

Selección léxica en la percepción de un texto

POR

RAMON ALMELA PEREZ

*Departamento de
Lingüística Española*

SUMARIO

I. CONDICIONES METODOLOGICAS DE LA REALIZACION DE LA ENCUESTA.

II.—DATOS Y ANÁLISIS.

- A) Elección y olvido de palabras.
- B) Conexiones entre palabras emitidas y percibidas.
- C) Las palabras gramaticales.

III.—CONCLUSIÓN.

I CONDICIONES METODOLOGICAS DE LA REALIZACION DE LA ENCUESTA

El texto base de la encuesta consta de dos páginas escritas a máquina y a doble espacio. Es la reproducción de un artículo aparecido en 1978 en una revista de alcance nacional y de ámbito e información generales, publicada en España.

La encuesta la realicé, entre los días 4 y 11 de mayo de 1978, sobre 420 individuos agrupados en 12 sectores, perteneciendo éstos, a su vez, a los siguientes ámbitos: E.G.B., B.U.P., Formación Profesional, Escuelas Técnicas Superiores y Filosofía y Letras.



Entregadas las dos hojas sin explicación previa alguna, dispusieron los encuestados de quince minutos para leerlas. Transcurrido este tiempo, les fueron retiradas las hojas que contenían el texto, y se les dio una nueva hoja, para que en ella, y en los lugares prefijados, fueran anotando las 10 primeras palabras, y por el orden cuantitativo de mayor a menor frecuencia que recordaran.

No se admitían grupos de palabras, por muy estrecha que fuera su unión, sino solamente palabras gráficamente individualizadas. En los casos en que fueron escritos como respuesta aquellos grupos de palabras, yo adopté como válida la primera de ellas.

Era inevitable que algunos —instintos infantiles— intentaron copiarse. Para conservar masivamente la independencia de respuesta, prohibí que hicieran preguntas en voz alta; sobre todo, para que las preguntas que algunos pudieran formular no sirvieran de sugerencias de respuestas para otros. Sin embargo, en algunos sectores del grupo de E.G.B. y B.U.P. no tomé intencionadamente esta precaución, decisión que se notó, como explicaré más abajo.

Adopté como unidad la «forma» y no el «lema», porque se trataba, no de obligarles a realizar tareas metalingüísticas de especialistas, sino de ponerles en ocasión de ejercitar acciones lingüísticas de oyentes, con lo que ello implica de frescura y naturalidad. *Estatal* y *estatales* son, por tanto, dos respuestas distintas. Tomo en el sentido de Ch. Muller los términos de «forma» (= vocablo) y de «palabra» (= aparición, ocurrencia).

II. DATOS Y ANALISIS

Presentar aquí y ahora una relación exhaustiva de datos sería pesadísimo y poco elocuente. Ofreceré sólo aquellos que vayan confirmando las reflexiones.

Analizaremos este contenido en tres apartados:

- 1.º Criterios de elección y olvido de palabras.
- 2.º Conexiones entre las palabras reales del texto y las percibidas por el oyente.
- 3.º Las palabras «gramaticales».

1.º ELECCIÓN Y OLVIDO DE PALABRAS

Plantearemos dos cuestiones. Primera: ¿con qué criterios prefiere el oyente unas palabras a otras, o deja de percibir algunas de las que

se le han transmitido? Segunda: ¿qué pesa más en el resultado de una mayor frecuencia efectiva de las formas elegidas?

Vayamos con la primera. La frontera que separa a los vocablos elegidos de los olvidados se apoya en dos pilares:

- el significado, lo semántico.
- la expresión, la forma.

Es curioso constatar que de las 234 formas del texto, justamente la mitad (117) han sido reseñadas y otras tantas han quedado silenciadas. Como número exacto, este hecho es anecdótico; como cantidad, es sintomático.

La fidelidad léxica es una utopía. En el camino de la comunicación verbal, gran parte de las unidades léxicas quedan en la cuneta de lo ignorado, y, por ende, inutilizado.

Desde el punto de vista de la expresión, la categoría que capitaliza casi la mitad del número de formas reseñadas es el sustantivo: 52 vocablos son sustantivos (44,84 por 100). Mucho menos pesa el adjetivo (26 vocablos) y muy poco el verbo (7 vocablos).

Por el contrario, entre los vocablos no elegidos predomina el verbo (39 vocablos); le siguen el sustantivo y el adjetivo, así como las restantes categorías. En cuanto al número de ocurrencias, el predominio del sustantivo es mayor todavía: el 70 por 100, frente al 21 por 100 del adjetivo y menores porcentajes de las restantes categorías.

No se trata de hacer apología del paralelismo entre las categorías lógicas y las categorías gramaticales; pero sí, de evidenciar que las realidades designadas por las unidades lingüísticas tradicionalmente llamadas sustantivos, son de carácter basilar y constituyen los elementos más recurridos de la experiencia.

En las etapas del lenguaje infantil se perfilan en primer lugar las entidades objetuales, y en segundo lugar las predicaciones de dichas entidades: *τό ὄνομα* precede a *τό ῥήμα*.

Desde el punto de vista del significado, el criterio es mucho más espermable. Agrupo los vocablos (sustantivos, adjetivos y verbos) en pertinentes y no pertinentes, según se relacionan o no semánticamente con uno de los seis conceptos que vertebran el argumento del artículo. Dicho argumento se puede expresar así: LA ECONOMIA DE NUESTRO PAIS, DEMOCRATICO, DEPENDE DE LA REFORMA FISCAL. Los conceptos angulares son, pues: -Economía -España -Estado -Democracia -Reforma -Fiscal.

Entre los 52 sustantivos elegidos, 32 —o sea, dos tercios— son pertinentes; pero, además, sucede que esas 32 formas copan el 97 por 100 de las ocurrencias de los sustantivos elegidos, es decir, prácticamente

todas. Ejemplos: *administración* (111 apariciones), *gasto* (139 ocurrencias), *control* (38 palabras), son formas pertinentes; mientras que son formas no pertinentes: *aparato* (1 ocurrencia), *forma* (4 apariciones), *tiempo* (2 palabras).

Confirmando estos resultados, y aplicando el mismo criterio, compruebo que ninguno de los sustantivos olvidados es pertinente: *obtención*, *manera*, *senderos*, *condiciones*...

Entre los adjetivos elegidos, 13 formas son pertinentes: *legislativo*, *español*, *público*... Otras 13 son no pertinentes: *definitiva*, *reacio*, *manejadas*... Pero si hay paridad en el número de formas pertinentes y no pertinentes, no la hay en la cantidad de frecuencia de unas y otras: aquellas tienen 601 ocurrencias (93,32 por 100) y estas tienen 43 apariciones (6,68 por 100).

Los verbos muestran signos menos elocuentes, pero no dejan de serlo en alguna medida: hay tres formas pertinentes que engloban el 80 por 100 de las ocurrencias, y cuatro formas no pertinentes que comprenden el 20 por 100 de las palabras de este mismo grupo.

Detalle significativo. Entre los adjetivos y los verbos no elegidos hay dos y tres formas pertinentes, respectivamente. Se trata de los adjetivos *estatales* y *fiscales*, y de los verbos *controlar*, *dirigir* y *gastarlos*. Lo curioso es que cuatro de ellos fueron sustituidos, entre el elenco de los vocablos elegidos, por sendos sustantivos: *estatales* por *estado*, *controlar* por *control*, *dirigir* por *dirección* y *gastarlos* por *gasto*.

Como conclusión podemos aseverar que el oyente selecciona unas palabras ateniéndose a su «capacidad prereceptiva», o «prerecepción»: «quidquid recipitur ad modum recipientis recipitur». En el orden formal atiende más al sustantivo, y en el orden significativo, a los vocablos emparentados temáticamente.

Es decir, en su experiencia personal virtual racio-verbal ha asimilado bien que las realidades objetuales —en un grado más o menos abstracto— son el punto de referencia de toda conceptualización. Y en su experiencia actual del acto de comunicación verbal, el oyente va contorneando el tema en los aspectos que considera más fundamentales.

Pues bien, la suma de estas dos experiencias, la virtual y la actual, es lo que yo llamo «prerecepción» o «capacidad prereceptora», de acuerdo con la cual percibe o no percibe el oyente.

La segunda cuestión es esta: ¿qué influye en la mayor o menor cantidad de ocurrencias de cada forma elegida? En la escala decreciente de frecuencias, las 16 formas con más de 50 apariciones suman 2.256 palabras. Pues bien, de esos 16 vocablos, 14 son sustantivos y abarcan 1.940 palabras, o sea, el 85,99 de la anterior cantidad.

Además, es digno de constatar el siguiente hecho. Considerando las familias morfológicas —concretamente me he fijado en diez—, es decir, grupos de formas con representantes en varias categorías y englobando también a las formas que figuraban en las respuestas sin figurar en el texto base, de nuevo aparece la preferencia por el sustantivo. *Gasto* y *gastos* alcanzan 214 apariciones, mientras que sus correspondientes formas verbales *gasta*, *gastamos* y *gastarlo*, y la adjetival *gastado* alcanzan todas juntas solamente 55.

Administración tiene 111 palabras, mientras que *administra*, *administrar* y *administrativa* consiguen 6. *Economía* alcanza 306 ocurrencias, y los adjetivos *económica*, *económicas*, *económico* y *económicos* tienen sólo 21. Lo mismo ocurre con *democracia*, *estado*, *España*...

Atendiendo ahora, no a la jerarquía de frecuencia, sino a las posiciones que ocupan, es evidente que los dos primeros puestos recibirán a los vocablos preferidos por el lector oyente. A partir de la posición tercera comienza a decrecer la significancia del puesto respecto de la predilección del oyente, y se van colocando formas por ser «formas», no por ser «tales» formas. No es casualidad que casi todas las posiciones primeras y segundas (739 sobre 840, esto es, el 88 por 100), están ocupadas por sustantivos; es débil la presencia del adjetivo en esos lugares, y casi nula la del verbo.

El factor, pues, que determina el puesto más o menos predominante en la jerarquía de las respuestas es el categorial: el sustantivo es la categoría privilegiada sin ambages y con clara ventaja desde todos los ángulos. No puede argüirse como causa de este fenómeno el hecho de que en el texto base el sustantivo predomine. Sí abundan, pero no predominan; su presencia no es tan abrumadora ni tan absorbente respecto a verbos y adjetivos: el 33 por 100 en formas y el 23 en ocurrencias.

Más aún, en relación con las categorías gramaticales, se presentan éstas en el texto globalmente en más cantidad que el sustantivo y son recogidas en menos.

2.º CONEXIONES ENTRE PALABRAS EMITIDAS Y PERCIBIDAS

Llamaremos «desvíos» a las palabras que figuran en la encuesta, pero no en el texto. Desde el punto de vista de las categorías gramaticales, hallamos de nuevo la presencia aplastante del sustantivo: más del 75 por 100 de las palabras son sustantivos.

Pero la cuestión más interesante es otra: ¿qué conexiones unen a las palabras que emite el hablante con las que retiene el oyente, en los

que no coinciden unas y otras? ¿Qué relación liga a los «desvíos» con las palabras reales del texto?

Es sorprendente que una notabilísima parte de los «desvíos» no tenga similitud alguna con el original. La mayoría, sin embargo, (el 87 por 100), sí está en las mismas «circunscripciones» de las palabras reales. ¿Cuáles son ellas? Aprecio dos clases de áreas y, dentro de ellas, cinco órbitas.

Las áreas son la del contenido y la de la expresión. El área del contenido consta de tres órbitas: la semántica, la pragmática y la lógica. El área de la expresión consta de dos órbitas: la fonológica y la morfológica.

Empecemos por la conexión lógica del área del contenido. Pocos vocablos, pero algunos, se encuentran conectados lógicamente con el argumento del artículo tal como lo describimos más arriba. He de aclarar previamente qué entiendo por órbita lógica.

Agrupo los conceptos, y las palabras que le están asociadas, en cuatro pares de ejes: teoría frente a praxis, razón frente a sensación, esencia frente a existencia, abstracto frente a concreto.

Pues bien, el argumento del texto se adosa nítidamente al segundo término de cada uno de esos cuatro pares. Por ello, a los desvíos que se articulan igualmente en torno al segundo grupo, les asigno el mismo campo lógico que a las palabras argumentales del texto. Algunos ejemplos: *dato, consumo, servicio, población, actuar...*

Por conexión pragmática entiendo la que existe si el «desvío» constituye una prolongación, análisis o aplicación del significado de uno de los seis conceptos fundamentales del argumento. Veamos unos ejemplos. A la órbita pragmática de *economía* pertenecen: *presupuesto, precio, renta*. A la de *reforma*: *saneamiento, crisis*. A la de *país*: *Europa, extranjeros*. A la de *fiscal*: *impuestos, recaudación*.

Este tipo de conexión es notable cuantitativamente, extensionalmente sivo de su concepto.

Menos número de casos —el 13 por 100— pertenecen a la órbita semántica. Esta se da entre el desvío y el vocablo del texto cuando ambos disfrutan de semas comunes. El lazo en la órbita semántica y en la pragmática es el significado, pero en aquella es más estrecho, y en esta es más laxo. Ejemplos: *economía* atrae a *financiación*; *país*, a *nación*; *pesetas*, a *monetario*; *estado*, a *gobierno*.

Casi la mitad de los desvíos conectan morfológicamente con las formas (el 21 por 100 de los desvíos), aunque escurridiza en el contorno intemas reales del texto: *cien* con *ciento*; *legislar* con *legislativo*; *preocupa-*

ción con preocupante; capitalismo con capitalistas; gastamos con gastarlos; democráticamente con democracia.

Casi inexistente es la similitud fonológica. Solamente constato un caso: *veracidad* por *voracidad*.

Como resumen podríamos representar en cuatro círculos concéntricos, y según intensidad decreciente de relación mutua, las diversas clases de conexiones entre los desvíos y las palabras originales del texto:

- círculo 1.º = órbita lógica
- » 2.º = » pragmática
- » 3.º = » semántica
- » 4.º = » fonomorfológica.

En sólido equilibrio, las áreas del contenido y de la expresión copan, cada uno, casi la mitad de las palabras y de las formas, siendo aquellas ligeramente más abundantes en la zona de la expresión, y estas en la zona del contenido.

3.º LAS PALABRAS «GRAMATICALES»

¿Qué papel juegan en el mecanismo inconsciente de interpretación léxica las palabras llamadas «gramaticales»? Las palabras llamadas «útiles», «sinsemánticas», «vacías», o como se prefiera, juegan un papel muy débil.

En este trabajo considero como tales a las preposiciones, conjunciones, artículos y a las formas simples del verbo *haber* que no van solas, como *he* y otras. De las 22 consideradas, 5 no aparecieron debido a su escasa frecuencia en el texto. Claro que otras dos, también escasas, sí que aparecieron. En conjunto, del casi 40 por 100 de palabras (189 sobre 493) que ocupan en el conjunto del texto, sólo fueron elegidas en un 6,05 por 100 respecto al total de las apariciones en las respuestas.

La más frecuente, *de*, con 28 apariciones, sólo fue recogida en un 4,33 por 1.000, mientras que las más frecuentes de las no gramaticales, *Estado* y *público*, con 7 cada una, esto es, 4 veces menos que *de* obtienen el 116,21 por 1.000 y el 28,63 por 1.000, respectivamente. Más simple: *de* tiene el 5,67 por 100 de frecuencia de palabra en el texto, y el 0,43 por 100 de las palabras elegidas; *estado* tiene el 1,41 por 100 (cuatro veces menos que *de*), y el 11,62 de las palabras elegidas (27 veces más). Pero hay más: mientras *de* aparece en primera posición dos veces y en segunda posición, *estado* ocurre 103 veces en primera posición y 79 en segunda posición.

Se trata de una consideración en discurso. En lengua no creo que sean «sinsemánticas», «vacías» estas palabras. ¿Cómo negar que *Y*, por

ejemplo, tiene un significado, y bien nítido, de conexión? ¿Acaso es inteligible una frase como «IBA CALLE IBA SOMBRERO», a la que faltan *el, por la, pero, sin*? Claro que esa construcción es «superada» en inteligibilidad por otra en la cual sólo estuvieran presentes los ausentes: «EL POR LA PERO SIN». En la práctica, sin embargo, son palabras de las que no hay conciencia comunicativa.

Pero, yendo más al fondo de la cuestión: ¿cuál es la causa de que no se sientan significativas? A mi juicio, este fenómeno se debe a que su referente no es objetual, sino una relación, un ente de razón, con más valor lógico que ontológico, más verdadero que real, menos plástico, más abstracto. Por citar sólo algunos casos: la ilación, la dirección del movimiento, la pertenencia, la instrumentalidad..., son significados tan necesarios en la lengua como en el discurso.

Pero incluso hay más. El grupo de encuestados, en algunos de cuyos sectores permití preguntas en voz alta, dio el porcentaje más alto de palabras gramaticales: casi dos tercios; el otro tercio lo comparten los restantes dos grupos. Entre todas las sesiones, que totalizaban 420 individuos, sólo a siete de ellos se les ocurrió hacer preguntas que se enderezaban a aclarar si esta clase de palabras también se admitían; mi respuesta, en todos los casos, les dejaba en libertad.

III. CONCLUSION

La rentabilidad y la fidelidad del empleo de la categoría «sustantivo» en el circuito de la comunicación verbal quedan evidenciadas.

Tomen nota los políticos que en sus mítines quieran hacer oyentes «fieles». Tomemos nota los profesores y conferenciantes para no rellenar con adjetivos el espacio de la cadena que debe vehicular el pensamiento, y establecer y mantener el contacto racioverbal entre hablante y oyente.

No he constatado diferencia alguna entre las respuestas de los tres grupos. En cualquiera de las cuestiones y apartados eran homólogas las tres clases de contestaciones, con la excepción de las palabras gramaticales, a cuyos hecho y causa me referí antes.

II. DATOS Y ANALISIS

A) ELECCION DE PALABRAS Y FORMAS

(Aspecto estudiado: categoría gramatical)

1. FRECUENCIA (EN %)

Cat. gram.	En el texto		En las respuestas	
	Formas	Palabras	Formas	Palabras
Adjetivos	19,66	13,19	23,61	21,55
Sustantivos	33,76	23,12	44,84	69,04
Verbos	19,66	11,77	3,39	2,18
Pal. Gramatic.	9,40	37,93	14,92	6,05
Otras	17,52	13,99	13,24	1,18

2. DISTRIBUCIÓN POSICIONAL (EN %)

Cat. gram.	1. ^a posición	2. ^a posición
Adjetivos	6,90	16,90
Sustantivos	88,09	75,48
Verbos	0,95	1,90
Pal. Gramatic.	3,58	3,33
Otras	0,98	2,39

B) CONEXIONES ENTRE LO EMITIDO Y LO PERCIBIDO

(Las cantidades hacen referencia a %)

1. AMPLITUD DE LA CONEXIÓN DESVÍO/VOCABLO REAL

	Están conectadas	No están conectadas
Formas	77	23
Palabras	87	13

2. LOS TIPOS DE CONEXIÓN

Areas	Orbitas	Formas	Palabras	Total área	
				Formas	Palabras
Contenido	Lógica	11,48	7,24	41,87	41,68
	Pragmática	23,64	20,83		
	Semántica	6,75	13,61		
Expresión	Morfológica	34,46	45,12	35,12	45,32
	Fonológica	0,67	0,20		

3. CATEGORÍAS GRAMATICALES DE LOS DESVÍOS

	Sustan- tivos	Adjetivos	Verbos	Otras
Formas	55,06	22,98	18,59	3,37
Palabras	75,40	14,32	8,24	2,04

C) LAS PALABRAS GRAMATICALES

1. PORCENTAJES TOTALES

	<i>En el texto</i>	<i>En las respuestas</i>
Formas	9,40	14,92
Palabras	37,93	6,05

2. COMPARACIÓN DE PORCENTAJES REPRESENTATIVOS

<i>Forma</i>	<i>Frecuencias</i>		<i>Distribución</i>	
	<i>En el texto</i>	<i>En las res- puestas</i>	<i>En 1.º posición</i>	<i>En 2.º posición</i>
De	5,67	0,43	0,47	0,23
Estado	1,41	11,62	24,52	18,80