

La sobirania tecnològica per al català: quines opcions tenim?

Mireia Farrús

Universitat de Barcelona

Les aplicacions tecnològiques amb alguna finalitat lingüística com ara traductors automàtics, sistemes de diàleg, reconeixadors de la parla, generadors de resums automàtics, etc. tenen normalment una arquitectura comuna a totes les llengües. Cada aplicació, aleshores, s'adapta mitjançant corpus específics per a cada llengua. Per tant, es necessiten dos components: l'arquitectura i les dades, i tant l'un com l'altre constitueixen recursos que poden ser de propietat o lliures. En el cas del programari lliure, l'usuari té accés al codi font i el pot modificar a la seva conveniència o distribuir-lo lliurement. També pot ser, però, que l'usuari tingui accés al codi font però no pugui modificar-lo ni distribuir el producte, de manera que tindriem un recurs de codi obert però no lliure. Els seus usos i possibilitats es determinen a través de la llicència que s'atorga a cada producte.

Els productes comercials de propietat tenen una gran tendència a afavorir les llengües dominants i majoritàries, mentre que les altres queden arraconades en l'àmbit tecnològic. Si les úniques opcions que tenim, aleshores, són els productes de propietat, la inversió necessària tant de les institucions com de l'usuari final és enorme, sobretot econòmicament. Per això els recursos lliures en llengües minoritàries i minoritzades són tan importants: permeten crear i gestionar recursos lingüístics sense dependre de grans empreses del sector; és a dir, gaudir de sobirania tecnològica. La sobirania tecnològica és un corrent social i polític que defensa i promou l'autogestió per part dels ciutadans de les infraestructures tecnològiques. En el cas de les aplicacions lingüístiques, rau en la creació i la gestió de programari i recursos lliures perquè els ciutadans a nivell individual, les institucions i també les empreses els puguin utilitzar per a crear productes en una determinada llengua. Però, quines opcions tenim per al català?

En el cas del programari lliure, l'usuari té accés al codi font i el pot modificar a la seva conveniència o distribuir-lo lliurement.

Softcatalà (<https://www.softcatala.org/>) és l'exemple paradigmàtic de sobirania tecnològica en català. Fundada en 1998 com a associació sense ànim de lucre, la seva missió principal és el foment de l'ús i la presència del català en la informàtica, internet i les noves tecnologies, amb prioritat per al programari lliure. Softcatalà treballa en la traducció al català de programari, i proporciona eines lingüístiques com correctors i traductors automàtics, diccionaris i tecnologies de la parla, entre altres. Recentment, l'associació ha llançat un nou traductor automàtic basat en xarxes neuronals, que és la tecnologia més puntera per a aquest tipus d'aplicacions.

Un altre exemple més recent de sobirania tecnològica és la cooperativa Col·lectivaT (<https://collectivat.cat/>), que

des de l'any 2017 ofereix recursos computacionals en reconeixement i síntesi de la parla, traducció automàtica i comprensió del llenguatge natural. Els seus membres també ofereixen tallers sobre tecnologies lingüístiques per a desenvolupadors, emprenedors i usuaris. Fa pocs mesos, Col·lectivaT va presentar el *Catotron*, el primer sistema de síntesi de veu lliure basat en xarxes neuronals, que, com en el cas de la traducció automàtica, és la tecnologia més avançada en aquest tipus d'aplicacions.

A un nivell més institucional, el Departament de Polítiques Digitals de la Generalitat de Catalunya ha impulsat molt recentment el projecte AINA —en homenatge a la filòloga Aina Moll— amb la col·laboració del Barcelona Supercomputing Center. L'objectiu de l'AINA és proporcionar recursos digitals en català perquè les empreses o centres de recerca les puguin utilitzar per a entrenar algorismes d'intel·ligència artificial basats fonamentalment en xarxes neuronals i poder crear així aplicacions avançades, sobretot en els camps de la traducció automàtica i de les tecnologies de la parla. De fet, llengües majoritàries com

l'anglès ja tenen al seu abast aquest tipus de recursos generats per institucions que se'n fan càrrec; així, grans empreses com Google, Microsoft i Apple utilitzen recursos digitals creats per la DARPA (l'Agència de Projectes de Recerca Avançada del Departament de Defensa dels EUA) per a generar els seus models d'intel·ligència artificial. Qualsevol país, nació o territori amb una llengua pròpia, per petit que sigui, ha de tenir aquest suport institucional lingüísticament, també en el camp digital.

La sobirania tecnològica, però, no és exclusiva de les aplicacions informàtiques, sinó que es pot estendre a altres sectors digitals com l'audiovisual. En un món dominat per les llengües majoritàries com el dels *youtubers*, iniciatives com la del Canal Malaia són cabdals. Els algorismes de les xarxes prioritzen les recomanacions automàtiques dels vídeos amb moltes visites, però com que en català hi ha pocs vídeos i es troben dispersats, és difícil que la xarxa generi recomanacions de vídeos basant-se en la llengua. Per tal de fer front a això, el Canal Malaia ajuda a visibilitzar altres vídeos en català quan se n'ha vist un en aquesta llengua, de manera que l'algorisme aprèn a

relacionar vídeos en aquesta mateixa llengua i pot generar les recomanacions adequades.

En un món digital i globalitzat com el que ara tenim, es farà difícil evitar l'extinció d'una llengua si no se n'evita també l'extinció digital, i sembla que en el perill d'aquesta segona ja s'hi troben unes quantes llengües europees. En aquest sentit, la sobirania tecnològica és un benefici per a qualsevol llengua, però sobretot per a les minoritzades, perquè la disponibilitat de recursos lliures els atorga una independència que és crucial a l'hora de fer front a reptes generats per la intel·ligència artificial i altres sectors digitals, en què la quantitat de recursos disponibles és determinant per a la qualitat del producte final. Les iniciatives sorgides de diferents àmbits de la ciutadania, cooperatives, voluntariat, etc. són importantíssimes, i sovint molt més properes a l'usuari final, però el suport institucional també és imprescindible si volem tenir una llengua moderna, plenament normalitzada digitalment i a l'abast de totes les nostres pantalles. Però al final és feina de tots; com a usuaris hem de reclamar el dret de poder viure en català també a través de la tecnologia, i utilitzar tots els recursos disponibles que ajuden a fer-ho possible. ■

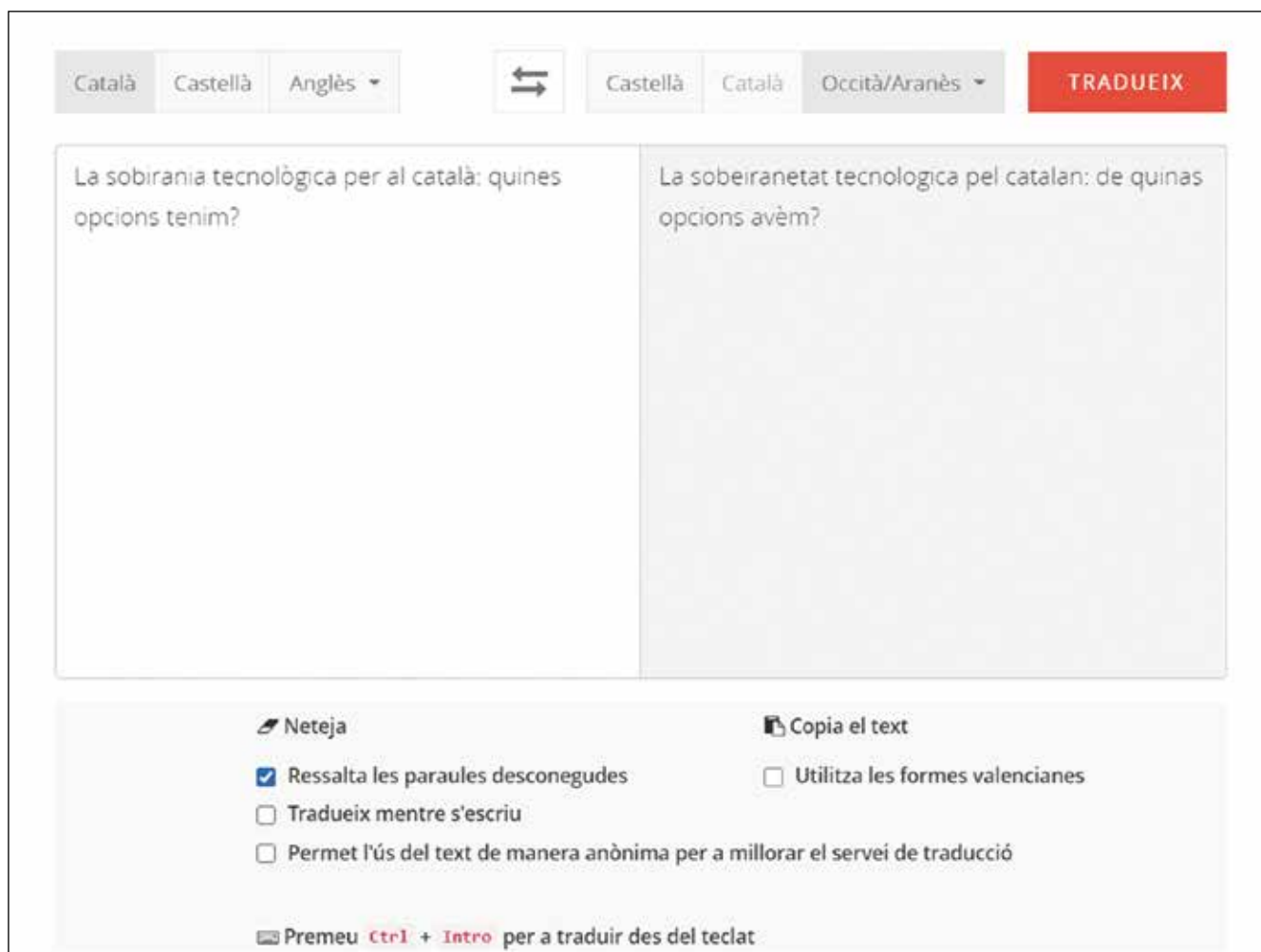


Figura: Traductor automàtic de Softcalà