



EDUARDO VALENCIA TIRAPU

La tirania del cas únic **Llengua i intel·ligència artificial**

Premi València d'Assaig 2026

 **Diputació
de València**

Cultura



**institució
alfons
el magnànim**

València, 2026



Obra guanyadora del Premi València d'Assaig 2026 convocat per la Institució Alfons el Magnànim – Centre Valencià d'Estudis i d'Investigació, i atorgat per un jurat presidit per Francisco Teruel Machí, diputat de Cultura de la Diputació de València, i integrat per Carme Navarro Dolz, Clara Colomina Martínez, Amadeo Serra Desfilis, Xavier Agustí Campos i, com a secretària, Immaculada Carbó Fons.

Edició composta amb lletra Andralis ND OSF i Andralis ND SC, impresa en l'interior sobre paper Prinset Ivori de 90 g/m² i en la coberta sobre cartolina Twil blanc natural de 300 g/m².

© 2026, Eduardo Valencia Tirapu

© 2026, d'esta edició:
Institució Alfons el Magnànim
Centre Valencià d'Estudis i d'Investigació
Diputació de València
Corona, 36 — 46003 València
Tel.: +34 963 883 169
magnanim@dival.es
www.alfonselmagnanim.net

ISBN: 978-84-1156-136-5

DL: V-4063-2026

Disseny de la col·lecció: Fèlix Bella
Redisseny de les cobertes de la col·lecció: Joanrojeski Estudi Creatiu
Disseny de la coberta: Carles Rodrigo
Maquetació: Quinto A Estudio Gráfico
Impressió: Impremta Diputació de València



Índex

Pròleg	9
El món ja era probabilístic	19
El llenguatge com a probabilitat	27
La tirania del cas únic	35
Per què ens costa pensar en agregat	41
La modernitat del cas	49
El paral·lel europeu: la regulació de la IA com a mirall	55
El mapa que la IA no veu	59
El valencià: massa gran per morir, massa petit per viure tranquil	67
Els monstres especulars al País Valencià	77
El <i>business case</i> que ningú no vol fer	85
El català a tots els territoris: l'agregat complet	93
L'èuscar: una supervivència improbable	101
El gallec, l'aragonés, l'asturià, l'occità: l'error agregat que ningú no mesura	107
Dos miralls: Navarra i el País Valencià	111
La Constitució del 78 com a sistema probabilístic mal calibrat	117
L'Alacant invisible: el sud que desapareix	123
Les Illes Balears: el turisme com a dissolvent lingüístic	131
La Catalunya Nord: quan la protecció desapareix	137
El tall silenciós: la transmissió que no es va fer	141
Les perifèries extremes: la Franja, l'Alguer i el Carxe	147
El català digital: corpus, models i supervivència algorísmica	151
La llengua com a actiu econòmic: el que no es mesura no existeix	155
La generació que decidirà	159
Assessor, no jutge: la IA en la política lingüística	169
Una ètica de l'agregat per a les llengües	179
Cap a una autoritat tècnica de la política lingüística	185
Epíleg	195



Pròleg

Una llengua que vaig veure morir

El dàlmata va morir el 10 de juny de 1898, a les 16.30 hores, a l'illa de Krk. L'últim parlant es deia Tuone Udaina (Antòni Udina, en la forma romanç del seu nom) i tenia setanta-quatre anys. No va morir de vell. Va morir perquè una mina terrestre li va esclatar a prop mentre caminava per un sender de l'illa. Amb ell es va extingir una llengua romànica que havia sobreviscut mil anys entre el llatí vulgar de la costa adriàtica i l'empenta progressiva del croat i del venecià. Una llengua sencera (amb els seus verbs, les seues cançons de bressol, les seues maneres de dir «tinc fred» o «fa bon dia») va desaparèixer del món en un instant concret, en un lloc concret, amb el cos concret d'un home vell que ja no s'hi sentia bé i que parlava una variant que, segons els lingüistes que el van entrevistar anys abans, ja no era del tot dàlmata ni del tot venecià, sinó una cosa intermèdia, erosionada, a mig camí entre la memòria i l'oblit.¹

Conec el dàlmata. No perquè el parlara ningú quan jo el vaig estudiar, feia un segle que era mort, sinó perquè la meua faena m'ha portat a passar temps amb llengües que ja no tenen qui les parle. Soc filòleg de formació. Un filòleg amb una trajectòria estranya, que ha saltat de les llengües als ordinadors i dels ordinadors a les llengües, sense acabar d'encaixar del tot en cap dels dos mons.

¹ Sobre l'extinció del dàlmata i el testimoniatge de Tuone Udaina, vegeu M. G. BARTOLI, *Das Dalmatische: altromanische Sprachreste von Veglia bis Ragusa und ihre Stellung in der apennino-balkanischen Romania*, Viena, Holder, 1906.



Als vint-i-dos anys vaig entrar a treballar a Euskaltzaindia, l'Acadèmia de la Llengua Basca, en el projecte del *Diccionari General*, l'*Orotariko Euskal Hiztegia*.² M'hi vaig quedar deu anys. El director era Koldo Mitxelena. Recorde un matí concret, a la seu de la Reial Acadèmia a Bilbao, amb les fitxes de paper escampades per la taula gran. Mitxelena m'havia demanat que buscara totes les ocurrences del sufix *-tasun* en la documentació anterior al segle XVIII. Jo n'havia trobat catorze. Ell les va mirar, en va descartar tres perquè la datació era dubtosa, en va qüestionar dues més per la font, i va dir: «Amb nou exemples no es pot afirmar res. Amb nou exemples es pot sospitar.» Aquella frase em va marcar més que qualsevol seminari. Era Popper traduït al mètode lexicogràfic: les dades et permeten sospitar, no afirmar. I la diferència entre ambdues coses és la diferència entre un lingüista i un propagandista.

Aquella dècada em va ensenyar què significa documentar una llengua des de dins: la paciència de les fitxes, la disciplina de l'evidència, la humilitat davant d'un sistema lingüístic que sempre et supera. Després, per raons que tenen més a veure amb la vida que amb la vocació, vaig acabar a Nova York treballant amb models d'aprenentatge automàtic. El primer dia que vaig veure com un model estadístic predeia la paraula següent d'una frase, vaig sentir un vertigen familiar: era el mateix que féiem amb les fitxes de Mitxelena, però a una escala que cap humà no podia somiar. I amb la mateixa fragilitat. Un model entrenat amb dades esbiaixades genera resultats esbiaixats, igual que un diccionari construït amb fonts esbiaixades genera un mapa esbiaixat de la llengua.

Aquest assaig naix d'aquesta doble experiència: les llengües i les màquines que les processen, i del fet que ambdues m'han

² Koldo MITXELENA (dir.), *Orotariko Euskal Hiztegia / Diccionario General Vasco*, Bilbao, Euskaltzaindia, 1987-2005.



ensenyat la mateixa lliçó: que el que no es mesura es perd. No ho dic com un eslògan.

Ho dic com algú que ha vist llengües desaparèixer dels textos i dels models amb la mateixa lentitud silenciosa. I que veu, als territoris de parla catalana,³ una llengua amb deu milions de parlants potencials que retrocedeix cada dia en ús real, en transmissió familiar, en presència pública, i on la intel·ligència artificial està a punt d'accelerar aquest retrocés d'una manera que molt pocs han començat a entendre.

Escric aquest assaig des de Dénia, a la Marina Alta. Hi he viscut llargues temporades al llarg dels últims anys. Dénia és un lloc que condensa, en un radi de trenta quilòmetres, tota la complexitat lingüística valenciana. Al nucli antic, les dones grans encara parlen valencià entre elles al mercat del Peix. A la platja de les Marines, els cartells dels restaurants són en anglés, alemany i castellà. El valencià no hi apareix.

A les urbanitzacions del Montgó, els residents britànics i alemanys viuen vides senceres sense saber que la llengua del lloc on viuen no és el castellà. I als instituts, els adolescents canvien de llengua a meitat de frase amb una fluïdesa que és alhora admirable i alarmanant: admirable perquè demostra competència,

3 Al llarg d'aquest assaig faig servir el terme *valencià* quan em referisc específicament a la varietat lingüística pròpia del País Valencià, i *català* quan em referisc a la llengua en conjunt, que inclou també les modalitats central, balear, septentrional, alguerès i nord-occidental. La unitat lingüística del català-valencià-balear és consens científic universal, recollit pels seus organismes normatius (Institut d'Estudis Catalans i Acadèmia Valenciana de la Llengua) i per la filologia romànica. L'AVL, organisme estatutari valencià, ho va declarar formalment el 9 de febrer de 2005: la llengua pròpia i històrica dels valencians és, des del punt de vista filològic, la mateixa que comparteixen les comunitats autònomes de Catalunya i de les Illes Balears i el Principat d'Andorra. La denominació *valencià* és l'oficial al País Valencià per imperatiu de l'Estatut d'Autonomia (Llei Orgànica 5/1982). Aquesta denominació estatutària no contradiu la unitat lingüística, com l'AVL va aclarir explícitament.



alarmant perquè el castellà guanya terreny a cada canvi. Dénia m'ha ensenyat a veure la substitució lingüística no com una abstracció estadística sinó com un fenomen quotidià, observable des de la terrassa d'un café qualsevol del carrer Marqués de Campo.

A la farmàcia del carrer Diana, un matí qualsevol, he sentit quatre llengües en cinc minuts: una turista alemanya que demanava alguna cosa amb un gest de la mà perquè no trobava la paraula, el farmacèutic que li responia en castellà amb accent del sud, un client britànic que pagava en anglés, i una dona gran que, en eixir, m'ha dit «bon dia» en valencià, l'única frase en la llengua del lloc que he sentit en tota la transacció. Al supermercat de les Marines, les etiquetes dels productes estan en anglés, alemany i castellà. El valencià no hi apareix. No per hostilitat. Per irrellevància. La llengua del territori és invisible en el propi territori. I ningú no ho veu com una anomalia. Això és el que la normalització de la minorització produeix: l'absència es torna normal. I quan l'absència és normal, la presència es torna estranya.

Vaig néixer a Pamplona el 1963. Navarra és un lloc que s'assembla al País Valencià més del que ningú no admetria còmodament. A Navarra, la Ley Foral del Euskera divideix el territori en tres zones lingüístiques: bascòfona, mixta i no bascòfona. A la zona no bascòfona: la Ribera, el sud, el meu sud, l'ús habitual de l'euscar és del 0,8%. Zero coma vuit per cent. L'àrab, en canvi, arriba al deu per cent, associat a comunitats d'origen magrebí que s'hi han establert en les últimes dècades. El castellà ho domina tot.

Al País Valencià, la situació és estructuralment paral·lela. Hi ha una zona valencianoparlant i una zona castellanoparlant. L'ús del valencià a casa ha caigut del 43,4% el 1992 al 19,5% el 2021. La capacitat de parlar-lo ha baixat dotze punts en tres dècades. A les ciutats, València, Alacant, l'erosió és avançada. A les comarques interiors, Castelló, el sud de València, la llengua resisteix, però cada generació transmet menys.



A Navarra, UPN (Unión del Pueblo Navarro) i el Partido Popular han utilitzat durant dècades el fantasma d'ETA per atacar l'euscar: «la llengua dels terroristes», deien, com si una llengua poguera ser culpable de res. Al País Valencià, la coalició conservadora utilitza el fantasma del «pancatalanisme» per atacar el valencià: «imposició lingüística», diuen, com si ensenyar una llengua als seus propis parlants fora un acte d'opressió. El mecanisme és idèntic. Només canvia el nom de la llengua i el nom del monstre.

Aquest assaig proposa una idea que pot resultar incòmoda: que la crisi de les llengües minoritzades als territoris de parla catalana no és un problema d'identitat, ni de voluntat política, ni d'educació insuficient. És un problema de racionalitat. Més concretament: és un problema de la nostra incapacitat per pensar en el conjunt quan el cas singular ens crida amb més força.

Anomenaré aquest mecanisme la tirania del cas únic. La seua formulació és senzilla: En sistemes probabilístics, la pressió per corregir l'error singular visible tendeix a deteriorar el rendiment agregat invisible.

Traduït a la política lingüística: una sola família que denuncia davant les càmeres que «l'han obligada a estudiar en valencià» pot desencadenar una reforma legislativa que empitjore les condicions de supervivència del valencià per a cinc milions de persones. El cas singular visible, dramàtic, narrable, mediàtic, destrueix el sistema que, en agregat, protegia la llengua. I ningú no mesura l'agregat, perquè l'agregat no té rostre.

La intel·ligència artificial entra en aquesta història no com a salvadora ni com a amenaça, sinó com a reveladora. Un model de llenguatge és, en essència, una màquina de distribució de probabilitat lingüística. Calcula quina seqüència de paraules és més probable donat un context. No comprén. Generalitza correlacions. I les correlacions que generalitza reflecteixen el poder lingüístic del món real. En un LLM (Large Language Model, model de llenguatge a gran escala), l'anglès és la



llengua per defecte. El castellà hi és amb força. El català hi és com una ombra. L'èuscar és quasi invisible. L'aragonés no existeix.

Cada vegada que un parlant de català canvia al castellà perquè «en català no funciona tan bé», la distribució es reforça. La IA no discrimina per ideologia. Discrimina per freqüència. I la freqüència és una altra manera de dir poder.

He tingut la sort, o la rara mania, d'escriure en llengües que no són la meua. Una novel·la en gallec. Recorde la nit que vaig intentar escriure una escena d'amor en gallec i em vaig adonar que no tenia les paraules per a la tendresa quotidiana: les tenia per a la descripció, per a la narració, però no per a allò que es diu a cau d'orella. Les havia après dels llibres, no de la vida. I aquella mancança em va dir més sobre la diferència entre saber una llengua i viure-la que qualsevol tractat de sociolingüística. Un text en èuscar que, per sorpresa meua, va guanyar un premi, una comèdia musical sobre Sabino Arana que vaig escriure mig de broma i que el jurat es va prendre seriosament. Textos en castellà, en anglés, en català. Cada llengua en què he escrit m'ha mostrat un límit diferent: el límit del que pots pensar però no dir, del que pots sentir però no formular, del que existeix en una llengua i no existeix en cap altra.

El que he descobert, escrivint i treballant amb models de llenguatge, és que una llengua no és un monument. No és una bandera. No és un dret abstracte. És un sistema viu que opera sobre probabilitats: la probabilitat que un pare parle als seus fills en valencià, que una mestra ensenye matemàtiques en català, que un metge atenga un pacient en èuscar, que un model d'IA genere un text coherent en aragonés. Cadascuna d'aquestes probabilitats és petita. Insignificant, si es mira de prop. Però el conjunt de milions de probabilitats petites és el que determina si una llengua viu o mor. El dàlmata no va morir d'un dia per l'altre. Va morir decisió a decisió, generació a generació, parlant a parlant, fins que només en va quedar un: un home vell,



mig sord, que barrejava paraules venecianes amb records d'una llengua que ja no tenia amb qui compartir.

El català no està en la situació del dàlmata. Té deu milions de parlants potencials, universitats, premsa, una tradició literària que va de Ramon Llull a Mercè Rodoreda. Però la tendència agregada (la que ningú no mesura, la que no té rostre) és de retrocés. I ara, la intel·ligència artificial accelera la tendència, perquè el món probabilístic de les màquines replica i amplifica la jerarquia lingüística del món humà.

Aquest assaig no és un pamflet identitari. No defensaré la llengua amb arguments sentimentals. He après (dels models estadístics, no de la ideologia) que les dades són més resistents que les proclames. El que defensaré és que necessitem una nova manera de pensar la política lingüística als territoris de parla catalana: una manera que siga capaç de sostenir la coherència agregada (la supervivència de la llengua com a sistema) sense abandonar la protecció del cas singular, el dret de cada parlant a ser atés en la seua llengua.

Anomenaré aquesta proposta una ètica de l'agregat. No és una substitució de l'ètica del cas que ha definit la modernitat europea. És el seu complement necessari en un món on la incertesa ja no pot ser amagada darrere de fórmules vagues.

La intel·ligència artificial no ha vingut a matar les nostres llengües.

Ha vingut a posar-les sota sospita. I la sospita és aquesta: potser el perill no ve de les màquines, sinó de la nostra pròpia incapacitat per pensar en el que no es veu.

He tingut la sort improbable d'haver treballat professionalment amb quatre llengües minoritzades. L'èuscar, durant els deu anys d'Euskaltzaindia. El gallec, quan vaig escriure una novel·la en la llengua del meu avi matern, una llengua que no havia parlat mai a casa però que portava gravada en la memòria de la infància, en les converses dels estius a Galícia. El català, que he après de gran i que escric amb la mateixa mescla de devoció i inseguretats