

El multilingüisme a Internet

Miquel Strubell Trueta

PID_00169661



Universitat Oberta
de Catalunya

www.uoc.edu

Índex

1. "Internet, chance ou menace pour la diversité culturelle et linguistique".....	5
2. Evolució de la presència de les llengües a Internet.....	7
2.1. La presència de les llengües a Internet	7
2.2. L'ús de diferents grafies a Internet	18
2.3. Els noms de domini internacionalitzats	20
2.4. Els dominis de primer nivell i el cas del domini català	20
2.5. La presència del català a Internet	23
2.5.1. Llocs i pàgines web	24
2.5.2. Espais interactius i participatius	26
2.6. Les llengües i els usuaris	28
Resum.....	35

1. "Internet, chance ou menace pour la diversité culturelle et linguistique"

Amrous expressa de manera clara la gran pregunta que es fan molts estudiosos del fenomen de les TIC: ofereixen noves possibilitats per al diàleg de les cultures i per a poder accedir a un coneixement cada cop més ric, certament; però a continuació afirma que –com en tota revolució tecnològica– susciten noves inquietuds:

"Il s'agit de menaces qui pèsent sur le respect des cultures et des langues et du risque d'uniformisation culturelle. En effet, la barrière langagière et les risques de prédominance d'une seule langue de communication représentent un des défis majeurs pour les réseaux électroniques."

Naila Amrous (2006, 5 de març). "Internet, chance ou menace pour la diversité culturelle et linguistique ?". *EspacesTemps.net*.

L'autora, que escriu des de l'òptica del món subdesenvolupat, proposa mesures de promoció en el camp tècnic, en el camp educatiu i en el camp de la cooperació internacional. Arriba a la conclusió que Internet afavoreix el multilingüisme sota certes condicions i, especialment, si les tecnologies es posen a disposició de tots els pobles –en particular, dels més dispersats demogràficament– mitjançant una política que afavoreixi i democratitzi l'accés a aquesta tecnologia.

Hi ha un altre discurs, però. Egia argumenta que, mirant el cas de la llengua basca a Internet, hom hi pot trobar avantatges pel fet de ser (demogràficament) petita. Exposa tres propostes, totes tres difícilment assumibles per una gran comunitat lingüística:

Vegeu també

Parlarem d'Internet en el món subdesenvolupat i de la divisió digital en el capítol "Informe ¿Cómo garantizar la presencia de una lengua en el ciberespacio? (2007)".

"¿Imaginan ustedes la ingente tarea que puede suponer el digitalizar toda la literatura escrita en lengua inglesa, española o francesa, pongamos hasta el año 1950? Para el inglés hay ya un intento serio y organizado, llamado Proyecto Gutenberg (<http://promo.net/pg/>), que seguro muchos de ustedes conocen. En cualquier caso, es empresa ardua, difícil de abordar y de resultados inciertos. Pero, ¿y para el euskera? ¿Cuánto tiempo, esfuerzo y dinero podría costarnos disponer de todos los textos publicados en lengua vasca hasta el año 1950? Relativamente poco, eso es seguro.

Otro tanto podríamos decir de la actividad de investigación académica desarrollada en lengua vasca. No creo que el disponer en dominio público de todas las tesinas de licenciatura y tesis doctorales realizadas en euskera suponga un trabajo ciclópeo. A cambio, ello podría proporcionar a la comunidad académica e investigadora un acervo de incalculable valor, capaz de generar a buen seguro una mayor actividad en lengua vasca en este terreno.

La difusión de información general en euskera por medio de Internet va a estar sujeta, sin duda, a los medios de comunicación públicos y privados que actualmente realizan su labor en lengua vasca. En este terreno, tampoco es esperable que nadie se vaya a ocupar de nosotros. Por tanto, contando con lo que ya hay, [...] bien podrían ensayarse fórmulas de colaboración mutua entre los diversos medios, para llegar a una especie de portal de la información en euskera, con contenidos multimedia provenientes de fuentes diversas, que aunara a la información general los contenidos de la información local."

Gotzon Egia (2002). "Las ventajas de ser pequeño: notas sobre la presencia de la lengua vasca en Internet". Congreso Mundial sobre Políticas Lingüísticas.

I, finalment, els especialistes mateixos han arribat a la conclusió que el risc d'una Internet unilingüe s'ha esvaït. Els dies 9 i 10 de març de 2001 es va celebrar a París un col·loqui internacional sobre "Le plurilinguisme dans la société de l'information", organitzat per la Commission Nationale Française pour l'Unesco, en cooperació amb la Unesco i l'Agència Intergovernamental de Països Francòfons. En el document de síntesi es constata que a Internet ocorre el següent:

<p>The number of languages has increased, and the fear of monolingualism on the Net has been prevented through its extension to a great number of linguistic communities, even if we are still far from a fair representation of the different languages on the Internet.</p>	<p>Les langues se sont multipliées, les craintes d'un monolinguisme de l'internet ont cédé devant son extension à un grand nombre de communautés linguistiques, même si nous sommes loin d'une représentation équitable des langues.</p>
---	--

Unesco (2001, març). *Colloque International le Plurilinguisme dans la Societe de l'Information. Synthese*. Commission Nationale Française pour l'Unesco. Paris: Unesco. (També disponible en anglès: *International Symposium Language Diversity in the Information Society. Final Report*.)

Referència bibliogràfica

Els interessats en la incorporació de la llengua basca en la tecnologia lingüística poden llegir l'article següent:

Itziar Aduriz; Arantza Diaz de Ilarraza; Kepa Sarasola (2009). "Basc i tecnologia lingüística". *Llengua, Societat i Comunicació* (núm. 7, pàg. 35-43).

2. Evolució de la presència de les llengües a Internet

Abans d'entrar a considerar els temes relacionats amb la gestió pròpiament dita del multilingüisme, serà interessant estudiar la posició relativa i, en la mesura del possible, l'evolució de la presència del català, el castellà i altres llengües dins el conjunt del ciberespai.

És notori que Internet va començar enllaçant institucions que funcionaven, a tots els efectes, de manera monolingüe en anglès. En veurem algunes de les conseqüències, començant per les limitacions de l'estàndard ASCII. Korpela manté un lloc web des de 1995 en què contesta afirmativament la pregunta que dona títol a la seva aportació, *English, the universal language on the Internet?*, i intenta donar-hi una explicació que no parteix, en canvi, d'aquesta evidència, segurament perquè escriu al cap de molt poc temps de l'aparició del World Wide Web, motiu pel qual no té una perspectiva històrica del fenomen.

2.1. La presència de les llengües a Internet

Entre els primers estudis publicats amb la intenció de veure el grau de diversitat lingüística a Internet trobem l'equip Babel, una iniciativa conjunta d'Alis Technologies i la Internet Society. L'informe d'aquest equip, que data de 1997, es basa en un mètode que inclou la verificació manual, d'una a una, de la llengua d'una mostra de 3.239 pàgines web: un procés primitiu i avorrit! Només van analitzar la presència de les disset llengües que el programari de detecció d'aleshores era capaç d'identificar. El resultat va ser clar: un 82,3% de portals web era en anglès, seguit d'un 4,0% en alemany, un 1,6% en japonès i un 1,5% en francès. A continuació venia l'espanyol amb un 1,1% del total.

Lloc web

Qui vulgui més informació sobre la presència del castellà i altres llengües a Internet pot consultar la llista de treballs al lloc web de Funredes.



Portada de l'estudi sobre mesurament de llengües a Internet

El mesurament de la realitat en el ciberespai és molt complex, com ho demostra un treball de Daniel Pimienta, en el qual subratlla que per molt que els científics i el món acadèmic facin la seva feina, les empreses de màrqueting (i potser hi podríem afegir els mitjans de comunicació) s'entesten a no actualitzar les seves dades:

"Quite paradoxically, the world of networks, born and developed in universities, in a way surrendered measuring the language situation to marketing companies, whose intentions are different from those of scientific publication, and who are therefore not very concerned with documenting their methods. Disorder and confusion regarding the state of languages on the Internet has been the result, which can lead to disinformation. Therefore, while the proportion of English-language speakers who use the Internet has gone from more than 80% in the year the Web was born to 35% today, the figures circulating in the media, against all evidence, are reported as stable between 70% and 80%! It is urgent that the academic world regains its role in this area along with public institutions, both national and international."

Daniel Pimienta (2005). "Linguistic Diversity in Cyberspace – Models for Development and Measurement". A: Paolillo; Pimienta; Prado. *Measuring Linguistic Diversity on the Internet* (pàg. 27-28). París: Unesco Publications for the World Summit on the Information Society.

Irònicament, la URL de la font que Pimienta dóna com a més vàlida per a calcular el pes de les llengües a Internet en el seu article de 2005 actualment es readreça automàticament a *Guava A/S, a dynamic digital marketing agency!*

En un altre article, Pimienta repassa un a un els criteris necessaris perquè un intent de quantificar la presència relativa de les llengües a Internet doni un resultat vàlid.

Referència bibliogràfica

Daniel Pimienta (2009). "Douze ans de Mesure de la Diversité Linguistique dans l'Internet: Bilan et Perspectives". Conferència inaugural del Forum International de Bamako sur le Multilinguisme. Union Africaine / Acalan, Maaya. Bamako (Mali).

Amb la perspectiva que dóna un període de dotze anys, Pimienta, Prado i Blanco estudien l'evolució de la diversitat lingüística a Internet. Grefenstette i Nioche, que ells mateixos citen, afirmen el següent:

Referència bibliogràfica

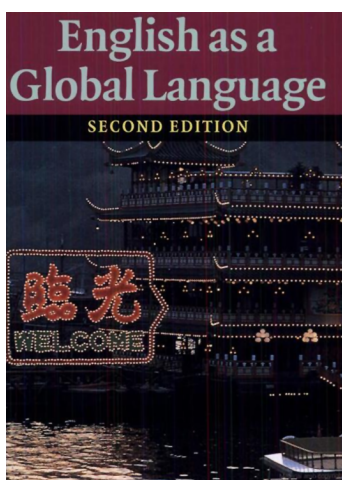
D. Pimienta; D. Prado; M. Blanco (2009). *Twelve years of measuring linguistic diversity in the Internet: balance and perspectives*. París: United Nations Educational, Scientific and Cultural Organization. (També disponible en castellà: *Doce años midiendo la diversidad lingüística en Internet: balance y perspectivas*.)

"One of the most easily remarkable aspects of the WWW from a content perspective is its multilinguality (Nunberg, 1996). A number of studies of language use, often involving some type of automatic language detection (Grefenstette, 1995), have been performed. David Crystal (1997) retrieved pages containing international words (such as Internet) and calculated that about 80% of those pages were English.

[...]

Jack Xu (2000) reports on experiments applying a language identifier to web pages indexed by the Excite portal that finds that 71% of the pages (453 million out of 634 million web pages) are written in English. About 29% is written in languages other than English. Breaking down the pages by language gives English (71%), Japanese (6.8%), German (5.1%), French (1.8%), Chinese (1.5%), Spanish (1.1%), Italian (0.9%), and Swedish (0.7%)."

G. Grefenstette; J. Nioche (2000). *Estimation of English and non-English Language use on the WWW*. Technical Report from Xerox Research Centre Europe.



Amb dades recollides el febrer de 2000, Grefenstette i Nioche utilitzen un algoritme de recompte de paraules en diferents llengües per tal d'arribar als resultats següents (noteu que en aquesta relació només hi ha trenta-dues llengües, totes escrites amb caràcters llatins):

		Word count estimate
1	Anglès	47.264.700.000
2	Alemanys	3.068.760.000
3	Francès	2.208.418.000
4	Espanyol	1.595.489.000
5	Italià	1.240.205.000
6	Portuguès	924.965.000
7	Suec	644.740.000
8	Neerlandès	622.063.000
9	Noruec	453.391.000

Estimació del nombre de mots disponibles en una sèrie de llengües escrites amb caràcters llatins per mitjà del portal Altavista. Font: G. Grefenstette, J. Nioche (2000). *Estimation of English and non-English Language use on the WWW*. Technical Report from Xerox Research Centre Europe (pàg. 8)

		Word count estimate
10	Txec	269.310.000
11	Hongarès	268.944.000
12	Polonès	235.726.000
13	Danès	206.167.000
14	Finès	192.105.000
15	Eslovac	140.909.000
16	Català	126.324.000
17	Malai	113.236.000
18	Turc	100.548.000
19	Eslovè	74.998.000
20	Croat	72.122.000
21	Romanès	63.846.000
22	Islandès	53.167.000
23	Irlandès	49.778.000
24	Estonià	43.257.000
25	Llatí	38.256.000
26	Èuscar	28.296.000
27	Esperanto	26.795.000
28	Letó	21.925.000
29	Lituà	20.927.000
30	Bretó	9.975.000
31	Albanès	9.203.000
32	Gal·lès	7.590.000

Estimació del nombre de mots disponibles en una sèrie de llengües escrites amb caràcters llatins per mitjà del portal Altavista. Font: G. Grefenstette, J. Nioche (2000). *Estimation of English and non-English Language use on the WWW*. Technical Report from Xerox Research Centre Europe (pàg. 8)

Veiem com predomina de manera molt clara la presència de l'anglès (en aquell moment tenia 15,4 vegades més paraules que l'idioma situat en segona posició, l'alemany). Però també és interessant veure l'espanyol, en quart lloc, molt per darrere de llengües amb una demografia força més reduïda, i el català, en setzè lloc, per davant de diverses llengües estatals europees, però per darrere de llengües europees amb una demografia comparable, com ara el suec, l'hongarès o el noruec.

Quina ha estat l'evolució d'aquestes llengües a la WWW? En un estudi que aplicava el mateix mètode de mesurament al llarg de tres anys i mig, Grefenstette i Nioche troben que mentre que l'anglès ha crescut un 800% en aquest període, les altres llengües encara han crescut més (l'alemany, un 1.500%, i l'espanyol, un 1.800%) en el mateix espai de temps.

	Octubre de 1996	Ràtio amb l'anglès	Agost de 1999	Ràtio amb l'anglès	Febrer de 2000	Ràtio amb l'anglès
Anglès	6.082.090.000	1,000	28.222.100.000	1,000	48.064.100.000	1,000
Alemany	228.938.428	0,038	1.994.229.409	0,071	3.333.127.671	0,069
Francès	223.316.023	0,037	1.529.795.169	0,054	2.732.221.327	0,057
Espanyol	104.319.158	0,017	1.125.646.460	0,040	1.894.966.981	0,039
Italià	123.555.682	0,020	817.270.444	0,029	1.338.351.674	0,028
Portuguès	106.167.245	0,017	589.391.943	0,021	1.161.898.076	0,024
Noruec	106.497.066	0,018	669.331.120	0,024	947.486.593	0,020
Finès	20.647.404	0,003	107.260.274	0,004	166.599.467	0,003

Estimació de la dimensió de cada llengua el 1996, al final de 1999 i al principi de 2002. G. Grefenstette, J. Nioche (2000). *Estimation of English and non-English Language use on the WWW*. Technical Report from Xerox Research Centre Europe (pàg. 9)

Fins i tot aleshores hi havia dubtes sobre la validesa de motors de cerca com Altavista, que segons alguns només arribaven a cobrir un 16% dels llocs web. En una comunicació personal (30 d'abril de 2010) Daniel Pardo afirma que Google ha deixat de recollir tots els llocs web del món, i que els criteris seguits per tal de fer la tria introdueixen biaixos segons la llengua que fan dubtar de la validesa de l'estadística lingüística actual.

En definitiva, després d'uns inicis en què gairebé tot era en anglès, diferents estudiosos conclouen que aproximadament un 71% de les pàgines estudiades (incloent-hi pàgines en llengües no llatines) eren en anglès i el 29% restant, en altres llengües.

O'Neill, Lavoie i Bennett troben una estabilitat relativa en la presència de l'anglès entre 1999 i 2002, amb un 72% del total. Entre les altres llengües, el japonès pujava del quart lloc (amb un 3%) al tercer lloc (amb un 6%), just darrere de l'alemany.

Un català de l'equip de Softcatalà, Jordi Mas, ha estudiat en diverses ocasions el que anomena la salut del català a Internet. Es poden consultar en línia els estudis amb dades recollides seguint un mateix mètode, respecte del 2003 i del 2005.

Mas utilitza les dades obtingudes per mitjà del lloc AlltheWeb en aquest període i, pel que fa al nombre absolut de pàgines web per llengua, el català passa de la vint-i-tresena posició a la vint-i-sisena i continua entre les trenta llengües

Referència bibliogràfica

E. T. O'Neill; B. F. Lavoie; R. Bennett (2003). "Trends in the Evolution of the Public Web: 1998 - 2002". *D-Lib Magazine* (vol. 9, núm. 4).

amb més presència a la xarxa. L'autor afegeix que "la popularització dels blocs en llengua catalana ha comportat també un increment important del nombre de pàgines disponibles".

	Cercador Google: Llengua	Pàgines l'octubre 2005	Pàgines l'agost 2003
1	Anglès	9.200.000.000	1.280.041.397
2	Francès	202.000.000	99.737.704
3	Alemanys	184.000.000	182.005.546
4	Japonès	150.000.000	69.730.375
5	Castellà	133.000.000	65.814.567
6	Xinès	74.700.000	65.730.212
7	Italià	68.400.000	41.849.365
8	Holandès	54.300.000	41.119.851
9	Portuguès	44.000.000	37.695.762
10	Rus	43.200.000	42.276.247
11	Suec	42.100.000	14.901.968
12	Noruec	28.700.000	8.123.301
13	Txec	26.700.000	15.580.583
14	Finès	26.200.000	5.678.599
15	Coreà	24.400.000	64.606.324
16	Polonès	23.800.000	22.154.325
17	Danès	22.200.000	12.107.133
18	Hongarès	15.100.000	8.540.941
19	Àrab	12.600.000	2.470.616
20	Turc	11.200.000	4.704.212
21	Hebreu	10.100.000	4.792.646
22	Grec	9.230.000	2.366.733
23	Indonesi	8.660.000	1.044.038
24	Croat	7.950.000	1.672.582

Referències bibliogràfiques

J. Mas i Hernández (2003). "La salut del català a Internet". *Softcatalà*.

J. Mas i Hernández (2005). "La salut del català a Internet". *Softcatalà*.

	Cercador Google: Llengua	Pàgines l'octubre 2005	Pàgines l'agost 2003
25	Romanès	7.250.000	2.052.990
26	Català	7.140.000	2.926.550
27	Eslovè	6.950.000	1.685.426
28	Estonià	6.510.000	1.464.539
29	Eslovac	6.130.000	5.077.965
30	Letó	5.750.000	559.913
31	Lituà	5.720.000	1.078.341
32	Islandès	5.250.000	1.391.302
33	Búlgar	3.810.000	1.120.713
34	Serbi	3.340.000	42.848

Respecte del català, el francès tenia 28,3 vegades més pàgines, l'espanyol 18,6, l'italià 9,6, el portuguès 6,2 i el romanès pràcticament les mateixes. He triat dades sobre llengües romàniques per una raó que descobrirem ben aviat.

Una font més recent fa referència a un altre sistema de mesurament, molt més parcial, l'Open Directory Project, que pretén crear el directori més ampli del web "confiant en un gran nombre d'editors voluntaris". La possible satisfacció de trobar el català en el dotzè lloc queda matisada pel fet que, per exemple, pràcticament no hi hagi webs en alemany dins aquest projecte. La seva classificació és la següent:

Vegeu també

Més endavant comentarem altres iniciatives multilingües a Internet basades en el voluntariat, en tant que exemples de gestió interessants en diferent llengües; parlarem de la *Wikipedia* en el capítol "La presència del català a Internet" i de Softcatalà, en el capítol homònim.

1	Francès	244.011
2	Italià	194.570
3	Japonès	188.003
4	Espanyol	150.900
5	Neerlandès	96.332
6	Polonès	77.122
7	Rus	74.417
8	Danès	46.293
9	Xinès simplificat	43.521
10	Turc	43.486

11	Suec	37.199
12	Català	35.621
13	Txec	24.861
14	Romanès	17.865
15	Portuguès	17.777
16	Noruec	14.541
17	Xinès tradicional	13.401
18	Finès	10.530
19	Hongarès	10.282
20	Àrab	7.611

De fet, sembla que Google s'ha apropiat de les dades de dmoz. El 22 d'abril de 2010 ofería dades sobre 47.441 pàgines web en català –és a dir, més que les que acabem de citar–, per categories, seguint el model i l'estructura de l'Open Directory Project:

Directorio de Google http://www.google.com/Top/World/Català/		
Categoria	Recursos	%
Regional	24.918	52,52%
Arts i Cultura	6.663	14,04%
Economia i empresa	2.775	5,85%
Societat	2.591	5,46%
Ensenyament	1.712	3,61%
Esports	1.596	3,36%
Ciència i tecnologia	1.535	3,24%
Llengua catalana	1.276	2,69%
Medicina i salut	1.025	2,16%
Política i govern	858	1,81%
Lleure	837	1,76%
Informàtica	651	1,37%
Mitjans de comunicació	554	1,17%
Referència	450	0,95%
Total	47.441	100,00%

Vegeu també

Comentarem amb més detall la presència del català en diferents categories, segons l'evolució de l'anàlisi mensual que en fa el WICCAC, en el capítol "La presència del català a Internet".

Referència bibliogràfica

Funredes (2008). *Observatorio sobre la diversidad lingüística y cultural en la Internet*.

Entre els estudis més recents n'hi ha un de Funredes de 2008, que concentra l'atenció en la posició d'algunes llengües romàniques amb relació a l'anglès i que, de fet, segueix una sèrie. El 2006 l'estudi *Lengua y cultura* es va transformar en l'*Observatorio sobre la diversidad lingüística y cultural en la Internet*. El maig de 2008, com dèiem, l'*Observatorio* va publicar les dades corresponents al mes de novembre de 2007, amb una novetat addicional: la incorporació del català al costat de les altres llengües de l'estudi. La manera de presentar les dades és curiosa, però s'entén bé:

Percentatge de pàgines web respecte de l'anglès	
Espanyol	8,45%
Francès	9,80%
Italià	5,92%
Portuguès	3,09%
Romanès	0,63%
Alemany	11,97%
Català	0,30%

Font: http://dtiil.unilat.org/LI/2007/index_es.htm

Funredes

La Fundació Xarxes i Desenvolupament és una ONG internacional dedicada a la difusió de les noves tecnologies de la informació i la comunicació (NTIC) en els països en vies de desenvolupament, en particular a l'Amèrica Llatina i el Carib.



Segons aquesta font, doncs, respecte del català, el francès tenia 33 vegades més pàgines, l'espanyol 28, l'italià gairebé 20, el portuguès 10 i el romanès el doble. Són xifres prou diferents de les obtingudes amb AlltheWeb per Mas i que hem vist més amunt, excepte respecte del francès. Comparat amb els casos de l'espanyol, l'italià, el portuguès i el romanès, el català sembla clarament més feble en aquest estudi de Funredes (que l'inclou, a més, per primer cop).

La mateixa fundació dóna dades interessants sobre l'evolució relativa de la presència d'algunes llengües a Internet entre 1998 i 2005. Reproduïm literalment la informació que ofereixen en el seu document de síntesi:

Presencia relativa de las lenguas neolatinas, y del alemán, con respecto al inglés en la Red

El siguiente cuadro muestra la evolución de la presencia **relativa**, con respecto al inglés, de las lenguas neolatinas y del alemán en la Red. Son resultados brutos y fiables del estudio.

	Para 100 páginas en inglés:					
	Español	Francés	Italiano	Portugués	Rumano	Alemán
Sep-98	3,37%	3,75%	2,00%	1,09%	0,20%	5,00%
Ago-00	8,41%	7,33%	4,60%	3,95%	0,37%	11,00%
Ago-01	11,24%	9,13%	6,15%	5,57%	0,35%	13,74%
Feb-02	11,80%	9,60%	6,51%	5,62%	0,33%	14,41%
Feb-03	10,83%	8,82%	5,28%	4,55%	0,23%	13,87%
May-04	10,19%	10,64%	6,15%	4,02%	0,31%	15,37%
May-05	10,23%	11,00%	6,77%	4,15%	0,37%	15,42%

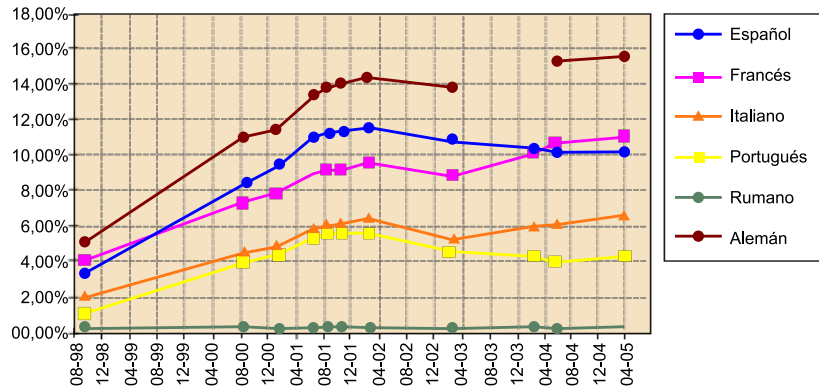
Presencia absoluta de las lenguas en la red

A partir de los datos **relativos** de las lenguas neolatinas, de la evolución de estos resultados y de una aproximación de la presencia de las lenguas no estudiadas, se deduce la presencia **absoluta** de las lenguas en la red. Ésta constituye una aproximación basada en una hipótesis no verificable sobre el peso absoluto del inglés en la red.

	Inglés	Español	Francés	Italiano	Portugués	Rumano	Alemán	Resto
Sep-98	75,00%	2,53%	2,81%	1,50%	0,82%	0,15%	3,75%	13,44%
Ago-00	60,00%	5,05%	4,40%	2,76%	2,37%	0,22%	3,00%	22,20%
Ago-01	55,00%	5,20%	4,34%	2,71%	2,44%	0,18%	6,29%	25,45%
Feb-02	50,00%	5,80%	4,80%	3,26%	2,81%	0,17%	7,21%	25,97%
Feb-03	49,00%	5,31%	4,32%	2,59%	2,23%	0,11%	6,80%	29,65%
May-04	46,30%	4,72%	4,93%	2,85%	1,86%	0,14%	7,12%	32,09%
May-05	45,00%	4,60%	4,95%	3,05%	1,87%	0,17%	6,94%	33,43%

Funredes resumeix aquestes i més dades en forma gràfica, així:

Tendencia de la evolución de las lenguas neolatinas en la red

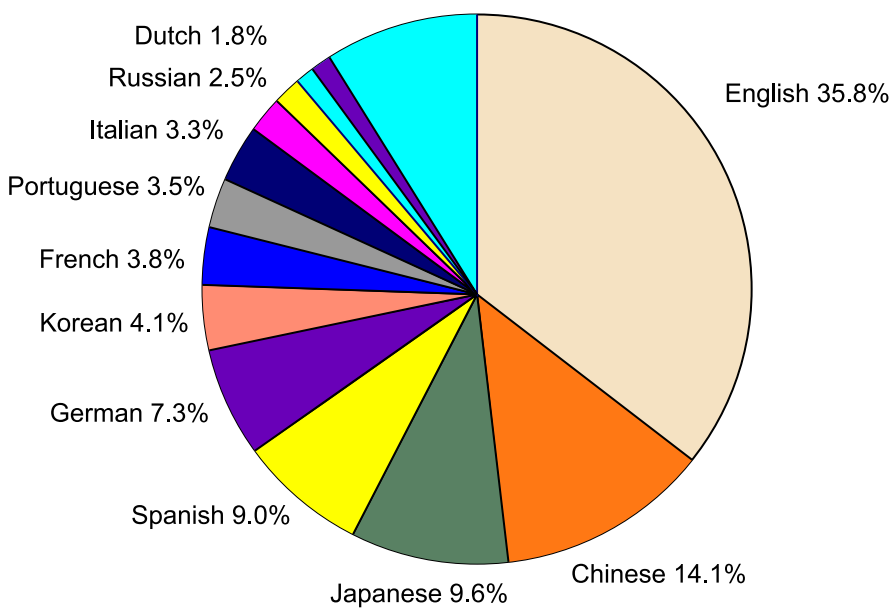


	09-98	08-00	01-01	06-01	08-01	10-01	02-02	02-03	02-04	05-04	03-05
—●— Español	3,37%	8,41%	9,46%	10,95%	11,24%	11,36%	11,60%	10,83%	10,30%	10,19%	10,23%
—■— Francés	3,75%	7,33%	7,89%	8,86%	9,13%	9,14%	9,60%	8,82%	10,18%	10,64%	11,00%
—▲— Italiano	2,00%	4,60%	4,93%	5,88%	6,15%	6,15%	6,51%	5,28%	6,09%	6,15%	6,77%
—■— Portugués	1,09%	3,95%	4,44%	5,40%	5,57%	5,61%	5,62%	4,55%	4,36%	4,02%	4,15%
—●— Rumano	0,20%	0,37%	0,33%	0,32%	0,35%	0,36%	0,33%	0,23%	0,41%	0,31%	0,37%
—●— Alemán	5%	11,00%	11,43%	13,42%	13,74%	14,08%	14,41%	13,87%		15,37%	15,42%

Tot i els mètodes criticables, val a dir que a hores d'ara no hi ha cap estadística vàlida sobre la presència relativa de les llengües a Internet.

Quant al nombre d'usuaris de cada comunitat lingüística, l'empresa PeakStats en fa una estimació. Al seu lloc web, però, les dades són de 2004 i estima en 729 milions el nombre d'internautes, una xifra superada llargament d'aleshores ençà:

Online Language Populations
Total: 729 Million
(March. 2004)



Font: http://www.peakstats.net/online_language_populations.htm

En general, els sistemes informàtics actuals fan servir el codi ASCII o una extensió compatible, amb dues finalitats. En primer lloc, com hem vist, s'utilitza per a la representació de textos i, en segon lloc, per al control de dispositius que processen text.

Pel que fa al català, el punt volat de la *ela* geminada (l·l) inicialment presentava problemes (com també en va tenir a les màquines d'escriure elèctriques fins que Olivetti va resoldre la qüestió ben entrats els anys vuitanta). Hi ha altres signes diacrítics i lletres, però, que el català comparteix amb la llengua francesa (*ç, à, è, ò, l'apòstrof...*) i/o amb el castellà (*é, í, ó, ú*). Per al castellà, el problema de la *ñ* és sentit de forma tan aguda que s'ha acabat convertint en un logo de la mateixa llengua.



La *ena* palatal (*eñe*)

Fa poc que ha començat a ser possible l'ús de la lletra *ñ* en els noms dels dominis (sembla que els primers apareixen el desembre de 2007, com ara <http://www.pañuelo.es>). Els podem trobar als llistats mensuals que publica el lloc web dominios.es en PDF, en l'apartat "Estadísticas" (tot i que el cercador d'Internet Explorer 8.0 –si més no, la versió en català– no hi detecta les *ñ*)! Hi consta, per exemple, <http://hostalcataluña.es/>, però també www.caixacatalua.es [sic]!

I si aquests problemes els detectem per a les llengües escrites amb lletres llatines, ens podem imaginar que la introducció a Internet de l'escriptura amb altres grafies ha presentat múltiples i variades problemàtiques. El repte és majúscul. Fixem-nos en les xifres que donen Mikami i altres sobre la distribució de la població mundial, segons els tipus d'escriptura de les seves llengües.

Esriptura	Llatina	Ciríl·lica	Àrab	Hanzi	Índica	Altres
Nombre d'usuaris en milions	2.238	451	462	1.085	807	129
[% del total]	43,28%	8,71%	8,93%	20,98%	15,61%	2,49%

Una ajuda per a l'escriptura xinesa

L'anomenat Los Angeles Chinese Learning Center ha hagut d'elaborar i penjar a Internet un full orientatiu per a explicar "How to Read and Type Chinese Characters on Your Computer and Chinese Character Input Methods".

La llengua grega i les llengües eslaves escrites amb grafia ciríl·lica són un primer graó en aquesta jerarquia de complexitat creixent (que reproduïx, en línies generals, l'evolució de la seva presència en teclats de màquina d'escriure). Tenen l'avantatge que s'escriuen d'esquerra a dreta i tenen un sistema de representació fonètic. L'àrab, a més de la seva família de signes, s'escriu i es llegeix

Referència bibliogràfica

Y. Mikami; A. Zaki Abu Bakar; V. Sonlert-Lamvannich; O. Vikas; P. Zavorsky; M. Zaidi Abdul Rozan; J. Göndri-Nagy; T. Takahashi (2005). "Language Diversity On The Internet: An Asian View". A: Paolillo; Pimienta; Prado. *Measuring Linguistic Diversity on the Internet* (pàg. 94). París: UNESCO Publications for the World Summit on the Information Society.

Altres alfabetes

Qui vulgui practicar el seu portuguès llegit, podrà trobar molta informació sobre aquestes i altres escriptures al lloc web de l'Institut Multimedia.

de dreta a esquerra; i el xinès, el coreà o el japonès, que tenen no pas dotzenes, sinó milers de caràcters que ja no segueixen el criteri bàsic fonètic, a més a més s'escriuen verticalment, en columnes.

En cada cas s'han anat preparant jocs de caràcters que o bé s'instal·len per defecte als ordinadors que es posen a la venda en els països corresponents, o bé es poden baixar d'Internet. Pel que fa als navegadors, es poden configurar per a llengües determinades, de vegades mitjançant el corresponent *language pack*, que s'ha de baixar.

Tot i que avui podem escriure en aquestes llengües sense més problemes (segons els estàndards fixats), cal adonar-se que fins fa molt poc les adreces electròniques i les URL¹ s'escriuen només i estrictament amb les vint-i-sis lletres de l'alfabet anglès, més el guionet, el guionet inferior, l'arrova (@) i el punt. És a dir, sense accents, apòstrofs ni diacrítics.

⁽¹⁾ **URL** significa *uniform resource locator*, mentre que **HTTP** significa *hypertext transfer protocol*.

2.3. Els noms de domini internacionalitzats

El projecte **Internationalized domain names** (IDN) té com a objectiu que es puguin registrar noms de domini representats per caràcters lingüístics locals, és a dir, noms de domini que puguin contenir lletres o caràcters d'escriptures fora del codi Unicode (i/o ASCII), com ara l'àrab o la xinesa. El promotor principal del projecte n'és el primer interessat, per la competència que hi té: l'ICANN.

Esforç per a l'estandardització dels noms de domini internacionalitzats

En el lloc web de l'ICANN podeu comprovar el ritme intens de reunions que manté aquest i altres organismes per tal de mirar d'establir els estàndards en l'àmbit dels noms de domini internacionalitzats.

Per tal d'assolir aquest objectiu, des del projecte es proposa que es pugui convertir la URL proposada en el format *punycode*² i que es permeti registrar aquesta versió.

⁽²⁾ El terme **punycode** és un joc de paraules en anglès. S'assembla molt a Unicode, però la paraula *puny* vol dir 'molt petit' i, en aquest cas, significaria que és molt més simple que la grafia en la llengua en qüestió.

2.4. Els dominis de primer nivell i el cas del domini català

Aquest no és el lloc per a explicar el sistema d'assignació dels dominis de primer nivell (TLD, *top level domains*), simplement perquè aquests dominis tenen molt poca relació directa amb la qüestió (multi)lingüística. Farem, però, una petita excepció per tal de presentar el cas del domini de primer nivell *.cat*.

L'esmentada assignació de dominis de primer nivell és tasca d'una organització internacional, la **Corporació d'Internet per a l'Assignació de Noms i Números** (ICANN, Internet Corporation for Assigned Names and Numbers).

Quin és el rol de l'ICANN? Ho expliquen al seu lloc web:



To reach another person on the Internet you have to type an address into your computer - a name or a number. That address has to be unique so computers know where to find each other. ICANN coordinates these unique identifiers across the world. Without that coordination we wouldn't have one global Internet.

ICANN was formed in 1998. It is a not-for-profit public-benefit corporation with participants from all over the world dedicated to keeping the Internet secure, stable and interoperable. It promotes competition and develops policy on the Internet's unique identifiers.

ICANN doesn't control content on the Internet. It cannot stop spam and it doesn't deal with access to the Internet. But through its coordination role of the Internet's naming system, it does have an important impact on the expansion and evolution of the Internet.

Per tant, no sembla que tingui competències en la gestió de la diversitat lingüística d'Internet. Aleshores, per què l'hem inclosa, aquesta corporació? Doncs perquè des de fa pocs anys ha reconegut una comunitat amb criteris lingüístics i culturals, com veurem tot seguit.

Els dominis de primer nivell es poden classificar en dues grans categories:

1) Els que volen un estat (com ara *.uk*, Regne Unit; *.it*, Itàlia; *.ad*, Andorra; *.cz*, República Txeca...) o un territori singular (*.gi*, Gibraltar; *.vg*, illes Verges britàniques; *.ps*, territori palestí ocupat...). En anglès es diuen *country code top level domains* (ccTLD).

2) Els que volen determinats tipus d'activitat o d'organització, els dominis de caire genèric (gTLD), com ara *.com*, *.net*, *.edu* o *.org*.

La idea d'un domini que representés la comunitat catalana (o catalanoparlant) va néixer pels volts de 1995, quan els primers internautes van començar a crear webs i grups de notícies, i a entendre el sistema de dominis. L'opció que inicialment semblava més fàcil, obtenir un ccTLD (un domini de dues lletres), es va veure obstaculitzada per la normativa internacional del moment. Finalment, quan l'ICANN va donar la possibilitat de crear dominis per a comunitats d'interès va ser factible demanar un domini de caire genèric (gTLD), com ara el *.com* o el *.net*, i s'optà pel *.cat* per a representar els catalanoparlants a Internet. La candidatura del *.cat* va superar amb escreix totes les exigències dels diversos comitès de selecció de l'ICANN. A més, de totes les candidatures que es van presentar aquell any, va ser la més ben puntuada per un grup independent d'experts japonesos.

El cas del català, doncs, trencava l'esquema gràcies a un procés de sol·licitud molt ben travat, liderat per un petit nucli d'organitzacions catalanes, entre les quals hi havia el Capítol Català de la Internet Society, bona part de la raó de ser del qual rau en la defensa dels interessos de la llengua catalana i els seus parlants. El procés ha estat descrit amb detall per Peter Gerrand (2006a, 2006b). La sol·licitud culmina la nit del 15 al 16 de setembre de 2005, quan l'ICANN aprova definitivament la candidatura del *.cat*. Al cap de pocs dies se signa el contracte entre aquest organisme regulador d'Internet i la Fundació puntCAT, la qual gestiona el registre del domini *.cat*. La Fundació, una enti-

Pàgina web

Hom pot accedir a la llista sencera d'aquests dominis a la pàgina web de l'Autoritat d'Assignació de Dominis d'Internet (IANA, Internet Assigned Numbers Authority).

tat privada sense ànim de lucre ni filiació política, havia estat fundada el 28 de desembre de 2004. La seva finalitat fundacional és promoure tota mena d'activitats relacionades amb la creació, la gestió i el control del registre per mitjà de *www.domini.cat* i, en general, promoure la llengua i la cultura catalanes en l'àmbit d'Internet i les noves tecnologies de la informació.

Per tal de reeixir, la sol·licitud del *.cat* no defensava pas un criteri territorial, perquè aleshores hauria caigut en criteris (geo)polítics, sinó més aviat un criteri lingüístic i cultural. El *.cat* es presentava com el domini que agrupava tots els llocs web que utilitzen la llengua catalana i/o s'adrecen a la comunitat catalanoparlant, o la representen.

Finalment, comentarem que un treball d'Atkinson (2006) explica el perquè del fracàs de la campanya per al *.ct*, que pretenia ser un domini territorial i que va ser bloquejada pel Govern espanyol, en contrast amb l'èxit del *.cat*.

L'èxit del *.cat* es pot mesurar en termes del nombre de dominis, o URL, atorgats per la Fundació puntCAT. Des que el registre del domini *.cat* va començar, el 13 de febrer de 2006, el nombre de dominis actius augmenta cada any i el 24 de gener de 2010 (és a dir, en menys de quatre anys) el *.cat* va superar els 40.000 dominis. Segons el lloc Pages Inventory, el domini *.cat* no queda gaire lluny, en el rànquing de dominis de primer nivell més populars, del domini *.ro* (Romania).

Data	Nombre
1/1/2007	19.693
1/1/2008	26.373
1/1/2009	33.410
1/1/2010	39.665
1/4/2010	41.537

Font: <http://www.domini.cat/es-tadistiques/index.html>

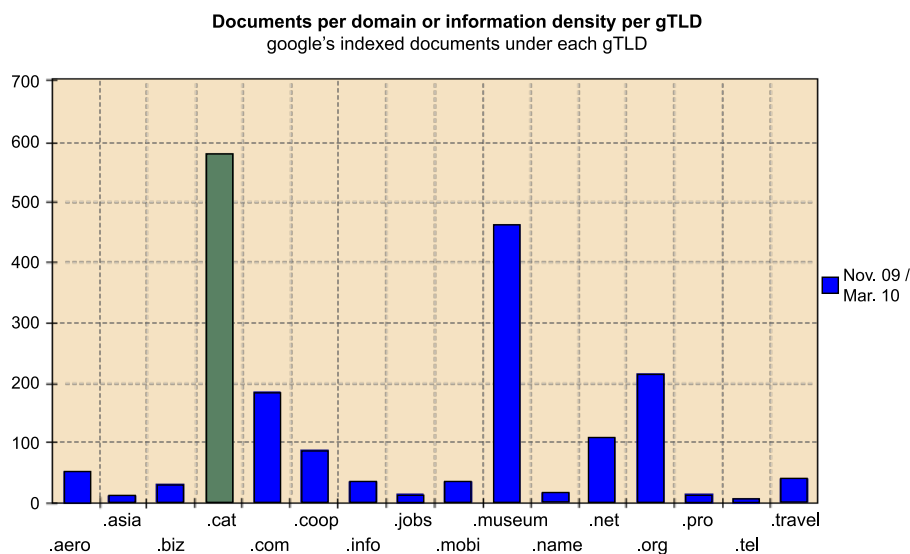
Un altre indicador de l'èxit del domini *.cat* és la qualitat del servei i de la informació que aporten els dominis. La Fundació puntCAT ha fet una comparativa de la densitat d'informació aportada pels diferents dominis genèrics (dividint la quantitat de documents indexats per un buscador sota el domini pel nombre de dominis d'aquest mateix domini) que mostra com, si s'agafen les dades dels documents indexats per Google el març de 2010 i els dominis reportats per l'ICANN el novembre de 2009, "el *.cat* se situa com a indiscutible número 1", molt per davant del *.com*, el *.org* o, fins i tot, el *.museum*.

Referència bibliogràfica

David Atkinson (2006). "Catalan on the Internet and the *.ct* and *.cat* campaigns". A: Claudio Menezes (editor). *Languages of the Internet. Special issue of Journal of Language and Politics* (pàg. 239-249). París: Unesco.

El domini *.es* a Catalunya

Segons un informe del Plan Avanza, el mes d'octubre de 1999 a Catalunya hi havia 167.700 dominis *.es* registrats.



Font: <http://www.domini.cat/estadistiques/index.html>

També es pot valorar pel nivell d'estímul que representa per a iniciatives semblants, com ara els gal·lesos, els bascos, els bretons i altres comunitats lingüístiques no identificades amb el poder de cap estat. Cal dir, però, que si la candidatura del *.cat* fou acceptada, en part va ser per l'excepcionalitat del cas, en termes demogràfics, i certament la incidència de les llengües de totes aquestes altres comunitats lingüístiques és molt inferior a la de la llengua catalana.

Així, els gallecs es demanen el següent:

Se Catalunya conseguiu o *.cat*... por que Galiza non ten un TLD de seu?

Os galegos levamos reclamando un dominio dedicado á nosa cultura e mais á nosa lingua desde hai máis de 10 anos (xa había campañas ao respecto en 1995).

Red Galega

2.5. La presència del català a Internet

Per què Vicent Partal, en parlar del català a la xarxa, pot parlar de la història i les raons "d'un cas d'èxit"? És una afirmació triomfalista o respon a una realitat? La seva conclusió encara és vàlida avui, després de més de sis anys?

En part atribueix l'èxit a l'evidència del pes de la presència del català a Internet:

"El descobriment, sensacional, era que si una pàgina web contenia les paraules *és* i com a mínim una de les paraules *dels*, *quan* o *què*, aleshores aquella pàgina estava sens dubte en català."

Aquest descobriment:

"Per una banda, va demostrar el que després seria una constant: que la iniciativa particular, la qualitat i l'enginy podrien fer un paper determinant en la batalla pel català a la xarxa. I d'altra banda, va demostrar a les empreses i als grans cercadors que seleccionar el català, i afegir-lo com una opció, no era ni costós ni difícil."

Vicent Partal (2004, 19 d'abril). "El català a la xarxa: història i raons d'un cas d'èxit". Pàgina web de Softcatalà.

dotCYM

dotCYM, una organització sense finalitat de lucre, fa campanya des de 2005 per a sol·licitar (i en cas d'èxit, gestionar) un domini de primer nivell per a la comunitat lingüística i cultural gal·lesa.



En definitiva, com esperem poder comprovar, la llengua catalana té una presència considerable a Internet.

Primer de tot, és important fer la distinció entre la presència d'una llengua a Internet i l'ús que els parlants de la llengua en qüestió fan d'Internet. En el primer cas, un indicador pot ser el nombre de pàgines que hi ha en aquella llengua (en termes absoluts i/o relatius). En el segon, pot ser la mesura, a partir d'enquestes, de les preferències i les conductes dels individus. Entremig, però, podem afegir indicadors d'activitat en els espais interactius de caire individual o d'esforç voluntari, com ara la *Wikipedia* o els blocs. Són espais més propis de la segona generació (2.0) d'Internet.

2.5.1. Llocs i pàgines web

Com assenyalen Castells i altres, i com ja hem avançat, hi ha quantitats molt diferents, segons el sector, de llocs web en llengua catalana. L'evolució d'aquesta presència ha estat objecte d'un estudi regular, des de l'any 2002, de **Webmàsters Independents en Català, de Cultura i d'Àmbits Cívics (WICCAC)**, una entitat que es defineix de la manera següent:

Referència bibliogràfica

M. Castells; F. Lupiàñez; F. Saigí; J. Sánchez (2007). *Projecte Internet Catalunya. Modernització Tecnològica, Canvi Organitzatiu i Servei als Usuaris en el Sistema de Salut de Catalunya. Informe de recerca* (vol. I). Barcelona.

"WICCAC és una agrupació de webmàsters que hem creat personalment, individualment o en grup, les nostres webs; que som independents, és a dir, que no estem lligats a organismes oficials ni entitats comercials ni professionals ni grups polítics; que tenim propòsits culturals o cívics, i que treballem en català. WICCAC va néixer el mes de febrer de l'any 2001.

Una de les activitats més conegudes de WICCAC és el *Baròmetre de l'ús del català a Internet*, un estudi sobre la utilització del català a les webs d'empreses i organitzacions radicades a Catalunya o a altres llocs dels Països Catalans, o que hi ofereixen els seus productes i serveis.

El *Baròmetre* es va crear el 12 de juny de 2002 i s'ha convertit en un estudi de referència. Cada mes es fa un *Resum de situació del baròmetre de l'ús del català a Internet*, amb percentatges d'utilització del català per sectors i de manera global. A partir d'una amplíssima mostra de llocs webs, creada amb criteris científics perquè sigui representativa del conjunt, WICCAC ha anat calculant aquest percentatge i presenta al seu lloc web aquests resultats mensuals des de l'agost de 2002.

Històric dels *Resums de situació del baròmetre de l'ús del català a Internet*

Mes/ Any	gen.	febr.	març	abr.	maig	juny	jul.	ag.	set.	oct.	nov.	des.
2010	58,89%	59,04%	59,13%	59,17%	59,24%	59,31%	59,31%	59,14%				

Font: http://wiccac.cat/hist_resums.html

Vídeo

Hi ha un vídeo breu, *El català, un exemple d'èxit a Internet*, de 02/10/2009, realitzat amb motiu del Simposi Internacional de Multilingüisme i Ciberespai organitzat per Linguamón, Casa de les Llengües.

Vegeu també

Hem parlat del nombre de pàgines per llengua en el capítol "La presència de les llengües a Internet". Les preferències d'ús dels individus, però, les tractarem en el capítol "Les llengües i els usuaris".

Mes/ Any	gen.	febr.	març	abr.	maig	juny	jul.	ag.	set.	oct.	nov.	des.
2009	56,41%	56,63%	56,83%	57,01%	57,09%	57,29%	57,58%	57,71%	58,12%	58,30%	58,55%	58,67%
2008	54,58%	54,73%	54,89%	54,98%	55,04%	55,13%	55,23%	55,48%	55,64%	55,82%	55,95%	56,16%
2007	52,52%	52,87%	53,09%	53,15%	53,34%	53,39%	53,53%	53,57%	53,70%	53,92%	54,19%	54,31%
2006	50,48%	50,56%	50,72%	50,78%	50,85%	50,95%	51,08%	51,09%	51,45%	51,57%	52,03%	52,23%
2005	46,51%	46,79%	47,13%	47,47%	47,70%	48,10%	48,06%	48,59%	48,97%	49,42%	50,02%	50,48%
2004	42,72%	43,50%	44,04%	44,07%	44,73%	44,87%	44,74%	45,43%	45,77%	46,44%	46,84%	46,96%
2003	40,48%	40,75%	41,43%	41,48%	40,96%	41,48%	41,84%	41,90%	41,86%	41,88%	41,67%	42,08%
2002								40,71%	38,75%	38,75%	39,22%	40,20%

Font: http://wiccac.cat/hist_resums.html

En aquest període de gairebé vuit anys, el percentatge de llocs web amb versió catalana ha pujat, doncs, del 40% al 59%. Si deixem de banda petites fluctuacions en els primers dos o tres mesos, podem dir que aquesta proporció ha anat pujant de manera pràcticament constant, però lenta.

És igualment interessant –i significatiu– estudiar com la presència del català varia segons els sectors o els grups. WICCAC també fa aquesta anàlisi. Al seu lloc web ens informa de l'estructuració seguida:

Grans Empreses i Multinacionals | Automòbil | Telecomunicacions i Telefonia | Caixes i Bancs | Portals | Cercadors i Directoris | Fires i Salons | Cinema | Teatre i Dansa | Diaris i Revistes | Viatges, Turisme, Transports i Hotels | Ràdio i Televisió | Esports | ONG | Jocs i Oci | Música | Universitats | Editorials i Llibreries | Assegurances | Alimentació i Begudes | Vins i Caves | Electrodomèstics i Fotografia | Informàtica i Internet | Immobiliàries | Moda i Roba | Neteja i Perfumeria | Varis | Laboratoris | Òptiques | Missatgeria i Transport | Església | País Valencià | Illes Balears | Empreses de treball temporal | Institucions de l'Administració de l'Estat i de la UE.

Aquest tema mereix un estudi específic en el pla docent de l'assignatura, gràcies a les dades disponibles al lloc web de WICCAC.

Altres fonts

Altres fonts estadístiques sobre Internet a Catalunya:

- Observatori de la Societat de la Informació (OBSI)
- Fundació Observatori per a la Societat de la Informació de Catalunya (FOBSIC)
- Web del Departament d'Innovació, Universitats i Empresa de la Generalitat de Catalunya

2.5.2. Espais interactius i participatius

Cal dir que a mesura que han anat entrant espais interactius a Internet, i sobretot espais que poden ser creats i gestionats per particulars sense una formació específica, la presència del català no s'ha quedat pas enrere. Al contrari, en l'anomenada Internet 2.0 o WWW 2.0, l'èxit en la comunitat catalana d'espais com Fotolog i Facebook i, en general, dels blocs ha estat considerable, fins al punt que algunes fonts, com Antoni Mir, director de Linguamón, Casa de les Llengües, estimen que el català és la vuitena llengua més usada en el blocs arreu del món.

Mercè Molist (2006) ha estudiat a fons la blocosfera catalana. Repassa els blocs pioners de casa nostra de caire tecnològic com Bulma, Cat-linux o PuntBarra, nascuts l'any 1998. Els primers blocs personals apareixen el 1999, i l'explosió es produeix entre el 2003 i el 2004.

La **Viquipèdia**, *Wikipedia* en català, va néixer el 16 de març de 2001. Segons s'explica en el mateix recurs (en la versió en espanyol), va ser la segona edició que es va crear, després de la *Wikipedia* en anglès.

Segons un apunt (*post*) de Xavi Villavilla, el català és la quinzena llengua en continguts a la *Wikipedia*. És relativament senzill investigar aquesta estadística, perquè es poden fer estudis de la presència, al llarg del temps, de les llengües a la *Wikipedia* mitjançant una eina estadística que el lloc ofereix i que conté un selector de llengua (la gent activa en cada llengua, la que fa aportacions, també hi apareix quantificada). A més, cada URL ofereix fins a dinou estadístiques diferents, relatives a cada versió lingüística. Vam posar a prova l'afirmació de Xavi Villavilla el 17 d'abril de 2010 amb la versió catalana de l'eina:

Nombre d'articles a la *Wikipedia*, per llengua. "Articles that contain at least one internal link"

1	Anglès	EN	3.200.000
2	Alemanys	DE	1.040.000
3	Francès	FR	922.278
4	Polonès	PL	681.985
5	Italià	IT	671.145
6	Holandès	NL	589.891
7	Espanyol	ES	571.449
8	Japonès	JA	553.502
9	Portuguès	PT	549.072
10	Romanès	RU	503.432
11	Suec	SV	348.367

Referència bibliogràfica

Antoni Mir (2009, 29 de setembre). *El País*.

Referència bibliogràfica

Mercè Molist (2006). "La blocosfera catalana". *Co-neixement i societat: revista d'Universitats, Recerca i Societat de la Informació* (núm. 10, pàg. 28-55).

Referència bibliogràfica

Xavi Villavilla (2009, 21 d'octubre). "El català, vuitena llengua amb més presència a la blogosfera".

12	Xinès	ZH	289.220
13	Noruec (bokmål)	NO	250.671
14	Finès	FI	232.166
15	Català	CA	229.614
16	Ucrainès	UK	192.487
17	Hongarès	HU	156.540
18	Txec	CZ	152.450
19	Turc	TR	143.174
20	Romanès	RO	140.545
21	Coreà	KO	129.821
22	Esperanto	EO	126.496
23	Danès	DA	124.774
24	Indonesi	ID	118.328
25	Vietnamita	VA	116.377
26	Àrab	AR	116.099
27	Eslovac	SK	113.323
28	Serbi	SR	112.389
29	Lituà	LT	102.057
30	Hebreu	HE	97.880
31	Búlgar	BG	92.744
32	Farsi	FA	91.344
33	Eslovè	SL	85.528
34	Croata	HR	80.985
35	Estonià	ET	71.888
36	Malai	MS	61.509
37	Anglès simple	SIMPLE	59.133
38	Thai	TH	57.854
39	Irlandès	GA	56.135
40	Noruec (nynorsk)	NN	54.585
41	Hindi	HI	53.756
42	Eusquera	EU	51.583
43	Grec	EL	48.915
44	Telugu	TE	44.413

45	Macedoni	MK	38.515
46	Llatí	LA	37.118
47	Georgià	KA	36.992
48	Bretó	BR	31.991
49	Àzeri	AZ	30.219
50	Bosnià	BS	28.861

Des de tots els punts de vista, el pes de la versió catalana de la *Wikipedia* és molt superior al pes demogràfic dels catalanoparlants. Que el francès només tingui quatre vegades més entrades, i el castellà i l'italià només tres (a tot estirar) dóna fe de com la comunitat lingüística catalana (i algunes altres) s'ha bolcat col·lectivament en aquesta empresa.

Reflexió

És interessant veure en quin mes comencen les estadístiques de la *Wikipedia* per a cada llengua. Per exemple, pel que fa a *Wikipedians who contributed 5 times or more in this month*, per a l'anglès la xifra supera els 10 per primer cop l'octubre de 2001; per al castellà, el gener de 2002; per al francès, el maig de 2002, i per a l'italià, el maig de 2003. Quin fou el primer mes, per a la versió en català?

2.6. Les llengües i els usuaris

Hem vist, a grans trets, quina és l'oferta de la llengua catalana a Internet. Ens podem demanar, però, quin n'és el consum a Internet i, en definitiva, a la resta d'eines tecnològiques. Dit d'una altra manera, més enllà de les iniciatives en xarxa, quanta gent fa servir el català i quanta gent el castellà o altres llengües?

Les dades són parcials. Per exemple, podem donar dades per al Principat de Catalunya, referides a l'any 2003, segons declaracions dels entrevistats de l'estudi Institut d'Estadística de Catalunya EULC/2003. La taula 4.4 dóna aquests resultats:

Ús del català a Internet: llengua en la pàgina inicial a Internet (població que utilitza la xarxa Internet habitualment). Catalunya. 2003. Àmbits territorials

	Català	Castellà	Anglès	Altres	Total
Àmbit metropolità	17,9%	68,5%	8,6%	5,0%	100,0%
Comarques gironines	34,0%	51,8%	6,4%	7,8%	100,0%
Camp de Tarragona	21,6%	68,6%	6,7%	3,0%	100,0%
Terres de l'Ebre	41,9%	55,1%	3,1%	0,0%	100,0%
Àmbit de Ponent	34,3%	58,2%	2,7%	4,8%	100,0%
Comarques centrals	33,2%	57,4%	6,3%	3,1%	100,0%
Alt Pirineu i Aran	31,7%	59,0%	6,4%	2,9%	100,0%

	Català	Castellà	Anglès	Altres	Total
Catalunya	21,4%	65,9%	7,8%	4,9%	100,0%

Pel que fa a la llengua de la pàgina inicial a Internet, el resultat indica una superioritat molt més gran del castellà que la que esperaríem del seu pes demogràfic dins la població estudiada. Fins i tot a les vegueries amb menys incidència de gent vinguda d'altres terres la diferència és sempre a favor del castellà.

Però ens podem demanar, per altra banda, quins usos es fan dels recursos. La gent fa servir el que hi ha a Internet? La pregunta és massa àmplia per a poder donar una resposta global. Per tant, i com a estudi de cas, ens continuarem centrant en la llengua catalana.

En el marc d'un estudi més ampli de la Generalitat de Catalunya sobre els usos lingüístics a casa nostra, i centrant-se en l'àmbit dels mitjans de comunicació i Internet, Bretxa assenyala la gran diferència que hi ha entre els mitjans de comunicació tradicionals (premsa, ràdio i televisió) i els nous relacionats amb les tecnologies de la informació i la comunicació, representats per Internet. Mentre que en el primer cas l'individu actua gairebé únicament com a element receptor, en el segon, i més que mai en la història dels mitjans de comunicació, ho fa també com a emissor, gràcies a la interactivitat del mitjà (augmentat, tot s'ha de dir, amb la Internet 2.0). En el cas d'Internet, Bretxa analitza els hàbits lingüístics de les persones com a usuàries i consumidores (ús de la xarxa, les pàgines web inicials i les més visitades) i també com a productores (usos privats: notes personals, correu electrònic i xats).

De l'enquesta oficial de la Generalitat de Catalunya feta el 2003 que he esmentat, en retindrem una taula, del bloc que fa referència als *Usos lingüístics en les tecnologies de la informació i la comunicació*:

Ús del català a Internet: llengua en els missatges de correu electrònic (població que escriu missatges de correu electrònic). Catalunya. 2003

Milers de persones / %						
Només català	Més català que castellà	Ambdues	Més castellà que català	Només castellà	Altres situacions	Total
286,0	270,6	254,3	191,0	369,4	278,3	1.649,5
17,3%	16,4%	15,4%	11,6%	22,4%	16,9%	100%

Cal afegir, però, que aquesta enquesta de l'any 2003 contenia un biaix que es va corregir en l'estudi següent. Es tracta d'un estudi, però, que no conté preguntes sobre els usos lingüístics en les TIC.

L'estudi de Vanessa Bretxa analitza cada mitjà de manera global, però també en funció d'unes variables clau, com l'àmbit geogràfic, l'origen familiar, l'edat, el nivell d'instrucció, la llengua inicial, la llengua d'identificació o el percen-

Referència bibliogràfica

Vanessa Bretxa i Riera (2005). "Els mitjans de comunicació i Internet". A: Joaquim Torres (coordinador). *Estadística sobre els usos lingüístics a Catalunya 2003. Llengua i societat a Catalunya en els inicis del segle XXI* (pàg. 145-170). Barcelona: Generalitat de Catalunya, Secretaria de Política Lingüística ("Publicacions de l'Institut de Sociolingüística Catalana", 8).

Referència bibliogràfica

Generalitat de Catalunya (2004). *Estadística d'usos lingüístics a Catalunya 2003*. Feta conjuntament per la Secretaria de Política Lingüística del Departament de la Presidència i l'Idescat.

Referència bibliogràfica

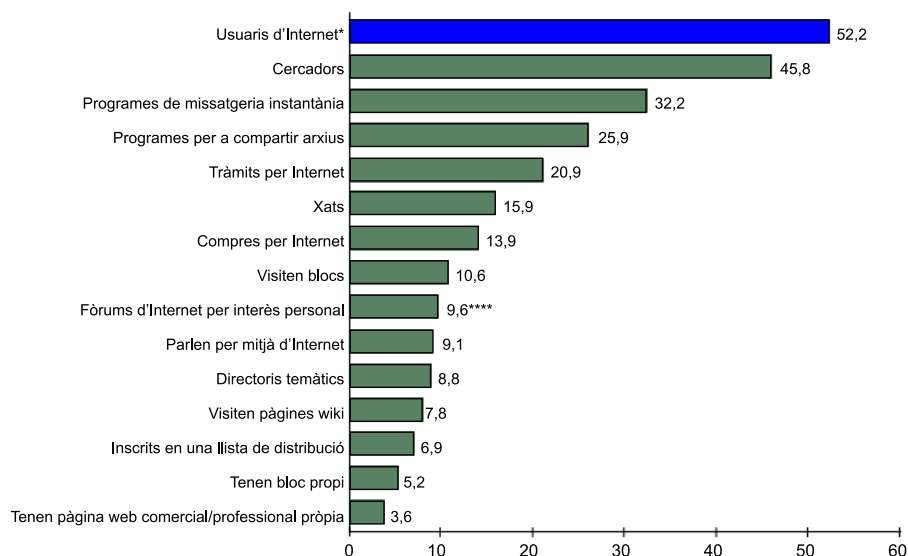
Generalitat de Catalunya (2009). *Enquesta d'usos lingüístics de la població 2008*. Enquesta feta conjuntament per la Secretaria de Política Lingüística del Departament de la Vicepresidència i l'Institut d'Estadística de Catalunya.

tatge d'ús de català. La llengua d'identificació destaca com a element decisiu en la tria lingüística, tant pel que fa al consum de mitjans de comunicació tradicionals com pel que fa al consum i l'exercici de pràctiques comunicatives a Internet. Els altres elements fonamentals són l'edat i el nivell d'instrucció.

Segons Torrijos, la meitat de la població de 15 anys o més utilitza Internet almenys un cop per setmana. Els serveis de la xarxa que es fan servir més són els cercadors (gairebé tots els usuaris han entrat en un cercador) i, a certa distància, els programes de missatgeria instantània, els programes per a compartir arxius i els tràmits.

Referència bibliogràfica

Anna Torrijos López (2008). "Perfils de consumidors i usos lingüístics en les TIC". *Llengua i Ús* (núm. 41, pàg. 70-88).



Font: Anna Torrijos López (2008). "Perfils de consumidors i usos lingüístics en les TIC". *Llengua i Ús* (núm. 41, pàg. 83)

La major part d'aquestes activitats es vertebren o es porten a terme per mitjà del llenguatge humà. És a dir, en una situació bilingüe o multilingüe serà de gran interès veure si el grau d'ús de les llengües en cada activitat, per part dels usuaris, s'acosta a la proporció d'aquestes llengües en diferents sectors de la societat.

Mas i Fossas afirma el següent:

"La tecnologia va arrelant cada cop més a la nostra vida quotidiana i, en aquest context, tant les autopistes de la informació i, especialment, Internet, com els programes informàtics han de poder funcionar en català en el món global intercomunicat d'ara. Els usuaris de la tecnologia han de poder utilitzar el català com a llengua vehicular i més encara quan això implica potencialment la societat sencera, atès que actualment el mercat informàtic i telemàtic no és un mercat tancat o limitat a un grup d'experts, sinó que s'ha fet extensiu a la ciutadania."

Gemma Mas i Fossas (2003). "La normalització de la llengua catalana i la tecnologia: recursos terminològics". *Revista de Llengua i Dret* (núm. 40, pàg. 104).

Els articles de Torrijos i de Ferran i altres recullen els resultats principals de l'estudi *TIC-Llengua 2007, ús social de la llengua a Internet i en les tecnologies de la informació i la comunicació*, encarregat per la Secretaria de Telecomunicacions i Societat de la Informació de la Generalitat de Catalunya i dut a terme per la Fundació Observatori per a la Societat de la Informació de Catalunya (FOBSIC) i l'Institut Apolda. L'estudi anava més enllà d'Internet. Tenia dos objectius:

- Identificar les llengües que hi ha en els diversos dispositius tecnològics que són a l'abast de la població catalana de 15 anys o més, i en els serveis accessibles a Internet que utilitza aquesta població.
- Quantificar i localitzar en el territori els diferents usos lingüístics i relacionar aquesta presència amb els aspectes següents:
 - Llengua pròpia de les persones que utilitzen els dispositius.
 - Llengua en què aquestes persones disposen del dispositiu.
 - Llengua en què aquestes persones voldrien disposar del dispositiu.
 - Similituds i diferències en relació amb els mitjans no digitals.

La mostra total de l'estudi va ser de 4.509 entrevistes, la suma d'una mostra aleatòria de 2.800 entrevistes i una mostra aleatòria complementària de 1.709 entrevistes més fetes només a persones usuàries d'Internet (definides com les que es connecten a la xarxa un cop cada dues setmanes pel cap baix). En conjunt l'estudi inclou 3.055 internautes.

A continuació, reproduïm els elements estudiats en el treball, per tal de veure el que en aquell moment semblaven les eines principals sobre les quals els usuaris podrien (o no) fer una opció lingüística.

Referència bibliogràfica

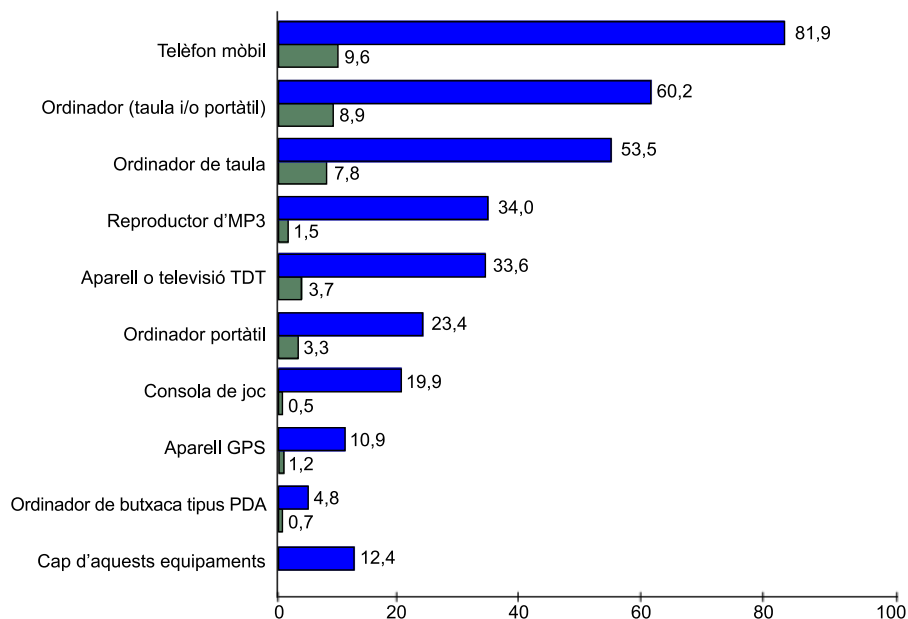
Jaume Ferran; Pilar Moreno; Marta Torres (2007). "TIC-Llengua: els usos lingüístics en les tecnologies de la informació i la comunicació a Internet". *Llengua i Ús* (núm. 39, pàg. 56-65).

Els productes i els serveis investigats

- el telèfon mòbil
- els reproductors d'MP3
- les consoles de joc
- l'aparell de TDT
- l'aparell GPS
- els traductors automàtics
- la pàgina inicial del navegador
- les pàgines web més visitades
- els cercadors
- els directoris i les pàgines wiki
- l'ordinador de sobretaula
- l'ordinador portàtil
- l'ordinador de butxaca tipus PDA o similar
- els sistemes operatius
- el bloc personal
- el web comercial propi de la persona entrevistada
- la participació en fòrums
- la participació en llistes de distribució
- la participació en xats
- les compres i els tràmits per Internet
- el programari lliure
- el programari de propietat
- el programa de processador de textos
- el corrector ortogràfic
- la comunicació mitjançant missatgeria instantània
- la parla mitjançant l'ordinador (tipus Skype)
- els programes per a compartir arxius
- els continguts compartits

Jaume Ferran, Pilar Moreno, Marta Torres (2007). "TIC-Llengua: els usos lingüístics en les tecnologies de la informació i la comunicació a Internet". *Llengua i Ús* (núm. 39, pàg. 59).

Ferran i altres indiquen el percentatge de persones que disposen d'equipaments TIC per al seu ús personal i el percentatge de persones que declaren que tenen el menú del dispositiu en català. Els percentatges en verd corresponen a l'ús declarat de català en el menú i, en el cas de l'ordinador, en el sistema operatiu:



Font: Jaume Ferran, Pilar Moreno, Marta Torres (2007). "TIC-Llengua: els usos lingüístics en les tecnologies de la informació i la comunicació a Internet". *Llengua i Ús* (núm. 39, pàg. 59)

L'estudi mostra, pel que fa a les persones que disposen de telèfon mòbil (81,9% del total), la desproporció entre la llengua en què se'n disposa i la llengua en què es preferiria disposar-ne:

	llengua en què se'n disposa	llengua en què es preferiria disposar-ne
català	11,8%	30,3%
castellà	82,7%	46,8%
anglès	2,5%	1,3%
altres	1,2%	0,8%
l'hi és indiferent		20,1%
NS/NC	1,9%	0,8%

Font: Jaume Ferran, Pilar Moreno, Marta Torres (2007). "TIC-Llengua: els usos lingüístics en les tecnologies de la informació i la comunicació a Internet". *Llengua i Ús*, número 39 (pàg. 60)

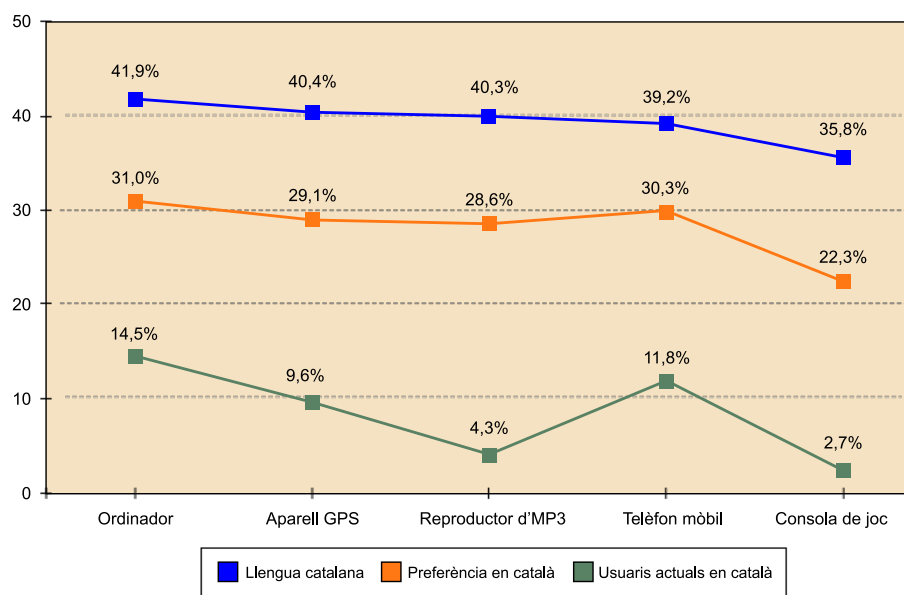
Citació

"Quant a l'exemple que hem analitzat amb més profunditat, el menú del telèfon mòbil, les dades mostren que tot i haver-hi el 38,8% de la població que resideix a Catalunya que té com a llengua pròpia el català i un percentatge elevat de grups amb altres llengües pròpies que declaren que voldrien tenir el menú dels seus dispositius tecnològics en català, només una mínima part de la població ha pogut accedir als dispositius en aquesta llengua. [...]

Així doncs, les preferències de les persones enquestades donen compte d'una demanda real, insatisfeta, de productes tecnològics en català que les empreses del sector han de tenir en compte pel volum de vendes que poden representar."

Jaume Ferran; Pilar Moreno; Marta Torres (2007). "TIC-Llengua: els usos lingüístics en les tecnologies de la informació i la comunicació a Internet". *Llengua i Ús* (núm. 39, pàg. 64-65).

Un segon treball derivat del mateix estudi revela la diferència entre (a) la proporció de catalanoparlants a la mostra (sobre els usuaris de cada tipus de dispositiu); (b) la preferència pel català entre aquestes persones; (c) els que usen (o poden usar) el català en cada tipus de dispositiu:



Anna Torrijos López (2008). "Perfils de consumidors i usos lingüístics en les TIC". *Llengua i Ús* (núm. 41)

La distància és molt notòria, però l'ús de les llengües s'acosta més quan l'usuari és lliure a l'hora de redactar. Així, i segons aquest mateix treball, en els procesadors de text, els correctors ortogràfics i els missatges curts de mòbil (SMS, *short message service*), l'ús del català se situa prop d'un 30% i l'ús d'ambdues llengües prop d'un 23-37% addicional.

Resum

En aquest mòdul hem mirat d'esbrinar quina és la situació del multilingüisme a Internet, centrant-nos sovint en el cas concret del català.

Hem constatat que no per a tothom Internet és percebut com a oportunitat. Hi ha comunitats lingüístiques, sobretot en parts del món poc desenvolupades, en què s'han de complir una sèrie de condicions per tal de poder-se'n beneficiar.

Hem presentat dades sobre l'evolució de la presència de les llengües a Internet, des de l'inici de la WWW fins ara, i hem constatat com s'hi ha anat produint la pèrdua lògica de domini absolut de què l'anglès gaudia en un primer moment.

Hem parlat de la problemàtica que comporta l'existència de grafies molts diferents de la llatina, i de com han anat entrant a Internet. En concret, hem esbossat els esforços per a garantir que els noms dels dominis es puguin escriure –i registrar, és clar– en grafies diferents de la llatina.

Hem parlat a continuació dels diversos tipus de dominis de primer nivell i del cas –únic– del domini per a la comunitat lingüística i cultural catalana, el *.cat*, que altres comunitats lingüístiques aspiren a obtenir.

Hem quantificat, segons diverses fonts, la presència del català a Internet d'acord amb recomptes de paraules i de pàgines web en diferents llengües.

Hem parlat també, agafant novament el cas català com a estudi de cas, de les llengües i els usuaris, en diferents eines informàtiques i tecnològiques, per tal de veure si hi ha una descompensació entre l'oferta i la demanda.

