Otra aplicación interesante de los listados de palabras en psicología, se da en la determinación de las redes de asociaciones semánticas y de la estructura de la memoria. En la mayoría de los estudios de asociaciones de palabras, se ha encontrado que los sujetos responden a la palabra estímulo con palabras que presentan una alta frecuencia en los listados (Osgood, 1953). Aun cuando no puede afirmarse que exista isomorfismo entre la memoria semántica y los listados de frecuencia de palabras, tal parece que la gente tuviese ciertas palabras "a mano" para ser empleados al surgir la ocasión, es decir, son sus "palabras preferidas". Zipf (1935) utilizó los listados en sus especulaciones sobre la biología del lenguaje. Él considera que las palabras más usadas tienden a hacerse más cortas, al igual que los movimientos físicos que se

practican mucho tienden a hacerse cada vez menos complicados, y a realizarse con mayor facilidad. Aparentemente las palabras “preferidas” de un grupo de personas, son precisamente, aquellas más usuales, y quizá tienen un mayor número de asociaciones entre ellas. Esto sugiere que los estímulos verbales tienden a desencadenar un conjunto similar de asociaciones entre los miembros de un grupo social; un resultado muy interesante, por cierto, para los propagandistas y manipuladores de opinión pública.

3. Usos en la investigación lingüística

Supongamos que se desea elaborar un diccionario para niños de primaria, ¿qué palabras deben incluirse? Una posibilidad es incluir las palabras más frecuentes en el tipo de literatura a que va a ser expuesto el niño de primaria. Lo ideal sería, tal y como sugieren en el **Word Frequency Book**, Carroll, Davies y Richman (1972), tener listas de diferentes fuentes: materiales escritos; las comunicaciones orales a que está expuesto el niño en la escuela, en la calle y en su hogar; el vocabulario que el niño utiliza y escribe. En esta forma estaríamos seguros de haber incluido no sólo aquellas palabras que él oír y escuchará, sino también las palabras que son importantes para los niños. Esto permitirá la elaboración de diccionarios realmente útiles y que incluyan tanto las palabras que comúnmente se emplean en el contexto social adulto que rodea a los niños, es decir, tanto el vocabulario con que el adulto habla con los niños, como las palabras con que el niño responde a ese contexto y las emplea para comunicarse con otros niños. Más aún, en el campo lexicográfico, las listas permitirían la determinación de un vocabulario común entre los diferentes grupos sociales y la importancia relativa de cada palabra dentro de cada grupo.

En la lingüística teórica hay también aplicaciones para las listas de frecuencia de palabras. Por ejemplo, Roberts (1965) realizó un estudio para determinar cuáles son los sonidos de mayor ocurrencia en el inglés americano, así como las secuencias de sonidos que más ocurren; trató de determinar la “sintaxis” fonética. Roberts empleó una lista de frecuencia de palabras (Horn, **Basic Writing Vocabulary**, 1962); les pidió a diferentes informantes que hicieran oraciones con esas palabras y grabó las respuestas de los informantes. En esa forma generó un nuevo listado de frecuencia de palabras y un listado de sonidos y grupos de sonidos. Su estudio corroboró las predicciones de Zipf de que las palabras más usuales son también más cortas, y que las palabras que ocurren más frecuentemente son también las que tienen mayor entropía, es decir son más fáciles de predecir cuando hay ruido en la comunicación. Estos resultados son importantes para el ámbito de la lingüística teórica, pero también tiene aplicaciones prácticas. Piénsese en la importancia que tiene para un niño que trata de aprender una lección que la información esté codificada en un vocabulario familiar para él.

VI. Conclusión

El análisis que se hace en este artículo de los listados de frecuencia de palabras existentes en español muestra que hay por lo menos dos listados bien elaborados y con características que garantizan su utilidad. Uno de ellos fue elaborado en Puerto Rico en 1952 y otro en España en 1953. En Venezuela no

se ha realizado, para conocimiento del autor, ningún trabajo de esta índole. Los dos listados mencionados anteriormente podrían emplearse, con ciertas modificaciones y precauciones, en el contexto venezolano, mientras se elabora una lista con las palabras que se utilizan en Venezuela. Esta tarea se hace hoy más simple que en el pasado debido a la existencia de las computadoras digitales, que simplifican el conteo rutinario de las frecuencias de aparición de las palabras.

El autor agradece la valiosa colaboración y recomendaciones hechas por el ingeniero Carlos Luis Niño C. y la Lic. Maríantonietta Castro Ruiz, durante la elaboración de este artículo.

Referencias bibliográficas

- Bloom, B. S.: **Handbook on formative and Summative Evaluation of student learning**. B. S. Blomm, J. Thomas Hastings, George F. Madans. New York, Mc. Graw Hill, 1971, 923 p.
- Bongers, H.: **The history and principles of vocabulary controls: as it affects the teaching of foreign languages in general and of English in particular**. Holland; Wocopi - Woerden, 1947.
- Buchanan, M. A.: **A graded Spanish Word Book**. Toronto, Am. and Canadian Committees on Modern Languages, III, 1929.
- Carrol, J. B.: **Word Frequency Book**. New York, American Heritage Publishing, 1972.
- Cartwright, C. W.: **A Study of the Vocabularies of Eleven Spanish Grammars and Fifteen Spanish Reading Texts**. Modern Language J. X., s.p.i. 1925.
- Céspedes, A. T. R.: **Investigación acerca las palabras usadas en castellano**. Panamá Star and Herald, 1929.
- Fries, Ch.: **English Word List**. Michigan, Ann Arbor, 1950.
- Gagné, R. M.: **The Conditions of Learning**. New York, Holt Rivechart and Winston, 1970.
- García Hoz, C.: **Vocabulario usual, común y fundamental**. Madrid, Instituto San José de Calasanz, 1953.
- Horn: **Basic Writing Vocabulary**. s.p.i.
- Jamieson, E. I. A.: **Standardized Vocabulary for Elementary Spanish Modern Language**. s.l.; s.n., 1924.
- Keniston, H.: **A Basic List of Spanish Words and Idioms**. s.l.; s.n., 1920.
- Meras, A.: **Pequeño vocabulario**. s.l., Heath and Co., 1927.
- Roberts, A. H.: **A Statistical Linguistic Analysis of American English**. The Hague Mouton, 1965.
- Rodríguez Bou. Ismael: **Recuento de vocabulario español**. Río Piedras, Universidad de Puerto Rico, 1925, 2v. Trabajo de investigación auspiciado por el Consejo Superior de Enseñanza de Puerto Rico.
- Thorndike, E. L.: **Difficulty, Reliability and Grade Achievement in a Test of English Vocabulary**, en "Teachers College Records", New York, Columbia University, 1923, p. 438-445.
- Zipf, G. K.: **The Psycho-biology of Language**. Boston, s.n., 1935.