

## El dret a viure plenament en la pròpia llengua, també «on line»

*Els parlants d'una llengua minoritària han de tenir el mateix dret a l'accés a les tecnologies de la informació que els parlants de les llengües més majoritàries | Cal potenciar la sobirania tecnològica*



Les noves plataformes de vídeo en 'streaming' presenten dificultats a l'hora d'oferir els seus continguts en català. | Adrià Costa.

*Segon article del dossier «El català al món digital»*



Quins són els punts forts i febles del català en l'àmbit digital?

El català sempre ha tingut una forta presència digital a internet. Només cal recordar que va ser la tercera llengua que va tenir la Viquipèdia per darrere de l'anglès i de l'alemany, i la segona que en va crear continguts, només per darrere de l'anglès. Actualment, és la vintena llengua en nombre d'articles, mentre que en nombre de parlants a nivell mundial es troba a la posició 128 segons el rànquing d'Ethnologue del 2021.

La consciència digital dels usuaris de la llengua catalana, doncs, és molt alta, i és un fet que s'ha reflectit en iniciatives molt importants com la de Softcatalà o Col·lectivaT, que promouen l'ús de programari en català, i també de serveis basats en tecnologies de la llengua, com ara reconeixadors i sintetitzadors de la parla, o traductors automàtics. **Aquesta consciència digital és un dels nostres punts forts, que si bé és molt alta en l'àmbit d'Internet, falta enfortir-lo encara més en el tema de productes, serveis i aplicacions.** En aquest sentit, no hem de desestimar el fet que Catalunya és un dels països més avançats a Europa en l'àmbit digital, i al capdavant de la llista de llocs més atractius del sud d'Europa per a captar inversió estrangera en tecnologia. Tot això hauria de jugar a favor de la llengua, també.

Els punts febles els compartim amb totes les llengües minoritàries i minoritzades. **El cost de generar recursos lingüístics que puguin donar forma a tots aquests serveis i aplicacions és molt alt**, si bé en algunes aplicacions només caldria un mínim de localització per tenir-los disponibles. En el cas del català, aquest punt feble encara s'afebleix més pel fet de compartir l'espai amb una llengua molt dominant i molt majoritària. En la gran majoria de productes digitals, les opcions lingüístiques venen determinades per grans corporacions que encara tenen la mentalitat d'*un estat, una llengua*, sense contemplar el multilingüisme existent dels països en els quals els comercialitzen, ja sigui per motius econòmics o per simplement per desinterès.

Quines accions s'haurien de promoure perquè fos una llengua disponible a tots els serveis digitals?

En primer lloc ho hem d'exigir com a usuaris. Viure plenament en la pròpia llengua és fer-ho també a les nostres pantalles, i és un dret fonamental que hem de reclamar. La consciència lingüística ha de passar per davant de la inèrcia d'utilitzar productes en castellà o en anglès 'perquè ja s'entén'. En segon lloc, les institucions han de contribuir a generar aquesta consciència lingüística entre els parlants, donant exemple, posant de manifest la riquesa que representa cada llengua, i tenint en compte que, avui en dia, **una llengua que perd el tren de la digitalització és condemnada a l'extinció.**

Les empreses acaben responant tard o d'hora a les demandes i exigències dels usuaris, i generaran productes en català si hi ha mercat. I en aquest sentit, hi ha dues accions principals que

# Pensem.

s'haurien de promoure. En primer lloc, apostar sempre que sigui possible per petites i mitjanes empreses locals, més properes culturalment i geogràficament, i per tant amb més consciència lingüística. Dependre dels gegants tecnològics i de la seva bondat perquè incloguin la nostra llengua en els seus productes no jugarà mai a favor nostre. En segon lloc, evitar sancions i, enlloc de sancionar, facilitar que aquestes empreses puguin produir els seus productes en català, generar recursos que es puguin compartir lliurement i que puguin ser utilitzats per les empreses, però també a nivell individual i institucional. En aquest sentit, hem de **potenciar molt més la sobirania tecnològica**, que ens permetrà tenir la paella pel mànec en l'àmbit lingüístic.



Quins actors o recursos caldria activar perquè fos possible?

La sobirania tecnològica rau en la gestió i **creació de recursos propis lliures i oberts**, amb independència dels programaris i d'altres serveis tecnològics de propietat. Aquesta sobirania té molts avantatges per a les llengües minoritzades com la nostra, perquè com deia més amunt, permet tenir la paella pel mànec en l'àmbit lingüístic sense haver d'esperar ni de suplicar a les grans multinacionals que incloguin la nostra llengua en els seus productes.

A nivell del carrer, hi ha algunes iniciatives molt interessants que cal potenciar i reconèixer. Softcatalà, per exemple, és una associació sense ànims de lucre que, des del 1997, fomenta la presència i l'ús del català a la informàtica, Internet i les noves tecnologies. Ofereix un ventall molt ampli de recursos, des de la traducció de programari i catalanitzadors de sistemes operatius i aplicacions, fins a diverses eines lingüístiques com correctors, diccionaris multilingües, diccionaris de sinònims, reconeixadors automàtics de veu o traductors automàtics. Recentment, ha incorporat un traductor automàtic entre el català i anglès basat en xarxes neuronals, que és la tecnologia més avançada en aquest tipus de sistemes.

La cooperativa Col·lectivaT, nascuda l'any 2017, també ofereix recursos computacionals en reconeixement i síntesi de la parla i traducció automàtica. Com a cooperativa, i amb el suport del Departament de Cultura de la Generalitat de Catalunya, també ofereix tallers sobre tecnologies lingüístiques per a desenvolupadors, emprenedors i usuaris. Fa cosa d'un any, Col·lectivaT va posar en marxa el *Catotron*, el primer sistema de síntesi de la parla lliure i obert, basat en xarxes neuronals, que com en el cas de la traducció automàtica, és la tecnologia puntera en aquest tipus d'aplicacions.

A nivell individual, també tenim l'opció de col·laborar en iniciatives com el projecte Common Voice, una iniciativa de Mozilla que recopila dades de veu en format lliure i obert per a desenvolupar sistemes de reconeixement de la parla, també de codi obert. El projecte Common Voice té com a objectiu **reduir les desigualtats digitals entre llengües en l'àmbit de les tecnologies de la parla**. Com que la majoria de dades que utilitzen aquests sistemes són de propietat i en moltes llengües són fins i tot inexistents, qualsevol parlant d'una llengua pot donar la seva veu, i també validar la precisió de frases enregistrades per altres usuaris. També es pot sol·licitar la inclusió d'altres llengües al projecte, una opció molt interessant per a les llengües més minoritàries.

A nivell institucional, el projecte AINA, que es va llançar a finals de l'any 2020 des del Departament de Polítiques Digitals de la Generalitat de Catalunya i amb la col·laboració del Centre de Supercomputació de Barcelona, té com a objectiu la generació de recursos per a entrenar algorismes d'intel·ligència artificial i poder desenvolupar totes aquestes eines lingüístiques que actualment es basen en xarxes neuronals. Tot i que de moment només s'ha alliberat un corpus de text, en un futur està prevista la creació de corpus de veu, de manera que es podran potenciar els serveis que utilitzen tecnologies de la parla com assistents de veu, reconeixadors i sintetitzadors.

L'explosió de les tècniques basades en xarxes neuronals ha fet que tots aquests sistemes presentessin un salt qualitatiu molt alt, de manera que el seu ús s'acosta cada dia més a ser una realitat en la nostra vida diària. Com a contrapartida, aquestes tècniques requereixen grans quantitats de dades, amb el **risc de deixar enrere les llengües minoritzades o amb menys parlants**. Però és a les nostres mans aconseguir que els recursos existents siguin a l'abast de tothom, i també desenvolupar cada vegada amb més insistència i com ja es comença a fer ara, tècniques punteres basades en xarxes neuronals específiques per a llengües amb pocs recursos.

[noticiadiariambautor]93/222[/noticiadiariambautor]

Com encaixa la promoció del català en el context d'un mercat global dominat per llengües majoritàries?

La promoció del català en el context d'un mercat global dominat per llengües majoritàries té les mateixes dificultats que qualsevol llengua que no tingui un estat propi, i encara més si l'estat en el qual pertany no li va a favor. **El mercat global es regeix bàsicament pel concepte d'un estat - una llengua**, i no sempre la pressió dels usuaris és prou forta per a canviar-ho. Però si al món hi ha uns 200 estats i unes 6.000 llengües, vol dir que n'hi haurà 5.800 que estaran condemnades a la bretxa digital a curt termini, i a desaparèixer a llarg termini? És un disbarat.

**Poder viure plenament en la pròpia llengua és un dret i una qüestió de qualitat de vida** a nivell individual i de comunitat lingüística. Però no només és això. La diversitat lingüística és un bé comú i una riquesa que cal preservar per a la humanitat. Cada vegada que desapareix una llengua ens empobrim tots. Per tant, el relat que s'ha de guanyar és el de la preservació de la diversitat lingüística, al mateix nivell i amb la mateixa contundència que la preservació de les espècies i d'altres moviments ecologistes, per exemple, i el debat ha d'anar per davant de les exigències del mercat global. Les tecnologies han de servir precisament fer fomentar i fer valer aquesta riquesa, i no per a destruir-la.

En aquest sentit, els Objectius de Desenvolupament Sostenible de l'Agenda 2030 de les Nacions Unides són molt clars: una de les fites de l'Objectiu 11, sobre 'ciutats i comunitats sostenibles', és justament protegir i salvaguardar el patrimoni cultural i natural del món, mentre que una de les fites de l'Objectiu 9, sobre la "construcció d'infraestructures resilients, la promoció de la industrialització inclusiva i sostenible i el foment de la innovació", diu que cal reduir la bretxa digital per a garantir la igualtat d'accés a la informació i el coneixement.

[noticiadiariambautor]93/225[/noticiadiariambautor]

Com encaixa internament la promoció digital del català en un context de diglòssia a favor del castellà? Un context de diglòssia a favor d'una altra llengua es pot veure igual que el context d'un mercat

global dominat per llengües majoritàries. Segurament, la majoria de les aproximadament 5.800 llengües de què parlava abans es troben en una situació de diglòssia respecte les llengües oficials de l'estat al qual pertanyen. En el cas del català, a més a més, **ens trobem ni més ni menys que a l'ombra de la quarta llengua més parlada del món**. Això vol dir que el relat ha de ser el mateix: la diversitat lingüística s'ha de preservar a nivell mundial i també al nivell de cada estat, i els parlants d'una llengua minoritària han de tenir el mateix dret a l'accés a les tecnologies de la informació que els parlants de les llengües més majoritàries. La minorització d'una llengua dins l'estat al qual pertany s'ha d'evitar en tots els àmbits, i això inclou l'àmbit de la digitalització. Que com a ciutadà et puguis dirigir a un assistent virtual en la teva pròpia llengua és tan important com que ho puguis fer amb un treballador públic.

En el cas de l'àmbit digital és també important analitzar i treure profit del contacte lingüístic amb el castellà. D'una banda, la gran similitud a nivell fonètic, fonològic, lèxic i sintàctic de totes dues llengües fa que la substitució del català pel castellà sigui molt pronunciada, i que el català cada vegada estigui més *castellanitzat*. Però de l'altra, **podem aprofitar totes aquestes similituds per a reutilitzar al màxim els recursos computacionals del català**. La precisió de la traducció automàtica, per exemple, serà molt més elevada si es fa entre llengües de la mateixa família lingüística, de manera que **la qualitat de la traducció entre català i castellà pot ser molt alta**. De la mateixa manera, tant per a sistemes automàtics de traducció com de diàleg, de reconeixement o síntesi de la parla, que són la base, per exemple, dels assistents virtuals, podem utilitzar tècniques actuals de transferència d'aprenentatge aprofitant la gran quantitat de recursos de llengües majoritàries, que funcionaran més bé entre llengües amb un alt grau de similitud lingüística.

Un altre tema de debat és el dels canals audiovisuals. No podem entendre el català al món digital sense tenir en compte els joves *youtubers*, que són un puntal per a l'ús i la difusió de la llengua. És un tema que requereix sobretot una anàlisi sociolingüística, d'autoestima de la llengua, i de mercat. Però no hem d'oblidar que fins i tot aquí la tecnologia hi pot tenir el seu paper. Sense anar més lluny, el Canal Malaia aprofita els algorismes d'aprenentatge automàtic de *YouTube* per a agrupar vídeos en català i aconseguir que les recomanacions d'altres vídeos que fa la plataforma es basin també en el paràmetre de la llengua. També, a nivell de recerca, cal fer una aposta ràpida i contundent perquè el doblatge automàtic sigui una realitat a mig termini. Però tot això segurament ja es mereix un altre espai de debat.

[noticiadiariambautor]93/223[/noticiadiariambautor]

#### Bibliografia rellevant:

- Farrús, M. (21/11/2020). "Sobirania tecnològica, també". *Vilaweb*.  
<https://www.vilaweb.cat/noticies/sobirania-tecnologica-tambe/>  
Junyent, Carme; Zaballa, Bel. *El futur del català depèn de tu*. Edicions la Campana, Barcelona 2020. ISBN: 9788416863754.
- Pinter, A. (13/10/2019). "Catalunya capta uns 200 centres tecnològics". *L'Econòmic*.  
<https://www.leconomic.cat/article/1679078-catalunya-capta-uns-200-centres-tecnologics.htm>
- Puigneró, J. (07/09/2020). "Petits, frugals i digitals". *El Punt Avui*.  
<http://www.elpuntavui.cat/opinio/article/1845323-petits-frugals-i-digitals.html>
- United Nations (2021). Transforming Our World: The 2030 Agenda for Sustainable Development. A/RES/70/1.  
<https://sustainabledevelopment.un.org/content/documents/21252030%20Agenda%20for%20Sustainable%20Development%20web.pdf>
- UPF (06/11/2020). "Catotron, el primer sistema de síntesi de veu lliure i obert, basat en xarxes neuronals". Notícies UPF. [https://www.upf.edu/recercaupf/-/asset\\_publisher/RVNXhLpxnc9g/content/id/238849374/maximized#.YKT3HagzY2w](https://www.upf.edu/recercaupf/-/asset_publisher/RVNXhLpxnc9g/content/id/238849374/maximized#.YKT3HagzY2w)
- Vilaweb (14/08/2020). "La revolució del nou traductor neuronal anglès-català". *Vilaweb*.

<https://www.vilaweb.cat/noticies/tecnologia-de-xarxes-neuronals-la-revolucio-de-la-traduccio-automatica-en-obert-a-softcata/>

Articles del dossier:

[noticiadiariambautor]93/234[/noticiadiariambautor]  
[noticiadiariambautor]93/233[/noticiadiariambautor]  
[noticiadiariambautor]93/229[/noticiadiariambautor]  
[noticiadiariambautor]93/232[/noticiadiariambautor]  
[noticiadiariambautor]93/226[/noticiadiariambautor]  
[noticiadiariambautor]93/231[/noticiadiariambautor]  
[noticiadiariambautor]93/230[/noticiadiariambautor]  
[noticiadiariambautor]93/227[/noticiadiariambautor]  
[noticiadiariambautor]93/223[/noticiadiariambautor]  
[noticiadiariambautor]93/224[/noticiadiariambautor]  
[noticiadiariambautor]93/222[/noticiadiariambautor]  
[noticiadiariambautor]93/225[/noticiadiariambautor]