



La llengua catalana en el nou paradigma digital. Els reptes compartits de les llengües minoritàries

Jordi ASCENSI i SALA



Resum executiu

Durant aquesta ponència, realitzada durant la 37a Diada Andorrana de la Universitat Catalana d'Estiu a Prada, es va abordar quina era la situació del català, en el seu context de llengua minoritària i minoritzada, quan ens enfoquem a la realitat del que associem com a paradigma digital, entenent aquest com el conjunt de sistemes i dispositius que manipulen i gestionen informació en format digital (discret), així com totes les interaccions que en el nostre context social s'hi donen. La llengua, com a element determinant d'aquestes interaccions, té un rol fonamental i crític i és necessari de comprendre'n com es afectada. Les particularitats del context digital, en el sentit social, econòmic o cultural, així com les lògiques específiques que s'hi donen, fan que les llengües minoritàries tinguin uns condicionaments diferents als que es donen en el context del que podem considerar com el món no digital.

La ponència va descriure, de forma resumida, l'evolució del paradigma digital a partir dels anys 70 i fins el moment actual, i com s'hi ha anat manifestant la llengua catalana i, per homologia, bona part de les llengües minoritàries del món occidental. També es va comparar la presència del català en l'àmbit del programari existent i del contingut generat dins del context digital amb altres llengües majoritàries i minoritàries, fent èmfasi en la preeminència de l'anglès com a llengua hegemònica que, a més a més, ha anat guanyant protagonisme en els darrers anys. S'hi van presentar quines són les lògiques de funcionament dins del paradigma digital, els riscos derivats d'aquesta situació, i quines situacions es presenten tenint en compte l'adveniment de la intel·ligència artificial i de l'ús massiu de dades. També es van apuntar a quines són les estructures i dinàmiques de poder existents en l'àmbit digital, i quins en són els actors rellevants. Finalment, es van proposar alguns àmbits de treball per fer que la llengua catalana es mantingui viva dins del context digital, i de quina manera es pot col·laborar amb altres llengües minoritàries per tal de mantenir-ne la rellevància.

El paradigma digital

Per poder emmarcar el contingut de la ponència cal posar un mínim de context al que anomenem paradigma digital i com ha anat evolucionant aquest amb el pas del temps. De fet, el lligam que hi ha entre el desenvolupament i ús de les tecnologies de la informació i de la comunicació (les TIC) i l'ús de les llengües minoritàries, també ha anat evolucionant.

La següent figura mostra, de forma molt resumida i acotada en dècades a partir dels anys 70 del s. xx, l'evolució de les TIC i algunes fites que serveixen com a referències.

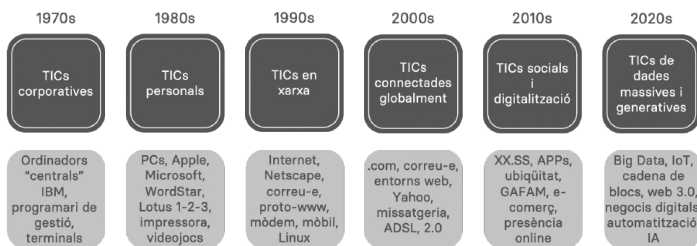


Figura 1: Evolució de les TIC des dels 70 fins al moment actual (Font: Jordi Ascensi)

El que podem comprovar és que hem passat d'unes poques aplicacions (programaris) i sistemes (maquinari) especialitzats i orientats a tasques de tipus administratiu, enfocant-se principalment a la gran empresa i a les grans administracions públiques, per anar poc a poc cap a una generalització de l'ús de les TIC en altres entorns, principalment el d'àmbit personal i el de les petites i mitjanes empreses, amb una expansió exponencial a partir de la dècada del 2000 i, un xic més endavant, amb l'ús massiu de les xarxes socials i els dispositius mòbils. No és objecte d'aquesta ponència parlar de la història de les TIC *per se*, però és important de tenir-la com a base de referència per a contextualitzar com ha estat aquesta evolució quan parlem del cas particular del català i, en general, de les llengües minoritàries.

Prenent com a base el cronograma anterior, la figura següent mostra, precisament, aquesta evolució quan parlem de l'ús del català en el context digital:

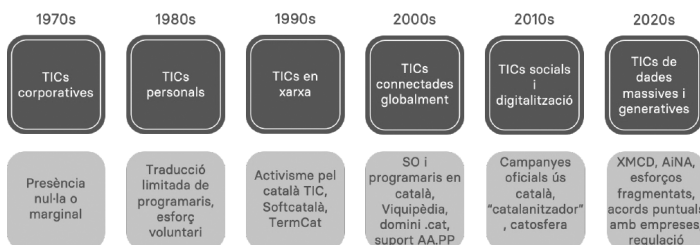


Figura 2: Evolució de l'ús del català dins de les TIC des dels 70 fins al moment actual (Font: Jordi Ascensi)

El que podem veure, de forma molt resumida, és una evolució contextualitzada de l'ús de la llengua catalana en funció del temps. Podem comprovar que en els inicis de les TIC, les aplicacions i sistemes emprats no estaven en català degut, principalment a la manca de mercat, a la quasi inexistent indústria de producció de maquinari i de programari a Espanya,¹ i al context polític i cultural de final de la dictadura del general Franco. En els anys posteriors, en un esforç de voluntarisme i associacionisme similar al que es va desenvolupar en d'altres sectors com el de la literatura o el dels continguts audiovisuals i també dins de l'àmbit de l'acadèmia (des de la Universitat Politècnica de Catalunya principalment), van anar apareixent activistes que anaven traduint al català els programaris més utilitzats i se n'anaven desenvolupant d'altres d'específics en l'àmbit educatiu. L'any 1998 neix Softcatalà,² una entitat sense ànim de lucre amb la voluntat de millorar la presència del català a les noves tecnologies, i que ha estat un referent en el desenvolupament d'iniciatives de programari en llengua catalana. La Generalitat de Catalunya també hi va posar del seu costat amb el centre de terminologia TermCat,³ que actua com a referència normalitzadora sobre l'ús dels termes de l'àmbit TIC en llengua catalana. Posteriorment va aparèixer la Viquipèdia,⁴ la tercera edició en llengua no anglesa en crear-se l'any 2001, i que actualment ocupa el 20è lloc mundial en nombre d'articles, o el domini '.cat',⁵ introduït l'any 2005, el primer domini d'internet d'àmbit lingüístic i cultural, acompanyant del domini '.ad',⁶ activat l'any 1996, i que és el domini de país del Principat d'Andorra. En la darrera dècada i escaig, i fins al moment actual, les activitats lligades a la promoció, difusió i reconeixement de la llengua catalana s'han anat multiplicant, fent del català un exemple de referència en l'àmbit d'ús de les llengües minoritàries en l'entorn digital, seguint de prop les tendències del mercat i abastant àmbits de rabiosa actualitat com la monitorització de la presència de les webs en català als cercadors com Google o Bing, a través de la Xarxa de Monitorització del Català Digital,⁷ o l'ús del català en els models grans de llenguatge amb la iniciativa AiNA.⁸

Contingut i continent

Quan parlem d'ús de la llengua en el paradigma digital o dins de les TIC, hem de pensar en dos àmbits diferents però que, en certs casos específics, s'interconnecten entre ells.

- **Àmbit de les aplicacions i sistemes amb interfícies en català:** aquí estem parlant que els sistemes operatius dels dispositius digitals, els programaris o les aplicacions que instal·lem als nostres dispositius mòbils estiguin en català i que, a més a més, les interaccions que hi puguem fer es puguin realitzar en català. Aquestes interaccions poden ser a través del teclat (comandes escrites en català) o amb el ratolí (en aquest cas concret parlem que els entorns on s'utilitza tinguin les icones o quadres de diàleg en català) o, un cop més, a través de la parla (que ens puguem adreçar al sistema a través d'un sistema de processament del llenguatge natural⁹). És el que anomenarem metafòricament com a *continent* en aquest document, i que aglutina tot l'entorn d'interacció amb l'usuari de les TIC.
- **Àmbit del contingut digital generat en català:** en aquest cas, aquí estem parlant de tot el contingut generat en l'àmbit digital, sigui text, imatge amb text, o contingut multimèdia que contingui la llengua catalana, estigui generat amb programari fet en català o no. Ho podem anomenar de forma directa com a *contingut*.

És important fer aquesta separació perquè quan parlem de context digital, cal definir si ens referim a un àmbit o a un altre. La generació d'interfícies d'usuari, el continent, en llengua catalana depèn, fonamentalment, de la iniciativa i voluntat de les entitats i persones que dissenyen i implementen els diferents tipus de programari, principalment empreses i persones que tenen coneixements de programació. La generació de contingut digital, el contingut, depèn principalment de les, precisament, persones i entitats que tenen la denominació de generadores de contingut i que no requereixen expressament de coneixements de programació per a generar-los. Hom pot generar contingut en llengua catalana a través d'un programari en llengua anglesa, o pot generar contingut en llengua francesa a través un programari dissenyat en català, i, així, amb totes les possibles combinacions que puguin resultar. La separació entre contingut i continent ens és prou útil per veure la presència de la llengua catalana en ambdós àmbits i quina és la comparació amb les altres llengües, siguin aquestes minoritàries o no.

L'European Language Equility¹⁰ (ELE) és un consorci format per diverses entitats de recerca europees i suportat pel finançament de la Unió Europea que té com a objectius estudiar i a fomentar la igualtat entre llengües, dins de la UE, en el context digital. Al seu lloc web proposen diferents eines per veure l'estat de la situació de les diferents llengües prenent, precisament la separació proposada anteriorment: *dataset* com a contingut i *software* com a continent.

La figura següent mostra l'estat comparatiu de diferents llengües, incloent la catalana, basada en nombre de conjunts de dades i programaris publicats en la llengua en qüestió i com es comparen entre elles (l'ELE disposa, al seu lloc web, d'una eina que permet fer aquest tipus de comparació).

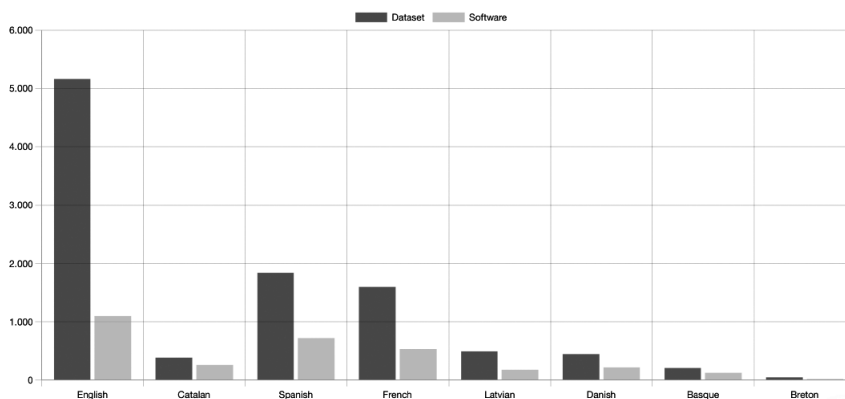
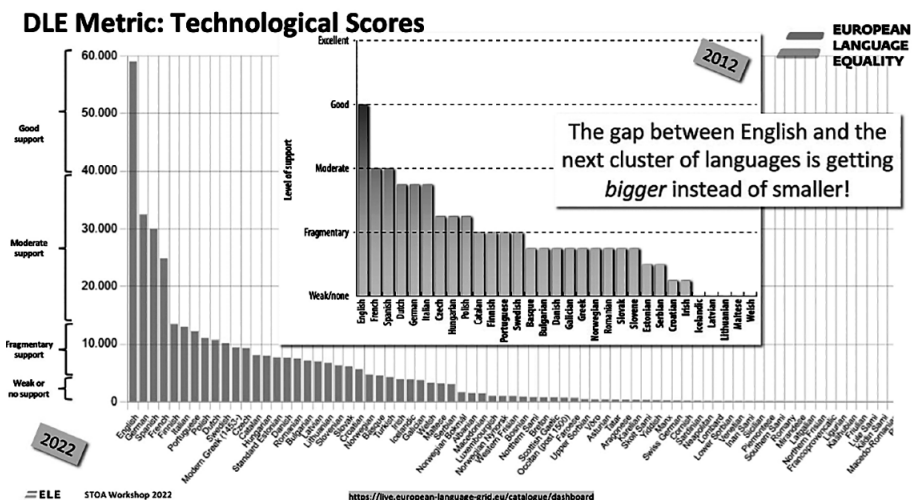


Figura 3. Font: European Language Equality (juny 2024)

Com podem observar, la llengua anglesa és predominant tant en el nombre de continguts generats com en el del programari existent, molt per sobre d'altres llengües majoritàries com poden ser el castellà o el francès. També podem veure la situació de la llengua catalana respecte d'altres llengües minoritàries europees com el basc, el bretó o fins i tot el letó o el danès. La comparació fa evident la preponderància de l'anglès com a principal llengua de publicació i, també, d'interfície d'usuari per als programaris. Aquest aspecte evidencia l'ús de la llengua anglesa com a *lingua franca* a Europa i, per extensió a la resta del món, també en l'àmbit digital, i que reflecteix les regles de mercat existents, on les entitats desenvolupadores de programari, i aquelles que fan desenvolupament de continguts, prioritzen l'ús de llengües majoritàries i, si el nombre de clients ho requereix, en desenvolupen el contingut i el continent en altres llengües existents.

En aquest sentit, aquestes entitats no només desenvolupen contingut i continent si hi ha mercat (si així fos, molt probablement no n'hi hauria en cap llengua minoritària, sobretot d'aquelles que comparteixen espai cultural amb llengües majoritàries), sinó que també responen a interessos de marca, a facilitats que es puguin posar per al seu ús o, en alguns casos, als incentius econòmics que se'n puguin derivar (p.ex. suport a la traducció de programaris o a la generació de continguts).¹¹

La següent figura mostra un estudi comparatiu del suport, en l'àmbit digital (entès com els recursos dedicats a la generació de contingut o a la generació de programari, i el seu resultat efectiu final) a les diferents llengües d'Europa, realitzat en dues onades, l'any 2012 i l'any 2022.



Podem veure immediatament la gran diferència entre la llengua anglesa i la resta de llengües, que s'amplia encara més en la darrera onada de l'estudi de l'any 2022. La llengua anglesa guanya rellevància, de forma preponderant, en l'àmbit digital i les principals raons es poden resumir en l'origen anglosaxó de les principals plataformes de programari i de generació de continguts, les lògiques de mercat que es generen amb la utilització de la llengua anglesa versus la relativa petita mida d'aquestes en les altres llengües (més contingut en anglès vol dir més públic objectiu i, per tant, més valor econòmic de la informació en qüestió), i l'amplificació d'aquest predomini a través de l'ús massiu de dades i la intel·ligència artificial en el moment actual.

El nou paradigma digital: dades massives i intel·ligència artificial generativa

El paradigma digital actual està virant ràpidament cap a l'ús de quantitats ingents de dades, digitalitzades a partir de continguts ja existents però, també, generades només en l'àmbit digital, i l'adopció accelerada de sistemes d'IA que són capaços de generar continguts per si mateixos. Aquesta situació genera tensions en l'ús i la presència de les llengües minoritàries, i provoca, de retruc, un conjunt de situacions que poden suposar un problema per a aquestes. Els principals riscos identificats són els següents:

- **Uniformització lingüística digital.** La generació massiva de dades i el desenvolupament d'IA es basen, gairebé exclusivament, en llengües amb gran presència digital i especialment l'anglès. Això provoca un procés d'uniformització on es privilegia l'ús de llengües dominants i es minimitza o ignora la diversitat lingüística.

Els sistemes de reconeixement de veu, traducció automàtica o assistents virtuals funcionen millor en llengües com l'anglès, el xinès o el castellà, perquè han estat definits i entrenats amb dades d'aquestes llengües, i sovint no incorporen llengües minoritzades com el català, el bretó o el gaèlic, per les mateixes regles que s'han explicat en el punt anterior. Plataformes globals (Google, Meta, Amazon, etc.¹²) no ofereixen serveis complets en llengües amb menys parlants digitals. Aquestes empreses estan aconseguint una hegemonia molt palesa dins de l'àmbit digital i les seves lògiques de funcionament, així com les estructures de poder que les generen, no responen a voluntats de diversitat cultural. A més a més, han anat ampliant la seva capacitat d'influència geopolítica els darrers anys, cosa que ha contribuït a la definició d'una societat *tecnocapitalista*.¹³ Això crea un entorn digital homogeni que afavoreix l'ús de llengües dominants i debilita la vitalitat digital de les minoritàries.

- **Llengües en risc d'extinció digital.** L'extinció digital és un fenomen real: moltes llengües que es mantenen vives en l'àmbit oral estan desapareixent de l'espai digital perquè no hi ha prou dades per entrenar tecnologies que les incorporin, o aquestes no tenen una representativitat de valor econòmic que serveixi d'incentiu per fer-les servir. Les llengües sense presència digital es tornen invisibles a la IA i als sistemes automatitzats, cosa que contribueix a la seva desaparició.

La manca de corpus textuais i orals digitals en llengües minoritàries fa que no es puguin aplicar-hi eines d'IA com la traducció automàtica, els xatbots o la síntesi de veu.

Aquest risc és encara més alt per a llengües amb pocs parlants o sense suport institucional.

- **Discriminació cultural digital.** Quan la IA i els sistemes digitals no tenen en compte la diversitat lingüística i cultural es produeix una discriminació estructural que afavoreix cultures dominants i invisibilitza les altres.

Els algorismes d'IA reproduïen biaixos culturals, ja que es basen en dades majoritàriament generades en cultures dominants, tant en el seu contingut com en la seva forma.

Els continguts en llengües minoritàries tenen menys visibilitat als cercadors, xarxes socials i serveis de *streaming* perquè els algorismes de classificació tendeixen a menystenir-los en tenir menys rellevància estadística global.

Això impedeix la transmissió de coneixements i valors culturals propis a través dels canals digitals, afectant la cohesió i la identitat col·lectiva dels parlants de les llengües minoritàries.

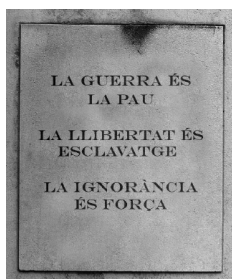
- **Fractura digital i exclusió social.** La manca de tecnologia i continguts en llengües minoritàries contribueix a una fractura digital que exclou comunitats senceres dels avantatges de la transformació digital i de la IA.

Les persones que volen o necessiten usar el català, el basc o altres llengües minoritzades es troben sovint amb interfícies no traduïdes, eines no accessibles o serveis que no entenen la seva llengua.

Això afavoreix l'abandonament de la llengua pròpia en àmbits digitals, especialment entre els joves, que són usuaris intensius de tecnologia però, també, en les persones grans, que poden sentir-se excloses quan les interfícies i els seus continguts no estan en la seva llengua pròpia, cosa que amplia la fractura digital generacional.

La manca de recursos en llengua pròpia pot generar barreres d'accés a l'educació, a l'administració electrònica i al mercat laboral digital.

Aquest riscos poden abocar-nos a una tendència a la uniformització i homogeneïtzació d'una forma única de cultura i d'entendre el món basada en les llengües majoritàries. Com a evocació literària lligada a tot l'anterior, podríem fer servir una imatge que resumeix el dogma de la 'Novaparla d'Oceània' que apareix a la famosa novel·la de George Orwell 1984. L'autor defineix una nova llengua dissenyada per fer impossible el pensament lliure i l'expressió d'idees, i que es basa en l'eliminació sistemàtica de paraules que puguin expressar conceptes contraris al règim imperant, la simplificació extrema de la gramàtica i el vocabulari, la creació de paraules compostes que substitueixen conceptes complexos, i l'eliminació de qualsevol ambigüitat o matisos en el llenguatge, fent valer allò que si les paraules no existeixen, es tornen literalment impensables.



Una mirada al futur

En un context on la digitalització i la intel·ligència artificial estan transformant radicalment la manera com interactuem i processem la informació, el futur de les llengües minoritàries com el català es troba en una cruïlla decisiva. La presència i vitalitat d'aquestes llengües en l'entorn digital requereix d'una estratègia coordinada i multidimensional que impliqui tots els actors rellevants, des de les administracions públiques fins al sector privat, passant per la societat civil i l'acadèmica.

A continuació es presenten sis possibles línies d'actuació per garantir la supervivència i el desenvolupament del català, i d'altres llengües minoritàries, en el nou paradigma digital.

- **Parlem-nos - Coordinació en l'àmbit del català.** És necessari coordinar els esforços que s'estan fent entre tots els territoris de parla catalana per mantenir i potenciar la presència del català en l'entorn digital. Cadascun dels territoris té una sèrie de capacitats i de particularitats que és d'interès poder activar en funció del cas d'ús. Per exemple, Andorra pot assegurar la presència de la llengua catalana en l'àmbit de les institucions internacionals; Catalunya i el País Valencià tenen una capacitat d'activar recursos de forma més massiva, i les Illes Balears, la Catalunya Nord o l'Alguer tenen un bon munt d'entitats activistes en favor de la llengua. S'han de mancomunar aquests esforços perquè puguin generar més impacte en l'àmbit digital, sabedors de les estructures de poder que el mouen i a on el treball coordinat és clau.

- **Parlem-ne - Coordinació entre llengües minoritàries.** Cal establir col·laboracions amb altres comunitats lingüístiques minoritàries europees per compartir recursos, experiències i estratègies en l'àmbit digital. La situació del català en l'àmbit digital és molt propera a la del danès, el basc o el maltès, i ben segurament, moltes de les iniciatives poden ser compartides entre les diferents llengües, així com els recursos necessaris per poder-les portar a terme. Mancomunar, un cop més, esforços és cabdal per poder tenir l'impacte desitjat.

- **Harmonitzar esforços.** És crucial la col·laboració entre administracions públiques, universitats, societat civil i sector privat, com s'ha demostrat amb iniciatives com Softcatalà, TermCat, i el projecte AiNA. L'aplicació de la coneguda col·laboració multipartida de la quàdruple hèlix,¹⁴ permet abordar la problemàtica des de diferents perspectives tot posant com a interès comú la protecció de la llengua.

- **Programa comú europeu.** Cal desenvolupar un programa coordinat a escala europea per a les tecnologies del llenguatge en l'àmbit digital, similar a les iniciatives de l'European Language Equality., enfocat a les particularitats de les llengües minoritàries i tenint en compte l'aparició de noves tecnologies com la IA i el nou paradigma del contingut digital.

- **Fomentar, facilitar, promoure, regular i generar mercat.** És necessari continuar posant recursos i aplicant esforços en la generació d'un ecosistema que incentivi l'ús del català en l'entorn digital, tant en la creació de contingut com en el desenvolupament de programari. L'incentiu de la compra pública de programari i de contingut en llengua catalana pot servir com a vector d'interès perquè les empreses en vulguin desenvolupar les interfícies necessàries. Al mateix temps, la llengua ofereix un aspecte de

diferenciació i de personalització que pot servir com a argumentari per fer-ne una proposta de valor.

- **Model publicoprivat amb dades obertes com a bé comú.** Cal establir un model de col·laboració entre el sector públic i privat que permeti l'accés i l'ús de dades en català com a recurs compartit. Això és especialment important en el context actual de l'ús de la intel·ligència artificial i del processament massiu de dades.

La cita del filòsof de la llengua alemany Ludwig Wittgenstein, "els límits del meu llenguatge defineixen els límits del meu món", adquireix una nova dimensió en l'era digital. Si el llenguatge determina la nostra capacitat de comprendre i interactuar amb el món, la presència o absència d'una llengua en l'entorn digital definirà, inevitablement, els límits del nostre món digital. En un moment on la intel·ligència artificial i les dades massives estan redefinint la manera com processem la informació i ens comuniquem, garantir la vitalitat del català, i per extensió la d'altres llengües minoritàries, en aquest nou paradigma no és només una qüestió de preservació cultural, sinó de garantir que les seves comunitats lingüístiques puguin participar plenament en la construcció del futur digital. Només a través de la continuació d'accions coordinades, continuades, adaptades i decidides entre totes les persones i entitats implicades podrem assegurar que els límits que van apareixent en el context digital al nostre llenguatge no esdevinguin una visió limitada de la nostra participació en el món digital del present i del demà.

Jordi Ascensi i Sala,

enginyer de telecomunicacions, coordinador de Tecnologia d'Andorra Recerca+Innovació

Bibliografia de suport

- ALFONSI, F. (2018). *Report on endangered languages and linguistic diversity in the European Union*. European Parliament Committee on Culture and Education.
- BASTARDAS-BOADA, A. (2019). *From Language Shift to Language Revitalization and Sustainability*. Edicions UB.
- CRYSTAL, D. (2022). *Language and the Internet*. Cambridge University Press.
- European Language Equality Project. (2022). *Digital Language Equality in Europe by 2030*.
- JUNYENT, C. (2020). *El futur del català depèn de tu*. La Campana.
- MOROZOV, E. (2018). *Capitalismo Big Tech: ¿Welfare o neofeudalismo digital?* Enclave de Libros.
- UNESCO. (2023). *Atlas of Languages in Danger in the Digital Age*.

Notes

- 1- Una honrosa excepció la trobem en l'ordinador Factor-P, desenvolupat l'any 1966 per Joan Majó, fundador de l'empresa Telesincro i que posteriorment va exercir el càrrec de ministre d'Indústria i Energia al Govern d'Espanya, i l'especialista en maquinari Jordi Vidal. Font: <https://www.enciclopedia.cat/tecnic-i-tecnologia-en-el-desenvolupament-de-la-catalunya-contemporania/la-historia-de-la>
- 2- <https://www.softcatala.org>
- 3- <https://termcat.cat/ca>
- 4- <https://ca.wikipedia.org/wiki/Viquipèdia>

- 5- <https://fundació.cat>
- 6- <https://www.domini.ad>
- 7- <https://fundació.cat/xmcd/>
- 8- <https://projecteaina.cat>
- 9- <https://www.termcat.cat/oc/cercaterm/fitxa/NDUxMjU3MA%3D%3D>
- 10- <https://european-language-equality.eu>
- 11- Podem trobar dos exemples d'ambdós àmbits dins del cas del català. El primer, amb el suport explícit de la Generalitat de Catalunya a l'empresa Microsoft per tenir els seus programaris en català (2006; <https://www.plataforma-llengua.cat/que-fem/noticies/546/el-windows-vista-ja-es-troba-a-les-botigues-preinstallat-en-importants-marques-dordi-nadors>), o, darrerament amb la iniciativa dels premis Crit de la Fundació.cat (<https://fundació.cat/premis-crit-creacio-digital-catala/>)
- 12- Són les conegudes com a 'Gafam': Google, Amazon, Facebook (Meta), Apple i Microsoft. <https://ca.wikipedia.org/wiki/GAFAM>
- 13- Evgenyi Morozov ho explica en aquest article publicat al web del CCCB. <https://lab.cccb.org/ca/el-preu-de-la-utopia-digital/>
- 14- <https://www.termcat.cat/ca/neoloteca/fitxa/NDg3Mjg4OQ%3D%3D>