El español en cifras

AQUILINO SÁNCHEZ
Universidad de Murcia

En esta sección de *Carabela* se ofrecen datos cuantitativos relativos al uso del español actual en todos sus ámbitos (geográficos, sociales, tipológicos...), en las modalidades oral y escrita.

Los datos cuantitativos son útiles para que el profesor pueda tomar algunas decisiones razonablemente fundadas en relación con lo que debe o conviene enseñar en determinados casos.

Los datos de esta sección están basados en el corpus *Cumbre* (propiedad de SGEL, S.A.). *Cumbre* consta de 20 millones de palabras (equivalentes a unos 200 libros de 300 páginas cada uno) y es representativo del español oral y escrito usado en todos los países de habla hispana en la última década del siglo XX. Este corpus ya ha servido de base para dos obras importantes: un diccionario del español actual (*Gran Diccionario de uso del español actual*, SGEL, 2001) y una gramática (*Gramática práctica del español actual*, SGEL, 2005). El corpus *Cumbre* está parcialmente etiquetado (subcorpus *Cumbre-2*, de 2 millones de palabras) y sobre esta recopilación se basan también algunos datos más precisos.

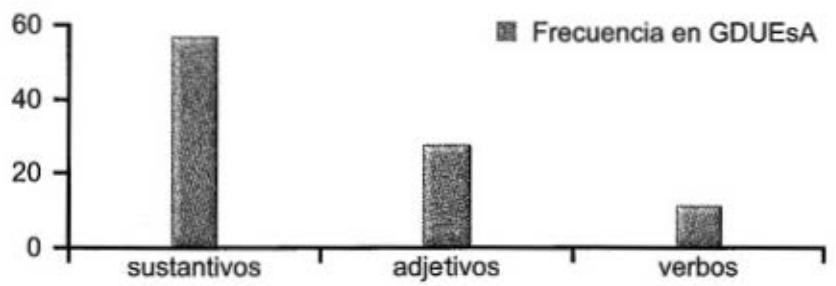
Las partes de la oración en los diccionarios y en el uso real de la lengua

François Gouin (1831-1896), un profesor de idiomas más original que conocido, afirmaba que el verbo es la parte más importante de la oración, porque es el que señala o significa la acción, y en la expresión de la acción se centra –dice– lo más importante de la comunicación mediante el lenguaje. En su método –el método de *las Series*– concreta sus convicciones destacando todas las formas verbales en columna aparte para que los alumnos las aprendan con mayor facilidad (Sánchez 1997:125 ss). Esta percepción no parece ser compartida por la generalidad de los hablantes e incluso por los autores de materiales: las palabras que aprenden los alumnos en los niveles elementales incluyen más sustantivos que verbos o adjetivos. ¿Existe alguna razón a favor de uno u otro punto de vista?

El cómputo de las palabras que integran un diccionario actual (*Gran Diccionario de Uso del Español Actual*, SGEL, 2001) atribuye el 95,5% del total de lemas a sustantivos, adjetivos o verbos, en las siguientes proporciones (el 4,5% restante incluye el resto de partes de la oración):

Sustantivos	56,64%
Adjetivos	27,02%
Verbos	11,87%

Los porcentajes se visualizan mejor en el gráfico equivalente:



Una de las funciones principales de los diccionarios es recoger las palabras usadas por la comunidad de hablantes de una lengua. De ahí que los lexicógrafos intenten

incluir todas las voces usadas en un determinado momento histórico, sin tener en cuenta su grado de uso. De los datos presentes en el GDUEsA cabe concluir que los sustantivos o nombres constituyen el componente más importante del español: los sustantivos duplican a los adjetivos y quintuplican a los verbos en cantidad. Este hecho avala lo que los hablantes suelen percibir, a la vez que revela que la referencia a todo lo que nos rodea se concreta de manera prioritaria en los nombres.

Los repertorios léxicos tradicionales, que no toman en consideración el uso que los hablantes hacen de las palabras, variarían de manera significativa si tuviéramos ese uso en cuenta. El mismo GDUEsA, que también ofrece –en su versión escrita y electrónica– información sobre la frecuencia de uso de las voces recogidas, permite constatar que las aproximadamente 1.400 voces marcadas como de frecuencia muy alta (más de 75 ocurrencias por millón) se reparten de la manera siguiente:

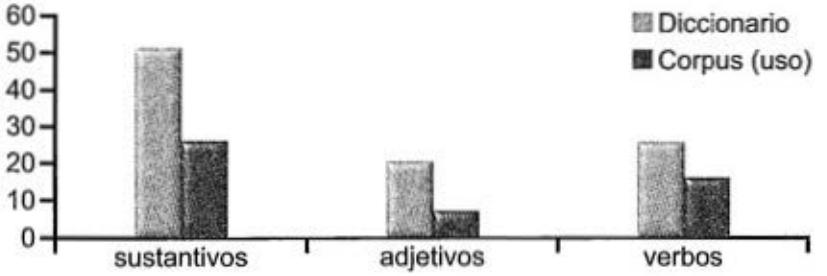
Sustantivos	50,83%
Adjetivos	20,13%
Verbos	**25,14%**

Atendiendo a esa franja de uso, los sustantivos siguen acaparando la frecuencia más elevada, pero los verbos duplican su presencia, mientras los adjetivos disminuyen un poco en número. Este hecho revela claramente que, analizando los lemas incluidos en los diccionarios desde la perspectiva del uso, los porcentajes sufren variaciones significativas a favor del verbo.

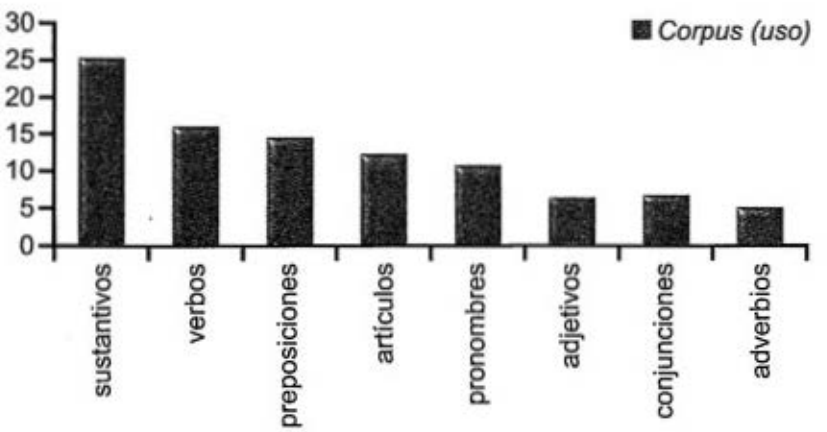
A los profesores de español como lengua extranjera les interesa todavía otro dato, relacionado con la rentabilidad comunicativa de las palabras que los alumnos aprenden o deben aprender. Y este dato se encuentra solamente en el análisis de las palabras que usamos los hablantes de español. De nuevo, el corpus *Cumbre* nos permite extraer información relevante e ilustrativa. No obstante, hay que tener presente que cuando analizamos el lenguaje usado en la comunicación diaria, ya no hablamos de palabras lematizadas –como hacen los diccionarios–, sino de formas o de variantes de los lemas, tal cual realmente aparecen en el uso (*mesa, mesas, hablo, hablaban, habló*, etc.). Teniendo, pues en cuenta estas precisiones, *Cumbre-2* revela que los porcentajes de las palabras usadas, agrupadas según la parte de la oración a la que pertenecen, difieren mucho de las cifras que ofrecen los diccionarios. Así, las tres partes de la oración predominantes en los repertorios lexicográficos quedan reducidas a las siguientes cifras cuando contabilizamos su frecuencia en el uso comunicativo:

Sustantivos	25,8%
Adjetivos	6,91%
Verbos	15,73%

La diferencia en los porcentajes se visualiza mejor en este gráfico combinado:



Comparando diccionario y corpus, la frecuencia de las formas sustantivas disminuye a la mitad en el uso comunicativo, las formas verbales aumentan, mientras que las formas de los adjetivos disminuyen notoriamente en cantidad. A este hecho hay que añadir otro dato importante: algunas de las categorías que en el diccionario tienen una presencia numérica muy baja, cobran una gran importancia en el uso diario (es decir, se usan con mucha frecuencia): son las formas llamadas 'funcionales', como los artículos, las preposiciones, los adverbios y las conjunciones. Si todas estas categorías juntas apenas cuentan con un 5% de presencia en los diccionarios, en el uso comunicativo de la lengua alcanzan el 50%. Las preposiciones, artículos y pronombres superan en frecuencia a los adjetivos y se acercan al verbo. El siguiente gráfico es muy revelador:



En términos comparativos, al sustantivo le corresponde en el diccionario un porcentaje del 56,6%, mientras en el corpus o uso real le corresponde solamente un 25,8%, es decir, casi **la mitad menos**. En cambio, las preposiciones reflejan un comportamiento muy diferente: en el diccionario cuentan con una presencia del 0,042%, mientras en el uso gozan de un 14,4%, es decir, **344 veces más**.

Al profesor de español le interesa saber respecto a las palabras funcionales (preposiciones, conjunciones, artículos...) que, aunque no son portadoras de significado por sí mismas, su incidencia en el uso es tan alta que importa aprenderlas ya desde los estadios iniciales. Así lo demanda su índice de rentabilidad comunicativa, aunque sólo sea para establecer relaciones entre palabras u oraciones.

También es preciso dar parcialmente la razón a Gouin en lo que se refiere a la mayor presencia de los verbos: entre los 50 lemas y formas más frecuentes –si agrupamos las categorías de sustantivo, verbo y adjetivo– 30 de los lemas son verbos, y en cuanto a las formas, el número es de 28.

Palabras (formas) funcionales más frecuentes del español	**Sustantivos, verbos y adjetivos más frecuentes del español (formas)**	**Sustantivos, verbos y adjetivos más frecuentes del español (lemas)**
de	es	ser
la	ha	saber
que	está	haber
y	hay	estar
el	ser	tener
en	era	hacer
a	había	poder
los	años	ir
se	tiene	decir
no	fue	unir
un	puede	año
las	han	comer
del	vez	ver
por	parte	bueno
con	tiempo	vez
una	vida	dar
lo	día	parar
para	hacer	querer
su	hace	día
al	bueno	deber
como	decir	grande
más	mundo	entrar
o	va	llegar
pero	estaba	pasar
me	estado	tiempo
le	sea	político
sus	ver	poner
si	están	país
yo	gobierno	vida
ya	casa	hombre
este	sido	sobrar

Palabras (formas) funcionales más frecuentes del español	Sustantivos, verbos y adjetivos más frecuentes del español (formas)	Sustantivos, verbos y adjetivos más frecuentes del español (lemas)
porque	dijo	parecer
muy	país	nuevo
todo	hombre	hablar
cuando	forma	dejar
qué	año	llevar
sin	caso	parte
son	hecho	casar
sobre	he	persona
también	momento	seguir
esta	general	quedar
sí	creo	encontrar
entre	cosas	mujer
mi	trabajo	mundo
dos	días	venir
nos	tienen	momento
hasta	política	general
desde	tenía	formar
te	dice	último
eso	vamos	señor

Referencias

SÁNCHEZ, A. (1997), *Los métodos en la enseñanza de idiomas*, Madrid, SGEL.