

## A.7. Métrica de las lenguas en internet

Por José Antonio Millán

Millán, José Antonio. "Métrica de las lenguas en internet". En: *Anuario ThinkEPI*, 2007, pp. 29-31.



**"El pionero de los estudios fue Daniel Pimienta desde la Fundación Redes y Desarrollo (Funredes)"**

**"La presencia relativa del inglés no ha dejado de bajar desde 1998 (cuando tenía el 75% de la red) hasta el 45% en la actualidad"**

**LA WEB PÚBLICA es un objeto de estudio abierto a todos (por definición).**

En una red dominada inicialmente por el inglés, los hablantes de otras lenguas con

pujanza cultural pronto intentaron descubrir cuál era su posición (Millán, 1997).

Los primeros estudios se desdoblaron en dos aspectos, los cuantitativos y los cualitativos. Desde el punto de vista cuantitativo interesaba ver qué cantidad de páginas había en cada lengua. Cualitativamente, importaba ver la presencia de determinados elementos de una cultura en las páginas web de distintas lenguas. El pionero de ambos estudios fue **Daniel Pimienta**, desde la *Fundación Redes y Desarrollo (Funredes, 1996a y b)*, inicialmente para el español y el francés (en relación con el inglés), y que hoy en día abarca (con el apoyo de *Unión Latina* y de la *Délégation générale à la langue française*) las otras lenguas romances y además el alemán. El catalán ha estado ausente de estas medidas, porque Unión Latina se ocupaba sólo de las lenguas oficiales de los estados miembros (en 2005 se adhirió Andorra) y porque ninguna institución catalana ha requerido hasta ahora la colaboración de *Funredes*.

INDICADORES	Inglés	Español	Francés	Italiano	Portugués	Ruso
Hablantes (millones)	630	375	130	60	190	30
Hablantes (%)	10,5	6,3	2,2	1	3,2	0,5
Internautas (millones)	295	72	34	30	24	4
Internautas por hablantes (%)	46,9	19,2	26,1	50,7	12,8	14,7
Internautas por población (%)	4,9	1,2	0,6	0,5	0,4	0,1
Internautas por lengua (%)	29	7	4	4	3	0,2
Webs (%)	45,0	4,6	5,0	3,0	1,9	0,2
Productividad de webs	1,57	0,66	1,14	0,81	0,55	0,70
Webs por hablantes	4,29	0,74	2,28	3,05	0,59	0,33

Fuentes: *Global Reach* y *FunRedes*  
[http://funredes.org/presentation/e\\_democratie/](http://funredes.org/presentation/e_democratie/)

[http://funredes.org/presentation/e\\_democratie/](http://funredes.org/presentation/e_democratie/)

Para el estudio cuantitativo, que es el que nos ocupará, **Pimienta** utilizó inicialmente el buscador *Altavista* y un algoritmo de reconocimiento de lenguas (mediante palabras exclusivas de cada lengua) para determinar las cifras totales (*Funredes*, 1996a). Estudios posteriores como el de *Alis Technologies* (1997), con el apoyo de *Internet Society*, utilizaron un muestreo muy pequeño de sitios.

Las medidas, más fiables, que se extraían utilizando buscadores generales tenían el problema fundamental de saber si el buscador indizaba la totalidad del espacio web, o si se centraba en el conjunto más antiguo de páginas, dando poca cuenta de las nuevas incorporaciones a la Web, y primando de esta manera las cifras de las páginas en inglés (para una perspectiva global de la situación en aquel momento de los buscadores con respecto a la totalidad véase **Millán**, 2000). El catalán –que se había iniciado en la Red muy tempranamente: **Partal**, 2004– acogió una iniciativa precoz para decir a los buscadores en qué lengua estaban sus sitios, mediante etiquetas meta (**Yzaguirre**, 1997), como forma de aumentar su visibilidad.

### Mejora progresiva de las mediciones

La evolución de los estudios con una misma metodología (los de *Funredes*, véase la página resumen: 2005) a lo largo de los distintos años ha conducido a refinar no sólo el uso simultáneo de varios buscadores y los procedimientos de cálculo, sino también las medidas absolutas, poniéndolas en relación con el número de hablantes y también con el número de personas con conexión a la Red (productividad de un espacio web). Estos parámetros se pueden utilizar también para cada país, lo que proporciona una visión muy rica de comunidades geográficamente dispersas como la hispanohablante.

Según el estudio en vías de publicación, del que he recibido un adelanto provisional (**Pimienta**, 2005), la presencia relativa del inglés no ha dejado de bajar desde 1998 (cuando tenía el 75% de la red) hasta el 45% en la actualidad. El resto de las lenguas estarían, RESPECTO AL INGLÉS, así: alemán 15,42%, francés 11,00%, español 10,23%, italiano 6,77%, portugués 4,15%, rumano 0,37%. La novedad de las cifras anteriores

(2003) es que el español pierde el primer lugar de las lenguas romances frente al francés.

Con casi una década de datos tal vez estemos en situación de aventurar evoluciones cuantitativas para las lenguas que se incorporan a la Red: inicios de crecimientos rápidos, seguidos de mesetas de estabilización. El inglés se estabilizó hace años, el español habría alcanzado ya una de esas mesetas y el francés estaría llegando a la suya. Pero para estas y otras muchas cosas es mejor esperar la publicación definitiva del último estudio de *Funredes*. Y en cuanto a los estudios cualitativos... pueden ser objeto de otra nota.

### Referencias

*Alis Technologies. Palmarès des langues de la toile*, 1997.

<http://alis.isoc.org/palmares.html>

*Funredes. El primer estudio de las lenguas y la Internet*, 1996a.

<http://funredes.org/LC/espanol/L1.html>

*Funredes. El primer estudio de la cultura y la Internet*, 1996b.

<http://funredes.org/LC/espanol/C1.html>

*Funredes. Observatorio de las lenguas y culturas*, 2005.

<http://funredes.org/espanol/conexos.php3/docid/270>

**Millán, José A.** "El español en las redes globales". En: *Actas del I Congreso Internacional de la Lengua Española*, 1997.

<http://cvc.cervantes.es/obref/congresos/zacatecas/tecnologias/ponencias/millan.htm>

**Millán, José A.** "El libro de mil millones de páginas. La ecología lingüística de la Web". En: *Revista de Libros*, Madrid, nº 45, 2000. Versión ampliada en la web:

<http://jamillan.com/ecoling.htm>

**Partal, Vicent.** *El català a la xarxa: història i raons d'un cas d'èxit*, 2004.

<http://www.softcatala.org/articles/article39.htm>

**Pimienta, Daniel.** *Comunicación personal*, 2005.

<http://funredes.org/pimienta/>

**Yzaguirre, Lluís de.** *Indexem en català*, 1997.

<http://www.iula.upf.es/altres/pub/indexem/indexem.htm>

**José Antonio Millán**

[librosybitios@jamillan.com](mailto:librosybitios@jamillan.com)

<http://jamillan.com/librosybitios/blog/>

<http://librosybitios.com>

Haga que los contenidos de sus documentos  
y metadatos sean verdaderamente  
visibles y recolectables en la Web con  
**Repositorios y Recolectores**  
basados en la

**Open Archives Initiative (OAI)**

que hemos desarrollado y podemos  
ajustar a sus necesidades



Empresa pionera en el campo del tratamiento de la  
información y el conocimiento aplicado a  
**Archivos, Bibliotecas y Museos** en un entorno  
Internet, mediante desarrollos informáticos propios,  
de **Gestión Documental**, basados en XML.

Producimos **Bibliotecas Virtuales** en colaboración  
con diversas instituciones.

Visítenos en

[www.digibis.com](http://www.digibis.com)