



Les tecnologies de la llengua en la normalització lingüística del gallec

1. Introducció: una aproximació a la situació sociolingüística del gallec

AUTOR

Xavier Gómez Guinovart
Universitat de Vigo

En aquesta ponència parlaré de tecnologies de la llengua i de la llengua en les tecnologies, centrant-me en la llengua gallega i en les conseqüències que aquestes dues vessants tenen en la seva normalització lingüística. L'esquema de la meua intervenció serà el següent:

1. Introducció: una aproximació a la situació sociolingüística del gallec
2. La llengua gallega en les tecnologies de la informació i comunicació
3. Les tecnologies de la llengua gallega
4. Conclusions: llengua, tecnologia i normalització lingüística.

Primer de tot, per tal de contextualitzar la meua exposició, començaré amb algunes dades sociolingüístiques que permetin fer-se una idea sobre la situació actual del gallec.

El *Mapa Sociolingüístic de Galicia* (MSG), elaborat sota els auspicis de la Real Academia Galega pels membres del seu Seminario de Sociolingüística, és l'estudi demolingüístic quantitatiu més ample sobre la situació i l'evolució del gallec a Galícia. L'estudi s'ha realitzat a partir d'enquestes realitzades en dos moments del temps amb dotze anys de diferència: el 1992 i el 2004.

El MSG s'edita en tres volums: un amb les dades de les enquestes sobre llengua inicial i competència lingüística, un altre amb les dades sobre els usos lingüístics, i un tercer amb les dades sobre actituds lingüístiques. D'aquest últim, només estan publicades les dades del 1992, mentre que dels dos primers tenim les dades del 1992 i del 2004.

Les dades del primer volum ens mostren com, entre 1992 i 2004, el gallec va passar de ser la llengua inicial d'una mica més de la meitat de la població a ser la llengua inicial de només una cinquena part de la població. Aquestes dades estan representades visualment en la següent figura, extreta del MSG-04:

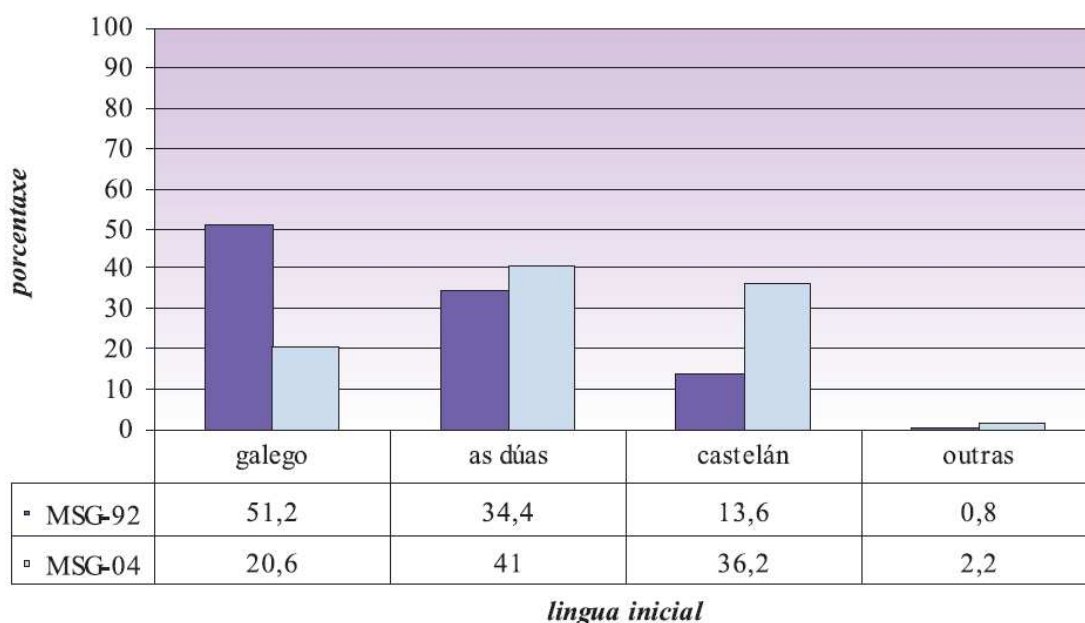


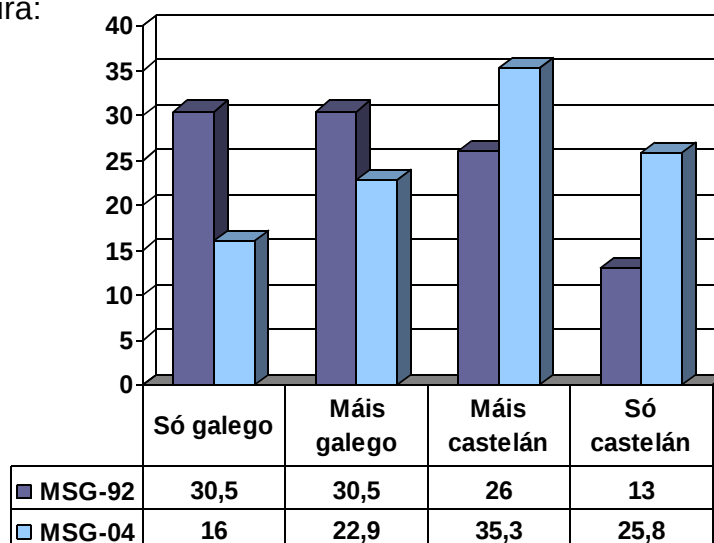
Figura 2.2. *Comparativa da lingua inicial no tramo 16-54 anos no MSG-92 e MSG-04*

En aquesta figura s'haurien de destacar els següents aspectes:

- que el percentatge de parlants inicials de gallec va patir una caiguda de 30,6 punts (del 51,2% en el MSG-92 al 20,6% en el MSG-04) al tram d'edat considerat
- que la proporció de parlants inicials de castellà va créixer 22,6 punts (del 13,6% en el MSG-92 al 36,2% en el MSG-04)
- que els bilingües inicials també van augmentar (6,6 punts) i són en el MSG-04 el grup més nombrós (el 41%)
- que la proporció de persones que van aprendre a parlar en d'altres llengües gairebé es va triplicar.

Comparant les dades del MSG-94 amb les del MSG-02 es pot apreciar la dimensió que està assolint el procés de substitució lingüística durant la última dècada. La distribució minvant de la llengua inicial és un indicador força clar del procés d'extinció d'una llengua. Quan es talla la transmissió intergeneracional de la llengua, ja es pot ben dir que el procés de substitució lingüística s'ha posat en marxa.

La rellevància d'aquesta caiguda de més de 30 punts en 12 anys del gallec com a llengua primera és encara més patent si tenim també en conta l'evolució dels indicadors sobre llengua habitual recollits al segon volum del MSG, com es pot veure a partir de la següent figura:



D'acord amb aquests dades, en aquest període es van produir alguns canvis substancials en les tendències d'ús lingüístic:

- es va duplicar el nombre de persones que no parlen mai gallec (del 13% al 25,8%)
- es va dividir a la meitat el nombre de monolingües en gallec (del 30,5% al 16%)
- es va reduir en 22 punts el nombre de persones que tenen el gallec com a llengua habitual
- (del 61% al 39%)

Amb aquests indicadors no és difícil concloure que el gallec a Galícia ha passat de ser la llengua minoritzada a ser la llengua minoritària i, encara més rotundament, que, si les tendències sociolingüístiques no s'alteren, el gallec a Galícia es troba en un clar perill d'extinció.

En aquesta situació sociolingüística de greu risc d'extinció per a la llengua pròpia, caldrien actuacions proteccionistes de política lingüística que permetessin aturar aquest procés en marxa i revifar la llengua. Tanmateix, la realitat política de Galícia és ben bé un altre, i totes les mesures de política lingüística del nou govern semblen apuntar en un sentit contrari: reforma de la Lei da Función Pública per la qual es suprimeix l'examen obligatori en gallec de les oposicions (juny de 2009); desmantellament de la Rede de Galescolas, xarxa pública d'escoles infantils que tenia com a llengua vehicular el gallec (febrer del 2010); elaboració de l'anomenat "Decreto para o plurilingüismo no ensino non universitario de Galicia", en fase de avantprojecte i d'imminent aprovació pel partit majoritari (presentat el 13 de març de 2010)...

Encara que el nombre de pàgines en la web constitueix un índex significatiu de la presència d'una llengua en Internet (i per tant de la seva vitalitat lingüística), la

rellevància d'aquest índex és molt més gran quan es pondera amb la grandària de la població parlant de cada llengua. L'índex ponderat de pàgines web per parlant constituïria un índex de la quantitat d'informació i continguts web als quals pot accedir una persona en la seva pròpia llengua. D'aquesta manera, el valor en el 2002 de l'índex de pàgines en la web per a la llengua gallega (0,014%), semblant al que obtenia l'euskera (0,011%), ha de ser interpretat tenint en consideració, entre altres variables, la diferència a favor del gallec entre el nombre de parlants de l'euskera i el de parlants del gallec. Observant aquesta variable, la llengua gallega no experimentaria variacions de classificació en el rànquing de les llengües amb respecte a l'anterior, mantenint-se en el trenta setè lloc, amb una proporció de 0,025 pàgines web per parlant, mentre que l'euskera ascendiria en la classificació del trenta vuitè lloc al dècim novè, amb un índex estimat de 0,138 pàgines web per parlant.

No obstant això, el nombre de pàgines web escrites en gallec va augmentar d'una manera sostinguda des de l'any 2002 en el qual es va dur a terme l'esmentat estudi. En l'actualitat, el nombre de pàgines web en gallec accessibles pel cercador "de referència" podria aproximar-se als 11 milions de pàgines, mentre que el nombre total de pàgines indexades per aquest mateix cercador podria ser de 13,5 mil milions de pàgines (dades extretes de WorldWideWebSize). Això és, en vuit anys, la presència del gallec en Internet hauria augmentat més de cinc vegades, del 0,014% al 0,08%. Comparat amb el català i amb l'euskera, el gallec seria la llengua de les tres amb major creixement percentual en els últims vuit anys. La presència actual del català en el cercador podria xifrar-se en uns 19 milions (del 0,094% al 0,140%) i la de l'euskera en uns 6 milions (del 0,011% al 0,044%).

	2002 <i>AllTheWeb</i>	2010 <i>Google</i>
<i>Pàgines en gallec</i>	98.998 0,014%	11.000.000 0,08%
<i>Pàgines en català</i>	681.768 0,094%	19.000.000 0,140%
<i>Pàgines en basc</i>	80.271 0,011%	6.000.000 0,044%
<i>Total pàgines</i>	728.948.154	13.500.000.000

Una raó que explicaria aquest creixement radica en el desenvolupament tardà de la teranyina gallega. D'acord amb l'última edició (de novembre de 2009) de l'Estudio General de Medios, la població gallega usuària d'Internet va passar del 23,6% de la població gallega en 2003 al 42% en 2009, mentre que a Catalunya en el mateix període es va passar del 30,8% al 57,1% i a Euskadi del 29,1% al 55,4%. La penetració d'Internet en els àmbits lingüístics del català i l'euskera, doncs, va uns anys pel davant amb respecte a la penetració de la xarxa en l'àmbit gallec, pel que té sentit que la producció de material textual per a la web en gallec s'iniciés amb un cert retard i amb una certa urgència, en un moment en què d'altres llengües com la catalana i la basca ja tenien el camí inicial fet.

Una altra raó del creixement relatiu del gallec en Internet podria tenir a veure justament amb les característiques sociològiques de la població gallega usuària activa en la teranyina. Dintre d'aquest sector, trobarem de ben segur els grups socials que més treballen en la xarxa per a la llengua: persones compromeses amb l'idioma del país i activistes lingüístics del ciberespai, galipedistas, redeiras i redeiros gallecs de fòrums i xarxes socials, blogueiros i blogueiras del blogomillo, falangulleiras i falangulleiros dels podcasts, periodistes digitals, la comunitat del programari lliure, creadores i creadors de continguts textuais i audiovisuals per a la web de tota mena de gèneres digitals i en la llengua pròpia, fent brollar amb força el gallec a l'Internet i augurant mil primaveres més per a aquest idioma en un camp, el del món virtual, on la presència de les llengües serà determinant per a la seva supervivència.

Les tecnologies de la llengua s'orienten al desenvolupament d'aplicacions informàtiques d'ús general que facilitin la utilització de la llengua, la seva traducció, el seu estudi o el seu aprenentatge. Tanmateix, el desenvolupament d'aplicacions de tecnologies de la llengua suposa una sòlida base prèvia de recursos lingüístics informatitzats i d'eines de processament del llenguatge. Per eines de TL entenem els sistemes informàtics orientats al desenvolupament d'aplicacions de TL, mentre que els recursos de TL constitueixen les dades lingüístiques sobre les quals es construeixen les aplicacions. Els recursos i la investigació lingüística constitueixen els pilars de les eines i de les aplicacions i, alhora, les eines i les aplicacions fan un

gran servei quan es tracta d'ampliar o millorar els recursos lingüístics adquirits. Eines, recursos i aplicacions són les peces necessàries de l'engranatge de les tecnologies de la llengua, i la seva interdependència és tan gran que tot sovint són concebudes com un conjunt indestruïble.

Algunes de les aplicacions més populars de les tecnologies lingüístiques són, la correcció ortogràfica, els diccionaris electrònics, la traducció automàtica i les tecnologies de la parla (particularment, en el telèfon, en el control vocal del mòbil, centraletes automatitzades i SMS; i en eines d'ajut a la conducció d'automòbil). El gallec s'hi troba representat en tots quatre àmbits, amb diferents graus de presència.

El gallec disposa de correctors automàtics de nivell ortogràfic i sintàctic, la qual cosa no és gaire freqüent en les llengües del seu entorn. També disposa d'un conjunt escaient de diccionaris electrònic, si bé en aquest terreny es troben a faltar, d'una banda, diccionaris plurilingües electrònics en un nombre més gran de combinacions lingüístiques i, d'altra banda, diccionaris electrònics de sinònims. En el camp de la traducció automàtica, es poden trobar per al gallec traductors bidireccionals amb el castellà, portuguès, català, anglès i francès, a més a més del traductor de Google que, des del febrer de 2009, posa al gallec en relació bidireccional amb uns altres 50 idiomes (entre ells, la major part dels europeus, el xinès, el rus, l'àrab, el japonès, el coreà i el swahili). Finalment, en l'àmbit de les tecnologies de la parla, es disposa de sistemes de conversió de veu a text i de sistemes de conversió text-veu de gran qualitat, com el de l'empresa torinesa Loquendo, amb veu de gran naturalitat i sorprenents efectes expressius.

La direcció actual de la investigació i desenvolupament en TL assenyala dos objectius prioritaris: d'una banda, la superació de les barreres lingüístiques que, especialment en l'àmbit d'Internet, impedeixen un accés igualitari als continguts i, de manera molt més rellevant per a la indústria, dificulten la implantació efectiva d'un comerç electrònic realment global; de l'altra banda, la superació de les "barreres arquitectòniques" d'accés a les noves tecnologies mitjançant la implementació d'interfícies d'interacció persona-ordinador en llenguatge natural oral. Les tecnologies clau que ens hauran de permetre l'ús de la pròpia llengua per accedir la

totes les possibilitats d'informació, comunicació i consum posades a disposició dels habitants del primer món per les noves tecnologies són la identificació automàtica de la llengua, la traducció automàtica, el reconeixement i síntesi de la parla, la recuperació interlingüística de la informació, els sistemes de diàleg i la web semàntica.

En aquest sentit, qualsevol comunitat lingüística que vulgui accedir en igualtat de condicions a aquesta infraestructura social i econòmica en la qual es configura la nostra societat de la informació, haurà de poder comptar amb aquestes tecnologies i haurà d'estar representada dintre d'aquesta infraestructura. Per la seva gran incidència social, la presència d'una llengua en l'àmbit de les TL és determinant per aconseguir o per conservar l'estat de llengua normalitzada.

4- Conclusions: llengua, tecnologia i normalització lingüística

La presència del gallec en les TL, en les TIC i en Internet pot ser important per a la seva normalització lingüística i pot ser un factor clau de resistència a l'avançat procés de substitució al qual apunten les tendències recollides al *Mapa Sociolingüístic*. Primer, perquè Internet és el mitjà de comunicació actual de major importància estratègica, i la presència del gallec en els mitjans de comunicació és imprescindible per a la seva normalització. Segon, perquè l'increment de la percepció social del prestigi i utilitat del gallec és fonamental per a la recuperació de la seva transmissió lingüística familiar (ara per ara, un dels principals esculls per a la supervivència de la llengua), i la presència del gallec en Internet garanteix l'increment necessari d'aquesta percepció de prestigi i utilitat. I, tercer, perquè el gallec precisa d'una identitat i una presència internacional reconeguda i respectada, i la presència del gallec en Internet i en les TIC assegura la difusió idònia per al seu coneixement i reconeixement a nivell global.

En síntesi, la presència del gallec en les TL, en les TIC i en Internet és important per a la seva normalització perquè pot contribuir a augmentar el seu ús, el seu prestigi i la seva utilitat, a més de constituir un factor simbòlic important en la percepció pròpia i aliena de la llengua.