El español en cifras. Las palabras y lemas más frecuentes del español actual

AQUILINO SÁNCHEZ

Las lenguas pueden describirse y presentarse al estudiante, extranjero o no, desde puntos de vista diferentes. Uno de ellos es el uso que hacen de su lengua los hablantes nativos de la misma. En esta sección de ***CARABELA*** se irán ofreciendo datos reales del español actual, tanto en su vertiente oral como escrita, y a partir de muestras de todos los ámbitos geográficos en los que se habla, es decir, España y los distintos países hispanoamericanos.

Los datos que se irán ofreciendo en este apartado han sido extraídos de un corpus representativo del español actual: el corpus *Cumbre*, de 20 millones de palabras (equivalente a unos 200 libros de 300 páginas cada uno), cuya recopilación finalizó antes del año 2000, y sobre el cual se basó el *Gran diccionario de uso del español actual* (*-GDUESA-*, SGEL, 2001), en su versión gráfica y electrónica. Algunos datos más específicos se han extraído de un *subcorpus* de 2 millones, igualmente representativo del conjunto, pero etiquetado morfológicamente y susceptible, por tanto, de aportar datos más detallados sobre aspectos morfológicos del español.

En la historia de la enseñanza de idiomas se han llevado a cabo varios estudios para determinar las palabras más usadas en una lengua concreta. Sobresalen los llevados a cabo en Estados Unidos e Inglaterra: Thorndike, E. L., (1932), *A Teacher"s Word Book of the 20,000 Words Found Most Frequently and Widely in General Reading for Children and Young People*, Columbia University, Teachers College.; (1944), *The Teacher"s Word Book of the 30,000 Words*, Columbia University, Teachers College. West, Michael, (1953), *A General Service List of English Words,* London, Longman.

También en español se han publicado algunos trabajos en el mismo sentido: García Hoz, Víctor, (1953), *Vocabulario usual, vocabulario común y vocabulario fundamental*, CSIC, Madrid; Márquez Villegas, L. (1975), *Vocabulario del español hablado*, SGEL, Madrid. Ambos estudios adolecen de falta de representatividad, no solamente porque la distancia temporal es ya significativa, sino también porque el método aplicado era o muy restrictivo en la tipología de muestras utilizadas (García Hoz), o poco ajustado a criterios objetivos (Márquez Villegas).

Pero la elaboración de vocabularios 'fundamentales' o de las listas de las palabras más frecuentes en cualquier lengua se ha hecho posible en los últimos años del siglo XX, con fiabilidad y rapidez, mediante la ayuda de los ordenadores y a partir de grandes recopilaciones de muestras lingüísticas. El corpus *Cumbre (*propiedad de *SGEL)*, de 20 millones de palabras en sus inicios, es por el momento la fuente sobre la que se basan los actuales listados de frecuencia del español. El libro recientemente publicado, *Frecuencias del español* (de Almela y otros, Editorial Universitas 2005) recoge con mayor amplitud listados léxicos básicos y otros estudios más detallados y enjundiosos. Aquí ofrecemos solamente dos listas útiles para el profesor de español: un listado de las 500 palabras (formas) más frecuentes en el uso del español y otro de los 250 lemas más utilizados. Conviene diferenciar entre *formas* y *lemas*. Por *forma* se entiende cualquier palabra que aparece en el uso, en cualquiera de las modalidades flexivas en que pueda darse (*un, una, es, tiene, tiene, tengo*...), mientras que el *lema* se refiere a lo que se entiende por voz o entrada en un diccionario, incluyendo, por tanto, cada lema todas las posibles formas flexivas de una voz (*ser -soy, eres, es, somos*...-, *tener -tengo, tenía, tendré*...-, etc.). Ambas listas son útiles para el profesor de español. El verbo ***tener*** es muy frecuente en español, pero no todas sus formas lo son: *tiene* cuenta con el 30% de todos los usos de 'tener', mientras que *tengo* sólo cuenta con el 8,5% y el uso de *tendréis* es casi irrelevante. De otra parte, todas las formas del presente de indicativo de 'tener' acaparan por sí solas el 66% del uso total de este verbo. Así pues, no es igualmente rentable aprender primero 'tiene' o 'tendréis', a pesar de que el verbo *tener*, en su conjunto, figure entre las voces más usadas.

A partir de estas reflexiones, juzgue el profesor de español por sí mismo sobre la importancia de rentabilizar la enseñanza y aprendizaje de palabras, apoyándose para ello en los dos listados ofrecidos, extraídos del corpus *Cumbre* y avalados por el uso que del español oral y escrito hacen actualmente los hablantes de esta lengua en España y en todos los países de Hispanoamérica.

1. Las 500 formas más frecuentes del español actual:

(Los números que siguen a cada palabra representan el número de veces que ésta aparece en el corpus *Cumbre*, de un total de 20 millones de palabras).

De	1216593	Dos	29334	Poco	14904
La	770656	Había	28013	Mucho	14724
Que	673494	Nos	27671	Usted	14452
Y	563139	Años	27604	E	14391
El	557608	Tiene	26605	Estaba	14236
En	522624	Hasta	26062	Estado	14165
A	429826	Desde	25753	Sea	14139
Los	327038	Te	25473	Ver	14097
Se	265864	Eso	25374	Gran	13961
No	233880	Fue	24739	Cómo	13897
Un	222746	Todos	24580	Aunque	13844
Las	215618	Puede	23953	Están	13762
Del	211604	Pues	23722	Tres	13726
Por	202115	Han	23513	Les	13662
Con	183781	Así	23160	Antes	13646
Una	181191	Bien	23031	Gobierno	13548
Es	176187	Vez	21350	Casa	13430
Lo	137537	Ni	21134	Sido	13402
Para	135083	Sólo	20690	Algo	13271
Su	127901	Ahora	20130	Dijo	13156
Al	113580	Él	19975	País	13094
Como	102577	Uno	19047	Sino	13050
Más	93074	Parte	18868	Hombre	13046
O	82455	Ese	18736	Forma	13030
Pero	72694	Tiempo	18128	Estos	12895
Me	69189	Vida	18086	Año	12873
Le	61247	Mismo	17886	Caso	12753
Ha	55331	Otro	17405	Hecho	12485
Sus	54089	Día	17357	Durante	12378
Si	53544	Cada	17225	Hoy	12334
Yo	47042	Hacer	17146	Luego	12268
Ya	45649	Siempre	17000	He	12245
Este	42028	Entonces	16891	Momento	12230
Porque	38804	Nada	16844	Unos	12154
Muy	38398	Donde	16768	General	11854
Todo	38351	Esa	16764	Creo	11841
Cuando	37518	Hace	16658	Ellos	11838
Qué	37472	Bueno	16574	Cosas	11783
Sin	36372	Decir	16543	Todas	11591
Son	36067	Tan	16504	Además	11440
Sobre	35467	Otra	16485	Mayor	11438
Está	34596	Esto	16388	Mejor	11364
También	33752	Después	15823	España	11362
Esta	33333	Ella	15733	Trabajo	11286
Hay	33141	Menos	15282	Toda	10967
Sí	33067	Tanto	15269	Días	10800
Entre	31747	Otros	15232	Tienen	10729
Ser	31317	Mundo	15203	Política	10648
Era	31293	Aquí	15182	Tenía	10614
Mi	30871	Va	15163	Contra	10468

Dice	10364	Tipo	7995	Aún	6370
Hacia	10356	Cosa	7904	Incluso	6345
Vamos	10315	Punto	7898	Fueron	6329
Tal	10278	Cualquier	7884	Persona	6322
Otras	10225	Primer	7883	Problema	6319
Veces	10113	Sociedad	7882	Bajo	6318
Gente	9841	Agua	7722	Países	6292
Ejemplo	9757	Presidente	7715	Da	6291
Lugar	9729	Embargo	7709	Realidad	6218
Tener	9648	Historia	7653	Madre	6198
Casi	9606	Pasado	7625	Partido	6153
Estas	9589	Grandes	7615	Posible	6118
Personas	9503	Nuestra	7585	Problemas	6093
Parece	9428	Noche	7521	Número	6091
Mujer	9409	Hemos	7505	Nombre	6088
Primera	9402	Cual	7489	Dar	6062
Según	9363	Importante	7460	Fuente	6034
Mientras	9221	Dicho	7390	Esos	6025
Mí	9182	Cuatro	7362	Mujeres	6014
Nosotros	9172	Estar	7251	Voy	6002
Poder	9168	Fuera	7225	Ojos	5983
Nunca	9160	Allí	7209	Cambio	5928
Tu	9128	Información	7207	Desarrollo	5928
Algunos	9116	Mis	7141	Través	5912
Pueden	9089	Siglo	7104	Cierto	5896
Claro	9034	Ello	7086	Gracias	5884
Cuenta	8979	Tenemos	7080	Cinco	5873
Sé	8925	Acuerdo	7058	Guerra	5827
Nuevo	8878	Padre	7037	Habían	5827
Debe	8870	Tarde	6995	Don	5814
Tú	8844	Mañana	6988	Relación	5812
Medio	8666	Haber	6978	Obra	5807
Nacional	8650	Unas	6943	Proceso	5793
Ante	8628	Nueva	6859	Ir	5774
Muchos	8586	Estamos	6846	Programa	5761
Señor	8472	Ciudad	6834	Hora	5746
Madrid	8457	Nadie	6777	Mal	5679
Millones	8425	Sentido	6673	Hizo	5675
Dentro	8393	Todavía	6644	Estados	5673
Nuestro	8382	Estoy	6629	Cuerpo	5652
Ahí	8302	Situación	6574	Tema	5649
Manera	8262	Grupo	6560	Primero	5645
Sistema	8257	Alguna	6532	Algunas	5597
Misma	8247	Ley	6522	Pueblo	5592
Fin	8237	Lado	6487	Horas	5589
Verdad	8227	Eran	6479	Quiere	5586
Quien	8215	Van	6469	Tras	5537
Muchas	8116	Será	6433	Tierra	5523
Tengo	8102	Frente	6406	Familia	5512
Social	8042	Mano	6388	Modo	5497

Centro	5461	Tampoco	4554	Papel	4053
Derecho	5452	Edad	4546	Razón	4013
Economía	5421	Ellas	4525	Político	3993
San	5406	Español	4519	Director	3964
Dios	5398	Voz	4513	Cara	3963
Hablar	5388	Semana	4505	Nuestros	3912
Quiero	5352	Libro	4493	Siguiente	3910
Cabeza	5335	Propio	4493	Casos	3909
Pasa	5325	Digo	4488	Dónde	3907
Determinar	5260	Seguridad	4488	Europa	3900
Algún	5255	Hijo	4477	Servicios	3894
Hombres	5232	Ningún	4472	Condiciones	3891
Sería	5200	Junto	4460	Eh	3883
Empresa	5184	Niños	4459	Ninguna	3867
Final	5171	Paso	4454	Orden	3864
Cuanto	5170	Pesetas	4447	Total	3861
Arte	5156	Podría	4443	Medios	3859
Quién	5144	María	4439	Seis	3853
Esas	5117	Universidad	4407	Tuvo	3850
Público	5107	Dado	4401	Estaban	3849
Iba	5085	Luz	4399	Media	3847
Llegar	5084	Vista	4382	Campo	3845
Igual	5074	Palabras	4375	Salud	3833
Meses	5068	Jóvenes	4348	Dirección	3824
Dinero	5036	Visto	4348	Difícil	3819
Amor	5013	Respecto	4335	Personal	3814
Mil	5007	Obras	4308	Viene	3808
Sabe	4984	Hijos	4301	Siendo	3803
Educación	4957	Comunicación	4297	Pasar	3796
Calle	4950	Real	4293	Proyecto	3795
Falta	4935	Puesto	4281	Pronto	3785
Soy	4904	Haya	4265	Joven	3781
Muerte	4894	Puerta	4257	Haciendo	3775
Buena	4877	Carlos	4248	Camino	3772
Podía	4866	Trata	4248	Existe	3756
Último	4858	Española	4244	Época	3749
Largo	4807	Mayoría	4227	Buenas	3748
Ayer	4797	Hubiera	4214	Consejo	3740
Música	4791	Tenido	4185	Diferentes	3739
Segundo	4782	Aquella	4180	Buenos	3736
Saber	4757	Principio	4150	Aire	3723
Cultura	4711	Podemos	4147	Quienes	3722
Aquel	4699	Buen	4146	Segunda	3720
Decía	4689	Bastante	4134	Fuerza	3713
Éste	4648	Idea	4119	Hacen	3710
Mercado	4636	Único	4109	Hacía	3706
Partir	4620	Interés	4088	Quería	3701
Demás	4618	Niño	4080	Rey	3701
Nivel	4614	Propia	4069	Derechos	3695
Manos	4608	Ustedes	4065	Empresas	3693

Interior	3687
Comunidad	3676
Dio	3670
Población	3660
Acción	3649
Medida	3641
Servicio	3636
Alguien	3630
Últimos	3630
Iglesia	3622
Económica	3621
Mediante	3615
Base	3605
Internacional	3601
Televisión	3597
Pesar	3593
Varios	3590
Sociales	3588
Datos	3580
Puedo	3576
Allá	3571
Palabra	3563
Última	3553
Valor	3552
Diez	3546
Relaciones	3544
Serie	3544
Especial	3536
Futuro	3531
Cine	3526
Producción	3523
Padres	3522
Fondo	3515
Anterior	3514
Favor	3494
Carácter	3489
Queda	3478
Ministro	3475
Actividad	3469
Grupos	3468
Ésta	3460
Nuevos	3448
Libertad	3441
Artículo	3440
Has	3439
Llama	3432
Alto	3427
Aquellos	3418
Salir	3416
Estudio	3414

2. Los 250 lemas más frecuentes del español actual:

(Los números que siguen a cada voz representan el número de veces que ésta aparece en el corpus *Cumbre*, de un total de 20 millones de palabras).

de	1216684	cuando	38854	vida	18721
la	986276	ver	38759	usted	18518
el	721130	muy	38399	hombre	18284
que	710981	mi	38016	sobrar	18243
y	563142	sin	36372	son	18077
en	522628	mucho	34593	parecer	17839
a	429846	también	33765	sobre	17733
ser	319626	vez	31468	quien	17626
un	316814	eso	31399	nuevo	17336
lo	301055	mismo	30482	cada	17226
no	233886	dar	30463	siempre	17015
del	211605	parir	29720	llevar	16964
por	202117	parar	29225	entonces	16893
con	183787	bueno	28937	hablar	16842
saber	165596	querer	28889	casar	16537
haber	139412	día	28159	tan	16507
se	132957	nos	27673	aquel	16483
su	127901	hasta	26062	esto	16389
al	113593	desde	25752	dejar	16212
estar	111401	te	25481	seguir	16064
tener	104468	deber	25461	quedar	15950
hacer	98029	bien	24747	entre	15874
más	93082	pues	23724	después	15824
este	89433	entrar	23655	mujer	15451
si	86639	alguno	22848	ante	15451
todo	85501	ello	22824	mundo	15428
o	82940	nuestro	22301	venir	15308
como	82280	esa	21881	menos	15283
su	81993	llegar	21558	encontrar	15259
poder	81111	ni	21136	aquí	15183
para	81056	partir	21102	momento	14881
ir	79162	sólo	20690	dos	14668
decir	76696	donde	20677	mayor	13984
pero	72713	pasar	20592	gran	13961
uno	71898	tiempo	20523	aunque	13844
me	70274	político	20359	tres	13726
le	68077	ahora	20132	general	13680
unir	58391	él	19981	formar	13521
otro	50682	primero	19844	último	13424
yo	47054	tanto	19741	señor	13351
ya	45656	poner	19733	llamar	13314
año	40492	poco	19628	crear	13290
porque	39674	ese	19569	algo	13273
comer	39033	país	19428	sino	13050

trabajar	12901	tratar	9673	casa	7748
estos	12895	nosotros	9650	gobernar	7715
volver	12855	noche	9618	fuente	7700
estado	12756	casi	9607	vario	7694
nada	12638	tomar	9590	información	7678
contar	12613	cualquier	9544	único	7647
problema	12413	mí	9503	utilizar	7616
durante	12379	producir	9459	relación	7593
hoy	12334	presentar	9438	servicio	7530
mejor	12323	realizar	9430	bajo	7474
pensar	12275	según	9369	cuatro	7363
luego	12267	cuanto	9259	mañana	7309
creer	12155	mientras	9221	posible	7231
conocer	11978	millón	9218	trabajo	7231
salir	11946	empezar	9209	abrir	7220
asir	11809	manera	9182	allí	7209
hijo	11789	nunca	9160	lado	7205
propio	11716	existir	9142	cuyo	7166
social	11634	preguntar	9024	considerar	7139
hecho	11617	coser	8960	oír	7134
así	11581	económico	8938	recibir	7093
además	11441	empresa	8878	escribir	7082
sentir	11438	sociedad	8857	desarrollar	7061
lugar	11358	tú	8845	pedir	7045
claro	11341	cambiar	8824	ojo	7040
hora	11338	fin	8709	resultar	7033
tu	11330	ciudad	8656	número	6941
tal	11315	permitir	8457	cuerpo	6842
público	11200	verdad	8427	antes	6823
derecho	11162	siglo	8416	gracia	6821
cosa	11161	dentro	8393	familia	6789
mano	11001	ahí	8302		
niño	10862	ley	8299		
parte	10847	presidente	8295		
grande	10809	forma	8216		
mirar	10760	esperar	8214		
importante	10652	determinar	8183		
ejemplo	10617	alto	8164		
gente	10609	joven	8131		
padre	10578	gustar	8076		
cierto	10566	buscar	7954		
contra	10470	personar	7951		
hacia	10356	palabra	7946		
nacional	10308	aún	7932		
caso	10291	persona	7913		
cual	10095	pequeño	7903		
tipo	10048	primer	7883		
grupo	10029	situación	7840		
vivir	9969	gobierno	7831		
medio	9674	español	7829		