El español en la Red

POR **GUILLERMO ROJO**

El estudio de la presencia del español en Internet muestra una evolución positiva, pero su peso con relación a su importancia mundial parece manifiestamente mejorable y exige esfuerzos urgentes para mejorar todos sus parámetros.

La rapidísima evolución de los computadores y, sobre todo, el vertiginoso desarrollo de Internet ha cambiado profundamente los hábitos y el modo de actuar de una buena parte de la humanidad en los más diversos aspectos de su vida cotidiana, la forma de acceder a la información, tanto en aspectos personales como profesionales, públicos o privados, el ocio, las relaciones con los demás, etc. Dado que sirve como vehículo para transmitir o recibir información, la comunicación a través de Internet supone, en cualquiera de sus usos posibles, el empleo de una lengua, de forma que la presencia que un idioma tiene en la Red deriva fundamentalmente del grado en que la comunidad lingüística correspondiente tiene acceso a ella y la medida en que la utiliza diariamente. Además, el empleo de las lenguas en Internet se ha convertido en el índice más llamativo mediante el cual se mide el peso mundial de un idioma, en sustitución de los empleados tradicionalmente, con lo que (el estudio de) la mayor o menor utilización de una determinada lengua en la Red se carga inevitablemente de factores culturales, políticos y económicos (1). En este trabajo me propongo examinar ?de modo forzosamente superficial? los datos y las cuestiones fundamentales que se relacionan con el uso del español en Internet.

Los datos básicos

Cuando nos referimos a ´la Red?, ´Internet´, ´la web´, ´la gran malla mundial´, etc., no lo hacemos habitualmente para designar el sistema de comunicación electrónica mediante el cual podemos acceder con gran rapidez desde un computador a otro, sino al conjunto de varios miles de millones de documentos de las más variadas características y de aplicaciones de todos los tipos situados en millones de sedes electrónicas que residen en cientos de millones de servidores. Con la prudencia exigible cuando se trata de este tipo de cuantificaciones, y sólo

Fundación Telefónica
Telefónica

para que los lectores puedan hacerse cargo de la naturaleza de la cuestión que estamos tratando, el *Internet Systems Consortium* estimaba en 440 millones el número de servidores en funcionamiento en julio de 2006 (2).

Mucho más difícil de calcular es el número de sedes electrónicas (*sites*, traducido habitualmente por ´sitios electrónicos´ o ´sitios web´) y, más interesante para nuestros propósitos, el número de páginas electrónicas existentes en la actualidad. Debido a factores en cuyo análisis no podemos detenernos aquí, como la diferencia entre páginas estáticas y páginas dinámicas (esto es, páginas que se construyen en el momento en que se hace la consulta y desparecen inmediatamente después), la existencia de sedes electrónicas reservadas, de sedes electrónicas que son simples copias (*mirrors*) de otras, limitaciones en el acceso de los buscadores generales, etc., resulta imposible conocer el número real de páginas electrónicas existentes, de modo que tenemos que conformarnos con estimaciones muy dependientes de la tecnología utilizada y, como es de esperar, bastante divergentes entre sí. Es bien sabido que la publicidad de Google señala que el buscador facilita el acceso al contenido de unos 8.000 millones de páginas que ha indexado previamente. Se trata, sin duda, de un número enorme, pero está suficientemente demostrado que Google, como cualquier otro buscador, solo ´ve´ (es decir, indexa) una parte de lo realmente existente en Internet, de modo que Gulli y Signorini (2005) calculaban que en 2005 la Red estaba constituida por unos 11.500 millones de páginas. Más sorprendente, sin duda, resultará el dato facilitado por Aguillo, Ortega y Granadino (2006) según el cual Google contiene en realidad casi 40.000 millones de páginas, cinco veces más que la estimación del propio sistema. Sería necesario revisar esta cifra para tener en cuenta enlaces rotos, redondeos de cifras y páginas repetidas, además de la evidencia de que el buscador no procesa una buena parte (el 50 por ciento, según estos autores) de la parte pública de la Red, con lo que podemos situarnos en una cifra comprendida entre 25.000 y 30.000 millones de páginas en 2006.

Lamentablemente, no todo el mundo puede acceder a esta ingente cantidad de información. Según estimaciones de diciembre de 2006, sólo el 16,8 por ciento de la población mundial (unos 1.100 millones sobre un total de 6.500 millones) tiene actualmente posibilidad de entrar en Internet. Ese acceso, además, se distribuye de forma muy diferente en función de las características socioeconómicas de cada zona, de modo que oscila entre el 3,6 por ciento de la población de África y el 70 de la de Estados Unidos y Canadá, pasando por el 38,7 por ciento de la europea (3). La evolución de los últimos años permite albergar cierta esperanza, ya que el aumento para África entre 2000 y 2006 se cifra en un 625 por ciento, pero es evidente que las cifras tienen que cambiar mucho para alcanzar una situación mínimamente satisfactoria.

El mundo hispánico

¿Cuál es la situación actual de la comunidad hispana en cuanto a su presencia en Internet? Comenzando por el último dato, tiene acceso a la Red el 18,7 por ciento de los habitantes de los países hispanohablantes (4), lo cual supone una media escasamente superior a la mundial. Naturalmente, hay también aquí fuertes oscilaciones, que se sitúan entre, por ejemplo, los casos de Cuba, Honduras o Nicaragua, todos ellos inferiores al 5 por ciento, y los de España (43,3), Chile (42,8), Argentina (26,4 por ciento), Puerto Rico (25,2), Costa Rica (22,7), Uruguay

Fundación Telefónica
Telefónica

(20,8) o la población hispana de Estados Unidos (19,1 por ciento), que son los únicos situados por encima de la media del conjunto de estos países. Los casos de España o Chile, claramente destacados en este bloque, quedan sin embargo muy reducidos en importancia si los comparamos con la media de acceso en, por ejemplo, Suecia (74,9 por ciento), Australia (70,7), Estados Unidos (69,9), Japón (67,2) o la Unión Europea de 25 miembros (51,9). Este último caso es especialmente significativo para España, que pasa de ocupar el primer puesto entre los países hispánicos a situarse en los últimos lugares de la UE de 25 miembros, ya que únicamente Chipre, Grecia, Hungría, Lituania, Malta y Polonia (5) están por debajo de ella en la tasa de usuarios sobre la población total. El notable crecimiento experimentado desde el año 2000 (256 por ciento) no ha sido suficiente para nivelar la situación y colocar a nuestro país a la altura media de los veinticinco que formaban la UE en 2006.

A pesar de la baja tasa de acceso a Internet, el volumen general de la comunidad hispanohablante la convierte en la cuarta en peso específico de los internautas, con un 7,5 por ciento del total, por detrás únicamente de las comunidades de lengua inglesa (29,7), china (13,3) y japonesa (7,9) (6). Aunque la relación general es evidente, es claro también que la utilización de una lengua determinada en Internet no tiene forzosamente que ir en paralelo a la cantidad de personas conectadas a la Red desde países en los que esa lengua es oficial o una de las oficiales. El dato utilizado más frecuentemente para medir el peso de un idioma en Internet es el número de páginas electrónicas escritas en esa lengua. A las dificultades mencionadas anteriormente para llevar a cabo estudios de este tipo y valorar sus resultados se une ahora la consistente en la necesidad de detectar automáticamente la lengua en que está escrito un documento en el interior de un conjunto formado por miles de millones de páginas (7).

Para lograr una primera impresión general del papel que juega el español en la Red resulta útil partir del muy conocido estudio que Funredes ha venido realizando a lo largo de los últimos años (8) (ver tabla (1)).

Las líneas generales de la evolución son bastante claras. El inglés, que era la lengua en la que estaba escrito el 75 por ciento de lo que se podía detectar en Internet en 1998, había bajado hasta el 45 por ciento en marzo de 2005. Naturalmente, eso no significa que el número absoluto de páginas en inglés haya descendido, sino que su aumento ha sido menor en proporción al experimentado por el conjunto de las demás lenguas, lo cual es perfectamente esperable a partir de la distribución inicial. El español crece de forma notable hasta 2002, año en que comienza a descender hasta llegar a la situación en la que tiene un peso algo inferior al del francés y muy inferior al del alemán, lenguas que tienen un número considerablemente menor de hablantes.

Los datos anteriores resultan de una proyección realizada a partir de la presencia relativa de las demás lenguas con respecto al inglés. Dada la inseguridad general que arrastran consigo estos datos y también por su posible envejecimiento por los casi dos años de antigüedad transcurridos desde la última recogida, no estará de más contrastarlos con otros más recientes o bien obtenidos mediante una metodología diferente. La tabla 2 muestra en la primera columna los resultados del estudio de Aguillo, García y Arroyo (2004) utilizando los recursos proporcionados por Google y Yahoo para hacer el recuento del número de páginas existentes

Fundación Telefónica
Telefónica

en diferentes lenguas a mediados de 2004. La técnica empleada, que cualquiera puede utilizar, consiste en lanzar al buscador la petición de páginas que contengan una expresión general (ellos utilizan la expresión *http*) y estén, a juicio del identificador lingüístico del buscador, escritas en una lengua determinada. La segunda columna contiene, en las casillas correspondientes, los resultados del estudio de Funredes; la tercera da los datos de otro estudio realizado por Aguillo con otros colaboradores (Aguillo, Ortega y Granadina, 2006) sobre los contenidos de Google en mayo de 2006 y la cuarta proporciona los datos que yo mismo obtuve de Google a comienzos de octubre de 2006 utilizando la misma técnica empleada por Aguillo, García y Arroyo en el trabajo mencionado. Además, para que se pueda valorar adecuadamente el volumen al que estamos haciendo referencia, añado en la quinta columna el número absoluto de páginas existentes en las diferentes lenguas usando también los servicios proporcionados por Google (ver tabla (2)).

La distribución por países presenta el perfil esperable: España supone, según los diferentes recuentos, entre el 30 y el 45 por ciento de las páginas que los buscadores consideran escritas en español y los cinco países de mayor peso (Argentina, Chile, España, Estados Unidos y México) suman en conjunto entre el 75 y el 90 por ciento.

Aun con fuertes diferencias entre sí, los datos que la tabla 2 proporciona para el español son sistemáticamente bajos en relación con el peso cuantitativo de la comunidad hispanohablante. Todavía un paso más allá, lo que podríamos considerar el ´índice de productividad´ de los internautas hispanohablantes, esto es, la relación entre la cantidad de páginas electrónicas y la de usuarios de Internet es de 0,99, mientras que alcanza 9,55 para los anglófonos, 2,21 para los hablantes de japonés, 2,33 para los de francés, etc.

A pesar de la incertidumbre que contienen y las divergencias que muestran, los datos sobre el número de hispanohablantes conectados a la Red y de páginas electrónicas en español dibujan un panorama manifiestamente mejorable. Por supuesto, ni son los únicos ni, probablemente, los más significativos. Tener la posibilidad de conectarse a la Red no es lo mismo que hacerlo o limitarse a hacerlo una vez al trimestre y entrar en Internet puede no ir más allá de ´bajar´ una canción o un vídeo. Por otra parte, los recuentos de páginas cuentan con independencia de la calidad, el número de visitas, la actualización de la información, etc. Todo eso es cierto, pero también lo es que las cifras generales de acceso, incluso las más burdas, dan una idea del grado de inserción de la comunidad correspondiente en la llamada ´sociedad del conocimiento´, de modo que han de ser atendidas y las entidades responsables deberían tomar las medidas oportunas para mejorar la relación que presentan.

Obtener índices más elaborados, que apunten a aspectos cualitativamente más elevados, resulta bastante más difícil. Por ejemplo, da la sensación de que la prensa digital tiene en el mundo hispánico un peso esperanzador, tanto en el número de publicaciones con versión electrónica (o versión únicamente electrónica) (9) como en cantidad de consultantes. Para dar únicamente un dato, los diez periódicos digitales más importantes del mundo hispánico recibían conjuntamente en octubre de 2006 el 1,1 por ciento de todas las visitas a sedes electrónicas de este tipo. Según Díaz Noci (2006: 309 y sgs.), en 2005 había en España 1.274 medios de comunicación electrónicos (alrededor de un 10 por ciento de los existentes) y con datos de la OJD, se situaban, en noviembre de 2005 en, por ejemplo, 35 millones de visitas para *El Mundo*,

Fundación Telefónica
Telefónica

casi 30 millones para *Marca*¸ 17 millones para *As* ó 3.800.000 para *Libertad digital* en el mismo período. Desde un ángulo diferente, la versión de la *Wikipedia* en español tenía en esa misma fecha 160.313 artículos y 234.792 usuarios registrados, lo cual supone la décima posición en cuanto a número de artículos, con el 11 por ciento de los artículos existentes en la versión inglesa y el 9,5 por ciento de sus usuarios registrados.

Conclusiones: El futuro del español en Internet

Como hemos visto en los apartados anteriores, la tasa de hispanohablantes con acceso a Internet es bastante baja, sólo ligeramente superior a la media mundial. Incluso los países hispánicos mejor situados en este aspecto están todavía lejos de alcanzar los niveles de los más desarrollados. Aunque la evolución de los últimos años ha sido muy positiva, los contenidos en español presentes en la Red están también a mucha distancia de lo que el peso cuantitativo de los hablantes de esta lengua podría suponer. También aquí, la situación de los países que ocupan la cabeza del mundo hispánico resulta muy distante de la que tienen los que ocupan los primeros puestos.

Sería una actitud absurda esconder la cabeza y no reconocer la necesidad de mejorar la situación de nuestra lengua en la Red de redes. No se trata, por supuesto, de preocuparse únicamente por colocar al español en un lugar mejor que el que ocupa actualmente, sino de actuar en todos los frentes necesarios para cambiar las causas que dan lugar a esa situación. En el marco general de la situación económica y cultural del mundo hispánico, hay tres grandes factores a los que es necesario prestar atención. De un lado, la mejora de las condiciones técnicas y económicas de acceso a la Red. En segundo lugar, un esfuerzo adicional en la situación de contenidos realmente atractivos para la población que, en buena parte, es un resultado que se obtiene como consecuencia de la mayor facilidad de acceso. Por último, en lugar de escudarnos en la idea de que los buscadores comerciales prestan poca atención a los contenidos procedentes del mundo hispánico ?lo cual puede ser cierto?, es necesario dedicar más medios a la producción de herramientas y recursos lingüísticos basados en el español que aporten a la documentación existente en la Red y a su recuperación la ´inteligencia´ lingüística de la que, en términos generales, carecen en este momento, lo cual implica tanto la investigación básica como la aplicada. Sólo la conjunción de esfuerzos en las tres direcciones permitirá que la Red en español esté al nivel que la comunidad hispanohablante quiere tener en el futuro.

Referencias bibliográficas

AGUILLO, I. F., ORTEGA, J. L. Y GRANADINO, B.: «Contenidos del buscador *Google*. Distribución por países, dominios e idiomas», *El profesional de la información*, 15 de mayo de 2006, págs. 384-389.

AGUILLO, I. F., GARCÍA, I. Y ARROYO, N.: «Regional and linguistic patterns in *Google*

Fundación Telefónica
Telefónica

positioning», www.csi.ensmp.fr/WebCSI/4S/download_paper/download_paper.php?paper=aguillo_garcia_arroyo.pdf. 2004.

DÍAZ NOCI, J.: «Orígenes y desarrollo del ciberperiodismo en España: balance de una década», en *Informe anual de la profesión periodística 2006*, Madrid: Asociación de la Prensa de Madrid, 2006, págs. 307-319.

DÍAZ NOSTY, B.: «Las ediciones digitales de la prensa diaria en lengua española», *El español en el mundo. Anuario del Instituto Cervantes* 1999, págs. 65-186.

DÍAZ NOSTY, B. (dir.): «Cuarto informe sobre la prensa en lengua española en Internet», VI Foro Eurolatinoamericano de la Comunicación, Ciudad de Panamá, 2000, www.infoamerica.org/documentos_pdf/nosty07.pdf.

FUNDACIÓN CAJA DE BURGOS: *La difusión del español en Internet*, Burgos, 2006.

GULLI, A. y SIGNORINI, A.: «The Indexable Web is More than 11.5 billion pages», *WWW* 2005, www.cs.uiowa.edu/~asignori/web-size/.

MARCOS MARÍN, F.: «La lengua española en Internet», *El español en el mundo. Anuario del Instituto Cervantes*, 2000, págs. 299-357.

———: *Los retos del español*, Madrid / Frankfurt, Iberoamericana / Vervuert, 2006.

MARTÍN MAYORGA, D.: «El español en la sociedad de la información», *El español en el mundo. Anuario del Instituto Cervantes* 2000, págs. 359-374.

MILLÁN, J. A.: *Internet y el español*, Madrid, Fundación Retevisión, 2001.

OTERO, J. Y PERDIGUERO, H. (coords.): *El porvenir del español en la sociedad del conocimiento*, Burgos, Fundación Caja de Burgos, 2006.

PIMIENTA, D.: «Presencia de las lenguas neolatinas en la Internet», Congreso internacional sobre lenguas neolatinas en la comunicación especializada, (México, D.F. 2002), dtil.unilat.org/cong_com_esp/comunicaciones_es/pimienta.htm.

PIMIENTA, D. Y LAMEY, B.: «Lengua española y culturas hispánicas en la Internet. Comparación con el inglés y el francés», II Congreso internacional de la lengua española, Valladolid, 2001, www.cvc.cervantes.es/obref/congresos/valladolid/ponencias/el_espanol_en_la_sociedad/4_internet_en_espanol/pimienta_d.htm.

Páginas electrónicas citadas

Fundación Telefónica
Telefónica

Funredes: http://www.funredes.org

Google: http://www.google.com

Ethnologue: www.ethnologue.org

Internet Systems Consortium, Inc. (ISC): http://www.isc.org

Internet World Stats: http://www.internetworldstats/com

Ministerio de Industria, Turismo y Comercio:

http://www.mityc.es/es-ES/Servicios/IndicadoresE/sociedadinformacion

http://observatorio.red.es/indicadores/index.html

Nielsen//NetRatings: http://www.nielsen-netratings.com

Wikipedia: http://www.wikipedia.org

Fundación Telefónica