

Frecuencia y Dispersión del Vocabulario Español

Hiroto Ueda

Universidad de Tokio

1987, 2017

Prefacio	...	1
I. Ordenación alfabética de vocablo	...	15
II. Ordenación numérica de suma	...	212
III. Ordenación numérica de uso	...	395
IV. Manuales de español. Ordenación alfabética de vocablo	...	592
V. Manuales de español. Ordenación numérica de suma	...	659

<https://lecture.ecc.u-tokyo.ac.jp/~cueda/kenkyu/goi/frec-disp/frec-disp-0.pdf>

<https://lecture.ecc.u-tokyo.ac.jp/~cueda/kenkyu/goi/frec-disp/frec-disp-1.pdf>

<https://lecture.ecc.u-tokyo.ac.jp/~cueda/kenkyu/goi/frec-disp/frec-disp-2.pdf>

<https://lecture.ecc.u-tokyo.ac.jp/~cueda/kenkyu/goi/frec-disp/frec-disp-3.pdf>

<https://lecture.ecc.u-tokyo.ac.jp/~cueda/kenkyu/goi/frec-disp/frec-disp-4.pdf>

<https://lecture.ecc.u-tokyo.ac.jp/~cueda/kenkyu/goi/frec-disp/frec-disp-5.pdf>

PREFACIO

INTRODUCCIÓN

Para la descripción de la lengua, se considera indispensable establecer la forma y estructura básicas de las unidades en distintos niveles. En el nivel léxico, se sabe que el inventario de las unidades es de carácter abierto y, por lo tanto, no se puede determinar su tamaño más que con cierta arbitrariedad. Esto se deberá a la cantidad de los vocablos existentes y también a la productividad en formación de nuevas palabras.

Con el objetivo de determinar el léxico fundamental, hemos elegido dos criterios: frecuencia relativa y dispersión en distintos campos de texto. Basándonos en el trabajo de 1987, donde preparamos un material para los análisis posteriores, expondremos en el siguiente la manera de cuantificación y la de la clasificación contrastiva del vocabulario español.

CORPUS

Se han venido realizando varios estudios cuantitativos sobre el vocabulario del español moderno. De ellos, si nos limitamos a enumerar las investigaciones sobre la frecuencia de los vocablos, se podrían considerar como importantes los siguientes: Buchanan [1941], Rodríguez Bou [1952], García Hoz [1953], Juilland and Chang Rodríguez [1964] y Márquez Villegas [1975]. Entre ellos, destacan las obras de García Hoz (que de aquí en adelante abreviamos con G.H.) y la de Juilland and Chang-Rodríguez (J.Ch.) por lo acertado en la elección del corpus, por lo explícitos que se dan los valores concretos de la frecuencia y también por la magnitud de la escala de la investigación. El recuento de G.H. es resultado de un trabajo manual con el uso de las fichas y J.Ch. es producto del procesamiento mecánico del ordenador. El primero fue llevado a cabo en España y el otro, una década después, en Estados Unidos. Como veremos más adelante, los campos objeto de la investigación de los dos se encuentran en relación casi complementaria, por lo cual la elaboración de un vocabulario contrastivo de ambos no dejará de ser un material lingüístico útil tanto para la enseñanza como para el estudio léxico del español moderno.

En 1984 hicimos, por nuestra cuenta, un recuento de los vocablos aparecidos en doce libros de texto publicados en los países extranjeros (Estados Unidos, Francia y España), y lo trataremos como el décimo campo para la comparación.

CAMPOS

G.H. ha seleccionado cuatro tipos de materiales correspondientes a diferentes

aspectos de la vida:

- | | |
|---------------------------------|--------------------------|
| (1) Vida familiar | ... Cartas privadas |
| (2) Vida social indiferenciada. | ... Periódicos |
| (3) Vida social regulada | ... Documentos oficiales |
| (4) Vida cultural | ... Libros |

El autor asegura que estos cuatro aspectos de la vida son «claramente diferenciables, porque son diferentes en sí mismos y porque tienen un medio propio de expresión.» (p. 18) Y añade que «también podríamos considerar que los cuatro tipos de material utilizados representan la vida desde un punto de vista dinámico, ya que los libros y los documentos oficiales expresan una vida ordenada, reposada, estratificada en algún sentido, mientras los periódicos y las cartas manifiestan la fluidez de la vida diaria.» (p. 20).

J.Ch. ha tratado, por su parte, los siguientes cinco campos, de los cuales el último coincide con un campo de G.H.:

- | | |
|------------------------------|----------------------------------|
| (5) Drama | Obras teatrales |
| (6) Ficción | Novelas, cuentos |
| (7) Ensayo | Ensayos, memorias |
| (8) Documentos periodísticos | Periódicos, semanales, etc. |
| (9) Documentos técnicos | Medicina, ingeniería, etc. |

Los textos de estos cinco campos han sido elegidos para asegurar que los ejemplos sean verdaderamente representativos de «standard written Spanish» y para facilitar el estudio de la dispersión de las unidades lingüísticas en «various worlds of the universe.» (p. xv)

Tanto G.H. como J.Ch. han realizado el cálculo a partir de 100.000 vocablos ocurridos en cada campo lingüístico, lo que es conveniente para poder ser comparados a base de una escala común. Ahora bien, nuestro campo, que será el décimo de la serie, está constituido por las frecuencias de las palabras aparecidas en doce textos del nivel medio. La elección de este material se debe al objetivo de determinar el vocabulario necesario para la enseñanza y, al mismo tiempo, de completar el vocabulario usual que no tiende a aparecer con frecuencia en los nueve campos tratados en G.H. y J.Ch.. Como el total del número de las palabras resulta 117.070, multiplicamos cada cifra de la frecuencia por $100.000 / 117.070$ para ajustarlo al volumen de los dos trabajos anteriores.

SEGMENTACIÓN

Establecidos los campos objeto de la investigación, ahora tenemos que buscar

la definición de la «palabra» como unidad conveniente para la léxico-estadística. El problema, que no es fácil de resolver en la práctica, se divide *grosso modo* en dos: segmentación y lematización.

El primero, la segmentación, se refiere a la cuestión de encontrar la línea de demarcación de las palabras en el texto linealmente continuo. Identificarlas por medio de espacio y signos de puntuación no sería muy correcto desde el punto de vista estrictamente lingüístico, por más conveniente que sea para el tratamiento por el ordenador. Hay combinaciones de dos palabras tales como verbos seguidos de enclíticos (p.ej. *levantarnos*) y amalgamas de preposición y artículo determinado (*al, del*), y por otra parte, casos en que dos palabras constituyen una unidad semántica, cuyos ejemplos serán los verbos en la forma compuesta (*he llegado*) y los numerosos modismos. Para el primer caso no hay problema, pues G.H. y J.Ch. concuerdan en segmentar y sumar las frecuencias. Nosotros también estamos de acuerdo y los hemos segmentado en nuestro décimo campo. En el criterio del segundo caso, al contrario, difieren notablemente. Mientras que G.H. independiza los modismos tales como *acerca de, en breve, a caballo, por cierto, con que, de conformidad, de continuo, etc.*, J.Ch. admite muy pocos: *en seguida, no obstante, por qué, y por tanto*. Nosotros hemos considerado los modismos no como una unidad independiente sino como una secuencia de más de dos palabras; y ello por las razones siguientes:

(1) Para la descripción del español, se pueden establecer por lo menos tres unidades independientes, que son «formante», «constituyente» y «significante» (véase Ueda, 1983), de los cuales el constituyente será la unidad objeto del presente análisis estadístico. El formante que se refiere a la morfología y el significante, que se encuentra en relación con los modismos, deberán ser estudiados aparte.

(2) No se puede establecer un criterio explícito para admitir o no los modismos. Efectivamente G.H. y J.Ch. adoptan criterios distintos y desde el punto de vista teórico es imposible demarcar una línea divisoria de las simples combinaciones de palabras y los modismos, entre las cuales existe una escala gradual y continua (véase Ueda [1990b]).

(3) G.H. ofrece las cifras correspondientes a los numerosos modismos, pero siempre con pocos valores de frecuencia:

Modismo	Frec.-1	Frec.-2	Frec.-3	Frec.-4	Total
<i>acerca de</i>	1	0	0	0	1
<i>en breve</i>	1	0	0	0	1
<i>a caballo</i>	3	0	1	0	4
<i>por cierto</i>	2	0	0	1	3

Es cierto que tenemos que aceptar los datos de poca frecuencia tales como se

presentan. Estamos, sin embargo, ante varios casos en los que se supone que el criterio no había sido observado con rigidez. Por ejemplo no nos puede convencer que el modismo *después de* tenga una sola ocurrencia a pesar de que figura *después* 405 veces. Hay más de esta índole: *general* (461) y *en general* (1); *menos* (324) y *por lo menos* (1); *mientras* (97) y *mientras que* (1).

Por consiguiente, en nuestro estudio descomponemos las cifras de G.H. y J.Ch. para distribuir las en los lugares adecuados de cada palabra constituyente.

LEMATIZACIÓN

El problema segundo es la «lematización». En español el vocablo presenta distintas formas, según el género y el número en el caso de nombres, pronombres, adjetivos, y según la conjugación en el de verbos. Está claro que hay necesidad de agrupar las cuatro formas de adjetivo, *bueno, buenos, buena, buenas* bajo una entrada representativa, BUENO.

bueno	}	BUENO
buenos		
buena		
buenas		

A este respecto no hay discrepancia entre G y J.Ch., pero sí la hay en el tratamiento de las formas masculina y femenina en los casos como *hijo* e *hija*. G los agrupa en un lugar y J.Ch. los separa como dos palabras distintas:

	J.Ch..	G.H..
hijo	}	HIJO
hijos		
hija		
hijas		

Comparemos los dos estudios con más detalle (L.: lematización; no-L.: no lematización):

Caso	Ejemplo	G.	J.Ch.
Masc. / fem.de los nombres	<i>hijo / hija</i>	L.	no-L.
Masc. / fem.de los adjetivos	<i>bueno / buena</i>	L.	L.
Sing. / pl. de los nombres y adjetivos	<i>hijo / hijos</i>	L.	L.
Nombre / adjetivo	<i>espan / ol / español</i>	L.	no-L.
Formas conjugadas de los verbos	<i>cantar / canto</i>	L.	L.

Verbo / participiopasado	<i>agitar / agitado</i>	L.* no-L.
Relativo / Interrogativo	<i>quien / quién</i>	L. no-L.
Conjunción / Interrogativo	<i>como / cómo</i>	L. no-L.
Forma breve / completa de los posesivos	<i>su / suyo</i>	L. no-L.

En G.H. los participios pasados no se lematizan con el infinitivo si el *Diccionario de la Lengua Española* de Real Academia Española ofrece una entrada independiente de los mismos.

En cuanto a los homófonos, J.Ch. hace distinción en tales casos como *acordar* 'llegar dos o más personas a estar conformes' y *acordarse* 'recordar'; *estado* 'condición' y *estado* 'cuerpo político'.

En resumen, la diferencia de los criterios tomados en los dos trabajos al hacer la lematización es la siguiente:

Caso	G.H.	J.Ch.
Modismos	no-L.	L.
Variación formal	L.	no-L.
Homófonos	L.	no-L.

En el presente trabajo, todas estas categorías las hemos unificado bajo una entrada. De otro modo no se podría hacer la comparación debido a los huecos de las cifras que se dan en las partes no lematizadas tanto en G, como en J.Ch..

	G.H..				J.Ch..				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
acostumbrado	-	-	-	-	2	4	3	4	0
acostumbrar	8	7	2	5	5	2	0	3	2
↓		↓				↓			
acostumbrar	8	7	2	5	7	6	3	7	2

PALABRAS FALTANTES

G.H. no distingue dos formas homófonas, por ejemplo, la de artículo y la de pronombre. Si los agrupáramos en este trabajo bajo una sola forma representante, el resultado no tendría mucho sentido. Además, en nuestro recuento de los manuales extranjeros, estas palabras no han figurado como objetos de investigación. Por estas razones los artículos, pronombres y conjunciones - interrogativos, por ejemplo, como y cómo, utilizamos el signo de asterisco (*) en las celdas correspondientes. Para el cálculo de media y desviación típica, utilizamos el número de datos no faltantes.

LÍMITE INFERIOR DE LAS FRECUENCIAS

El presente trabajo enfoca las palabras de relativamente alta frecuencia, y, por lo tanto, hemos de determinar el límite inferior de los valores. Mientras que G.H. presenta todos los vocablos incluso los de una sola ocurrencia, J.Ch. se limita a los de la frecuencia de más de 5. Como no podemos menos de escoger la parte común de los dos, omitimos la porción de los datos de G.H. de la frecuencia de menos de 4, a menos que tengan correspondientes en J.Ch.. Por nuestra parte, de los vocablos de los manuales ofrecemos los datos que tienen la frecuencia de más de 5 ocurrencias.

DISPERSIÓN

Para la cuantificación de dispersión, necesitamos hacer unas operaciones matemáticas: la media, la varianza, la desviación típica y la regularización.

La media (m) se obtiene por la división de la totalidad por el número de datos (n):

$$\text{Valor medio } (m) = 1/n \sum_i x_i \quad (i=1,2, \dots, n)$$

La varianza es cociente de la división de la suma de la diferencia entre el valor en cuestión y el valor medio, por el número de datos:

$$\text{Varianza} = \{ [x_1 - m]^2 + [x_2 - m]^2 + \dots + [x_n - m]^2 \} / n$$

La desviación típica (DT) es raíz cuadrada de la varianza:

$$\text{Desviación típica (DT)} = \sqrt{\text{Varianza}} = (\text{Varianza})^{1/2}$$

La DT, que se utiliza para conocer el tipo de la distribución, tiene el inconveniente de no ser su oscilación, de 0 a 1. Aquí, lo que nos interesa es la escala relativa de amplitud sobre una base común. El valor máximo de DT se presenta en el caso en que la distribución se concentre en un solo dato, por ejemplo {10, 0, 0, 0, 0}.

Representando la frecuencia concentrada con X , el máximo de DT es:

$$\begin{aligned} \text{DT.max} &= (\{ [x_1 - m]^2 + [x_2 - m]^2 + \dots + [x_n - m]^2 \} / n)^{1/2} \\ &= (\{ [X - m]^2 + m^2 + \dots + m^2 \} / n)^{1/2} \text{ (porque } x_2 \dots x_n = 0) \\ &= [(mn - m)^2 + (n - 1) m^2]^{1/2} \quad \text{(porque } m = X/n) \\ &= [(n - 1)m^2 + (n - 1)m^2]^{1/2} \\ &= m (n-1)^{1/2} \end{aligned}$$

Para que el valor máximo se haga 1, en forma de Desviación Típica Regular (DTR), hay que dividir la DT por la DT.max:

$$DTR = DT / DT.max = DT / [m \sqrt{(n-1)}] \quad 0 \leq DTR \leq 1$$

Cuando DTR = 0, la distribución se hace perfectamente igual, y cuando DT' = 1, perfectamente desigual, por lo cual DTR será un índice de concentración más bien que de dispersión. Para obtener la cifra representativa de la dispersión (Disp), sustraemos DTR de 1:

$$\text{Dispersión (Disp)} = 1 - DTR = 1 - DT / [m \sqrt{(n-1)}]$$

Ahora veamos algunos ejemplos de la distribución de frecuencias junto con su dispersión.

VOCABLO	C.G.	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	Suma	Disp.
<i>coma</i>	n	0	0	1	47	0	0	0	0	0	2	50	.065
<i>domicilio</i>	n	6	21	22	1	0	0	0	0	0	0	50	.437
<i>paciencia</i>	n	22	0	2	0	12	6	0	3	0	5	50	.551
<i>útil</i>	aj	4	1	12	7	0	2	2	3	17	2	50	.653
<i>sospechar</i>	v	5	6	1	4	14	10	1	2	4	3	50	.738
<i>atrás</i>	av	6	4	1	5	7	7	10	3	2	5	50	.831

Y finalmente, el índice de Uso (U) es un valor sintético del Total y la Dispersión:

$$U = \text{Total} \times \text{Dispersión}$$

Las palabras de alta frecuencia, si poseen un alto índice de parcialidad en su distribución, la cifra D se baja notablemente y el U, en síntesis, se convierte en un valor reducido. Por esta razón, el grado de importancia de la palabra en cuestión no se representa en el Total mismo de las frecuencias, sino más bien en el Uso, puesto que la importancia se definirá no sólo por frecuencia sino también por la nivelación de la distribución de los valores.

NUEVA EDICIÓN

En 1987, preparé como un informe interior una lista contrastiva de los vocablos españoles, a partir de las sumas que ofrecen dos estudios anteriores bajo el título de Frecuencia y Dispersión de Vocabulario Español en el Instituto de Estudios e Lenguas Extranjeras de la Universidad de Estudios Extranjeros de Tokio. Como se trata de un informe, no se ha hecho público y ahí quedó sin mucha repercusión con única excepción de Morales (1989), a la que contesté en 1993 (véase más adelante en ADDENDA donde reproducimos la nota). Posteriormente he recibido las preguntas de los profesores

interesados sobre el mismo informe, probablemente a través de las personas a las que obsequié un ejemplar. Se imprimieron solo unos pocos ejemplares. Aunque se trataba nada más que de una lista de frecuencias, parece que ha llamado la atención de los investigadores.

Ahora, 30 años después, ofrecemos a través de internet al público en general una nueva edición del mismo trabajo en forma algo distinta, basada en nuestras prácticas de estadística léxica. También hemos subsanado algunos datos faltantes que había en la edición antigua. El cambio grande es ofrecer distintos índices numéricos referentes a la frecuencia y dispersión de cada vocablo, en lugar de un simple lista contrastiva de frecuencias.

ADDENDA

Este trabajo es la versión continuada de nuestra obra mimeográfica de 1987; *Frecuencia y dispersión del vocabulario español*. En aquel entonces, intentábamos realizar un estudio cuantitativo del léxico español, para el cual necesitábamos los materiales básicos de ídole contrastiva. Lo preparamos y lo hicimos circular interiormente con contados ejemplares. En 1989, en la revista *Lingüística*, apareció una reseña muy detallada de la profesora Amparo Morales de Universidad de Puerto Rico. Se trata de una crítica explícita, objetiva y sumamente dura, lo cual nos ha resultado muy útil para corregir y ampliar algunos puntos no muy bien precisados en su introducción. Gracias a sus indicaciones hemos podido incorporar en esta edición más datos en los párrafos anteriores ya que "la escasez de material introductorio constituye una de las deficiencias de la obra, (...)" (Morales, 1989, p. 282).

Como este trabajo es básicamente el mismo que el de la versión anterior de 1987, nos sentimos obligados a contestar a todos los puntos de sus críticas para justificar la reedición. Nos prescindiremos de la parte explicatoria de la reseña y nos limitaremos a las partes concretas de sus críticas. El lector comprenderá que esto no es para hacer una refutación a su revisión. Lo hacemos porque creemos oportuno aclarar los puntos no mencionados en la dicha introducción con los detalles suficientes. Por lo tanto agradecemos profundamente a la profesora Morales su generosidad de revisar nuestro material desde el punto de vista científico, y por lo tanto, muy imparcial, lo cual nos ha posibilitado contestar a sus indicaciones de la manera más objetiva posible.

(1) "La escasez de material introductorio constituye una de las deficiencias de la obra, pues indica una ausencia de planteamientos teóricos sobre este tipo de trabajos." (loc. cit.)

No entendemos lo que quiere decir en este contexto el 'planteamiento teórico', ya que como dijimos anteriormente, el objetivo de la preparación de este material no ha sido establecer ninguna teoría lexicométrica, sino simplemente ofrecer una lista

contrastiva, lo cual sería de carácter bastante práctico. Nos parecía entonces que no era lugar para desarrollar ampliamente el planteamiento teórico de lexicometría.

(2) "Ueda no ofrece ningún argumento que justifique la elección de los diccionarios de G.H. y J.Ch. como fuentes del tratamiento estadístico posterior, ni tampoco hace planteamiento alguno sobre los propios cálculos estadísticos." (p. 282)

La justificación de elegir los dos diccionarios estaba en la pág. vii de la obra (en la pág. 2 de este trabajo). Lo repito: "(...) Entre ellos, destacan las obras de García Hoz y la de Juilland and Chang-Rodríguez por lo acertado en la elección del corpus, por lo explícitos que se dan los valores concretos de la frecuencia y también por la magnitud de la escala de la investigación." Las otras obras como la de Buchanan [1941] o la de Arias et alii [sin fecha] nos han sido inconvenientes por no ofrecer las cifras concretas de frecuencia sino sólo la clasificación cuantitativa de las palabras.

(3) La revisora hace referencia al índice de dispersión (D) comparando con el otro utilizado por Juilland y Traversa [1973]. "Este índice varía de 0 a 1, es cero o nulo cuando la palabra aparece solamente en uno de los campos léxicos, e igual a la unidad cuando se reparte equitativamente en todos. Quisimos comparar los resultados de esta fórmula con la de Ueda. Como siempre cuando se hacen estas comparaciones hubo diferencias; en este caso obsérvese los dos ejemplos considerados:

	FRECUENCIAS										UEDA	JUILLAND
<i>ácido</i>	0	5	2	9	0	0	0	0	0	0	.394	.093
<i>afortunado</i>	2	2	2	3	0	2	3	1	2	2	.854	.782

Se diría, por estos datos, que la fórmula de Ueda es mucho menos sensible a los ceros que la de Juilland y Traversa." (p. 284, nota 2)

Nos da la impresión de que la revisora no nos ofrece simplemente los datos. Si es que alega que nuestra fórmula tiene defecto de ser 'mucho menos sensible a los ceros', contestaríamos que la de Juilland es 'demasiado sensible' a los ceros. La cifra .093 es prácticamente nula y no parece convincente dar esta cifra a la distribución de frecuencias que tiene en los tres campos entre los diez (5, 2 y 9).

Por otra parte, nos sentimos muy agradecidos por indicarnos nuestra falta de explicación del índice D (dispersión). Se trata de un error tipográfico –falta de un paréntesis en la fórmula matemática– y de la omisión de los procesos intermedios para sacar el coeficiente del denominador, lo cual "hace totalmente imposible su posterior uso por parte de los lectores interesados." (p. 284) Respetando a sus indicaciones hemos incluido, en esta obra actual, todas las explicaciones que creemos necesarias para cada caso (véase la pág. 7).

(4) La revisora, después de presentar los cuatro campos de G.H. (cartas, periódicos, documentos oficiales, libros) y los cinco de J.Ch. (drama, ficción, ensayo, documentos

técnicos, documentos periodísticos), dice lo siguiente: "Con estos materiales los índices finales de dispersión y uso pierden parte de su rigor estadístico o, por lo menos, carecen del significado que normalmente se les atribuirá. Especialmente, es muy particular el caso de la obra que se reseña aquí, puesto que su corpus le ofrece, de base, ventaja al léxico del periodismo (este campo aparece repetido en G.H. y J.Ch.; igualmente, el campo léxico LIBROS de G.H. contendrá, forzosamente, en su interior, parte del léxico de los campos de J.Ch..)" (pág. 285)

Ciertamente para hacer esta lista contrastiva, podría haber prescindido de uno de dos campos del léxico periodístico. Pero nos interesó hacer la comparación –no se olvide que se trata de un estudio 'contrastivo'– de los recuentos independientemente hechos a partir de los materiales distintos. No es que ofrezca 'ventaja al léxico del periodismo', ya que no se repiten las mismas palabras en los dos campos. Los que se coinciden no son las palabras sino el nombre del campo.

Es cierto que el campo léxico LIBROS de G.H. contiene forzosamente, en su interior, parte del léxico de los campos de J.Ch.. Pero esto no se limita al caso del campo LIBROS. En todos los campos hay coincidencias parciales. Lo que la revisora querría indicar sería que una parte del léxico de LIBROS de G.H., se repite en DRAMAS, FICCIONES y ENSAYOS de J.Ch.. Pero, tampoco importa esto, ya que nuestro plan no es hacer de nuevo una lista contrastiva, sino utilizar los materiales ya hechos para comparar sus resultados.

(5) La revisora ataca la inclusión del léxico de MANUALES, que es nuestra aportación a la lista contrastiva, diciendo: "Ueda olvida que manuales representan generalmente la 'metalengua' y no la lengua usual." Y respecto a nuestro objetivo pedagógico del estudio léxico, dice: "Por otro lado, no dejan de ser planteamientos circulares los que consideran que los resultados que se obtengan en este nuevo campo léxico serán, a la vez, los que se tomarán como índice para 'determinar el vocabulario necesario para la enseñanza.'" (p. 285-286)

Lo que entendemos por el término 'metalengua' es 'la lengua para describir la lengua', y generalmente se aplica a la terminología lingüística. Creemos que la revisora quería decir la 'lengua artificial'. Y esto no es sólo cuestión de los MANUALES. Las obras literarias (como dramas y novelas) son, sin duda alguna, un tanto artificiales. No son la manifestación real de la lengua. Incluso en cartas privadas, no se puede evitar esta artificialidad. Pues, no son manifestaciones de 'la lengua usual'. El que no sea lengua usual no perjudica, de ninguna manera, la validez de la comparación. Estamos comparando varios estilos de la lengua y no es que estemos haciendo la investigación estilística averiguando que un estilo es natural o artificial.

En cuanto a la circularidad del planteamiento, creemos que hacía mucha falta la explicación en nuestra obra anterior. No pensábamos, claro está, determinar el

vocabulario básico para la enseñanza, partiendo directamente del recuento del léxico aparecido en los manuales. Esto resultaría, como dice la revisora con toda la razón, un caso completo del círculo vicioso. Aquí también entra la palabra 'comparación', que tantas veces hasta ahora hemos repetido. Queríamos hacer la comparación del léxico usado en varios campos, entre los cuales, pensábamos imprescindible la existencia del campo MANUALES, para poder conocer sus características relativas a los otros campos. Ha sido falta de palabras de nuestra parte si no nos ha entendido mal. Reproducimos aquí la parte correspondiente: "(...) hicimos, por nuestra cuenta, un recuento de los vocablos aparecidos en doce libros de texto publicados en los países extranjeros (...), y trataremos como el décimo campo para la comparación." (p. vii) "La elección de este material se debe al objetivo de determinar el vocabulario necesario para la enseñanza y, al mismo tiempo, de completar el vocabulario usual que no tienden a aparecer con frecuencia en los nueve campos tratados en G.H. y J.Ch.." (p. viii-ix)

(6) Por lo que se refiere a la segmentación, la revisora critica nuestra supuesta incongruencia del criterio: "Estas palabras [La explicación nuestra del caso 'levantarnos'] nos permiten deducir que *levantarnos*, lo mismo que *mirarla*, se desglosarán en sus unidades respectivas e irán a englosar con sus datos numéricos los totales de su cabeza o tema. En realidad, esto no es así, porque, si bien *nosotros* y *él/ella* aparecen en el listado y cabe pensar que incluyen los resultados de *nos* y *la*, respectivamente, por otro lado, el autor señala en un aparte posterior, titulado 'Palabras no tratadas en este material', que ha excluido 'las palabras gramaticales de alta frecuencia: artículos determinados e indeterminados (*el, los, la, las, lo, un, una*) y pronombres personales átonos (*me, te, lo, la, le, nos, os, los, las, les, se*), lo cual indica la exclusión de estas formas en el conteo y, con ello, deja sin posible interpretación lo expresado por el autor en la página ix de su Introducción." (págs. 287-288)

Ha habido la confusión bastante complicada de las palabras "segmentar y sumar". Hemos 'segmentado' la combinación *levantarnos* en las dos palabras: *levantar* y *nos*. Y sumamos en el conteo la frecuencia de *levantar* desglosado de *levantarnos* con la de *levantar* no reflexivo. Como hemos excluido los pronombres 'átonos' de nuestro listado, debido a la restricción de parte de la obra de G.H., hemos desechado el pronombre reflexivo (*nos*). Los pronombres 'tónicos' (por ej. *nosotros, él/ella*), por aparecer en ambas obras tratadas, los hemos incluido aparte. Por lo tanto, no cabe pensar, en contra de lo que dice la revisora, que hayan incluido los resultados de *nos* y *la* en los apartados de *nosotros* y *él/ella*, respectivamente.

(7) Finalmente desearíamos desembrollar su malentendido de la revisora, quien cree que habíamos intentado establecer una 'muestra representativa' del léxico español. Como hemos dicho varias veces, en esta modesta obra no se trataba de establecer la 'muestra representativa' del léxico español, sino de hacer un estudio comparativo de los

materiales sacados de las grandes obras lexicométricas. Reproducimos la parte correspondiente a su conclusión:

"Como dijimos en párrafos anteriores, los datos más útiles son los de frecuencia que aparecen en las primeras listas. En cuanto a los resultados de dispersión y uso generales, conviene no olvidar que el gran paso de avance en léxico-estadística para estos diccionarios ha sido la creación de conciencia respecto a la necesidad de partir de una muestra representativa. Éste es un proceso que ya nadie pone en duda; las deficiencias en este aspecto son las fallas más importantes que presenta el diccionario de Ueda." (parte final de su reseña, p. 289).

Creemos haber podido explicar todos los puntos de su crítica. Debemos admitir que la introducción de 1987 ha sido sucinta y reducida para poder aclarar las dudas que se les ocurran a los posibles lectores. Repetimos que agradecemos a la profesora Amparo Morales quien nos ha ofrecido la oportunidad de explicar las cuestiones y de corregir algunos puntos tan importantes como las fórmulas matemáticas de la dispersión.

REFERENCIAS

- Arias, F.R., Pallares, B y Alegre, J. M. [sin fecha] *Vocabulario básico del español*. [sin lugar de publicación].
- Buchanan, M. 1941. *A Graded Spanish word book*. Toronto Univ. Press, Toronto.
- García Hoz, Víctor. 1953. *Vocabulario usual, vocabulario común y vocabulario fundamental*. C.S.I.C., Madrid.
- Juilland, A. and Chang-Rodríguez, E. 1964. *Frequency dictionary of Spanish words*. Mouton, The Hague.
- Marquez Villegas, L. 1975. *Vocabulario del español hablado*. S.G.E.L., Madrid.
- Morales, A. 1989. "Reseña de: Hiroto Ueda, *Frecuencia y dispersión del vocabulario español*." *Lingüística, Publicación de la Asociación de Lingüística y Filológica de la América Latina*, 1, págs. 282-289.
- Rodríguez Bou, I. 1952. *Recuento del vocabulario español*. Ed. Universitaria, Río Piedras.
- Ueda, H. 1993. "Notas sobre lexicometría del español", en *Lingüística (ALFAL)*, 5, pp. 147-154.

MANUALES DE ESPAÑOL

- Adams, N. B. and Frey, H. J. 1964. *Spanish for today*. Holt, Rinehart and Winston, New York.
- Bolinger, D. L., Bowen, J. D., Brady, A. M., Haden E. F., Poston Jr., L. Sacks, N. P. 1960. *Modern Spanish*. (3a ed.) Harcourt Brace Jovenovich, New York.
- Dabele, L. 1972. *¿Adónde?* Librairie Armand Colin, París.
- École Berlitz. 1969. *Libro español*. Societé Internationale des École Berlitz, París.
- Hesse, E. W. 1960. *Spanish conversational review grammar*. American Book Company, New York.
- LaGrone, G. G. 1957. *Basic Conversational Spanish*. Holt, Rinehart and Winston, New York.
- _____, Sendon McHenry, A. y O'Connor, P. 1961. *Entender y hablar*. Holt, Rinehart and Winston, New York.
- _____, Sendon McHenry, A. y O'Connor, P. 1962. *Hablar y leer*. Holt, Rinehart and Winston, New York.
- Neale-Silva, E. and Nicholas, R. L. 1977. *¡Adelante!* Scott, Foresman and Co. Glenview, Illinois.
- Sánchez, A, Ríos, M., y Domínguez, J. 1977. *Español en directo*. S.G.E.L. Madrid.
- Turk, L. H. 1965. *Foundation course in Spanish*. D. C. Heath and Co., Boston.
- ____ y Allen, E. M. 1956. *El Español al día*. (2a ed.) D.C. Heath and Co., Boston.