

# HIZKUNTZA-TEKNOLOGIAK EUSKALDUNON ARTEAN

XABIER  
ARREGI IPARRAGIRRE

HITZ HIZKUNTZA TEKNOLOGIAKO EUSKAL  
ZENTROA - IXA TALDEA  
UPV/EHU EUSKAL HERRIKO UNIBERTSITATEA

## 1. ADIMEN ARTIFIZIALA ETA HIZKUNTZA-TEKNOLOGIAK

Puri-purian dago adimen artifiziala. Termino teknikoa zen duela gutxi arte; gaur egun, hitz oso zabaldua da. Adimen artifizialaren arloak unibertsitateetako eta ikerkuntza-zentro espezializatuetak mugak gainditu ditu, eta hainbat erabilera praktiko ditu gure eguneroko jardunean. Hedapen hori, itxura guztien arabera, are eta bizkorragoa izango da hemendik aurrera.

Adimen artifiziala, gogora dezagun, makinek giza adimena imitatzeke duten gaitasuna da. Hortaz, edonolako gailu informatiboak giza pentsaerari eta portaerari dagozkion prozesuetarako gaitzea du xede. Oinarritzko ideia baten eraldaketa dakar horrek: adimena ez da engoitik izaki bizidunen ezaugarri eskusiboa, ezpada, makinei ere antzeman dakiekeen gaitasuna.

Hain zuzen ere, ulermena eta mintzamenaren giza adimenaren adierazpen nabarmenenak baitira, esan beharra dago hizkuntzaren prozesamendua, makinek giza hizkuntza ulertzea eta erabiltzea helburu duena, muin-muineko adimen artifiziala dela.

Nabari iritziko dio zenbaitek baieztapen horri, baina ez dago soberan. Hizkuntzaren

prozesamenduaz ari garenean, edo hizkuntza-teknologiez, adimen artifizialaz ari gara. Eta adimen artifiziala aztertzean, ezin diogu begiratua ukatu hizkuntzaren prozesamenduari.

Bi arloen bilakaerak horixe erakusten du, 1940ko eta 50eko hamarkadetan lehen azterketa teorikoak egin zirenetik. Sinboloen eta arrazoitze logikoaren moldera jo zen gero, eta 1990eko hamarkadan estatistikan oinarritutako ikasketa automatikoa etorri zen. Azken paradigma-aldaketak –azken bospasei urteetan izugarri hedatu denak– datu masiboen erabilera eta ikasketa sakonean (*deep learning*) du oinarria. Eredu neuronal honek emaitza harrigarriak ekarri ditu, adibidez, itzulpen automatikoan. Jauzi kualitatiboa eragin dute emaitza horiek, eta zinezko erabilgarritasuna eman diote aplikazio eta tresna linguistiko askori.

Adimen artifizialaren aztertze-eremu askotan igarri da hobekuntza, besteak beste, ezagutzaren adierazpidean eta ustiapenean, robotikan, ikusmen artifizialean, edo hizkuntzaren prozesamenduan. Erabilgarritasun funtzionala hobetzearekin bat, aplikazioen eskala eta sozializazioa nabarmen hazi da, eta arlo askotara iritsi da. Erabilera militarra edo etekin handiko erabilera industrialia izan ohi da

hegemonikoa, aurrerakuntza zientifiko-teknikoa lehen-lehenik horietara lerratzen delako ia beti, bistakoa denez. Baina, esparru horietatik at ere bada adimen artifiziala, eta gaur egun oso presente dago sistema prediktiboetan edo erabakiak hartzeko laguntza-sistemetan, fabrikazio- eta ekoizpen-prozesuetan, ibilgailu autonomoetan, domotikan, medikuntzan, eta, oro har, ezagutzaren ustiapena eskatzen duten prozesuetan. Hedatua dago, orobat, hizkuntzaren azterketari lotutako zereginetan, hala nola, itzulpen automatikoan, pertsonen eta makinaren arteko elkarrizketa-sistemetan, irakaskuntzan eta humanitate digitaletan, bideo-jokoetan, era guztietako dokumentuen digitalizazioan eta analisisetan, eta abar.

Adimen artifiziala eta hizkuntzaren prozesamendua batera badoaz ere, kostatzen da hizkuntza-teknologiak garapen ekonomiko eta industrialaren terminoetan planteatzea. Badirudi industria eta hizkuntza bi errealitate disjuntu direla. Baina hizkuntzaren prozesamenduko ekoizpen zientifikoak adimen artifizialekoaren % 30-40 hartzen du, eta, hori horrela, hizkuntzaren prozesamenduaren irismenari, eragiteko ahalari eta alderdi sozio-ekonomikoari serio heldu beharko genioke. Inpaktu ekonomikoaz eta inpaktu sozialaz ari gara, gure bizimoduan eragiteko ahalmenaz.

## 2. EUSKARA TESTUINGURU HORRETAN

Informazio eta Komunikazio Teknologien (IKT) aro honetan, euskaraz bizi nahi duen komunitateak ez du bazterrean geratu nahi. Ahalegin horretan azpimarra daitezke, adibidez, puntueus (domeinuak.eus) edo euskal wikipedia (eu.wikipedia.org) bezalako ekimenak. Bien arteko desberdintasunak desberdintasun, ekimen horiek badituzte nabarmentzeko moduko ezaugarri komunak: ez dira puntualak, ibilbide luzekoak baizik; administrazioaren eta gizarte-eragileen arteko lan-kidetza dute oinarrian, eta auzolanean hazten ari dira, bokazio kolektiboa elikatuz norbanakoen inplikaziotik.

IKTetik hizkuntza-teknologietara etorrira, zalantza berehala sortzen da: Ez ote da areagotuko handienen nagusitasuna? Hizkuntza gutxi batzuek izan ezik, izango al dute besteek garapen teknologikorako aukerarik? Izatekotan, nola? Esan beharra dago euskara baino hizkuntza boteretsuagoak ere kezkatuta daudela jadanik mundu mailako hizkuntza hegemonikoen nagusitasun guztizkoaren aurrean. Arrakala digitala, minimizazio digitala, heriotza digitala... termino horiek erabiltzen dira hitzetik hortzera.

Gurera etorrira, euskarak egin al diezaioke aurre teknologiaren apustuari? Dagoeneko uste zabaldua da oso: edo euskarak lortzen du bere espazioa teknologiak markatutako mundu honetan, edo *modernitatetik* kanpo geratuko da. Eta modernitatetik kanpo geratzeak galera izugarria lekarkioke normalizazio-prozesuari.

Badugu zer egina, beraz; eta aspaldi hasi ginen zeregin horretan, izan ere, ibilbide luzea du euskararen prozesamendu automatikoak. 1990eko hamarkadaren hasierarako garatuak ziren oinarriko baliabide eta tresna linguistiko zenbait –xuxen, esate baterako–, eta, haiek oinarri, orduz geroztik etengabea izan da garapena, garaian garaiko teknologiak eta ereduak baliatuta. Ibilbide horretan, txikiak izanda ere, jakin izan da nazioarteko sistema eta eredu berritzaileenetan ikertzen, eta euskarari aplikatzen. Parte hartu da proiektuen deialdi lehiakorrenetan, sarritan euskara ez beste hizkuntzetarako izan badira ere. Ikasitakoa, beti, euskarari aplikatzeko xedez aritu izan gara, eta horrela lortu da nazioarteko hizkuntza-teknologien mapan euskara agertzea.

Gaur egun, baliabide eta hedadura urriko hizkuntza izanagatik, prozesamendu linguistiko automatikoaren esparruan, munduko hogeitaz hizkuntza nagusien artean kokatu du euskara Andras Kornaik<sup>1</sup>.

Nazioartean, gurearen antzeko egoeran dauden hizkuntzentzat aitzindaria eta erreferentea da euskararen garapen teknologikoa.

1. Kornai, A.. *Hizkuntzen heriotza digitala*. BAT Soziolinguistika aldizkaria, 11. 2019 (2). 11-32. ISSN: 2659-5176.

Azken zazpi-zortzi urtean izandako hobekuntza orokorrak, gainera, potentzialitate handia du, eta eragina eduki lezake euskararen erabileran. Unea da, beraz.

## 2.1 EGOERA: AUKERAK ETA ARRISKUAK

Abagunea da hizketaren tratamenduan, itzulpen automatikoan edo testuen analisietan lortutako emaitza onen joerari eusteko ikerkuntza mailan, eta ikasitakoa gizarte- eta lan-mundura transferitzeko. Xedea garbia da: euskararen prozesamendu automatikoa presentzia handikoa izatea aro digitalean, eta euskaldunon esku jartzea errendimendu oneko baliabide eta tresna teknologikoak.

Badira aldeko alderdiak euskarak abagune honi behar bezala hel diezaion:

- Europar hizkuntza-teknologiak inbertsio-ardatza izateko bidean dira.
- Euskarak kalitate handiko baliabideak eta tresnak ditu.
- Arloko teknikarien eta ikertzaileen prestakuntza ondo errotuta dago. Hizkuntzaren prozesamenduan diharduten irakasleen eta ikertzaileen jarduna aspaldikoa da, eta etengabea da gazteen prestakuntza –aipagarriak dira UPV/EHUko HiTZ zentroak eskaintzen dituen HAP/LAP (ixa.si.ehu.es/master) eta EMLCT europar masterrak (ixa2.si.ehu.eus/erasmusmundus)–. Horrela sortu da hizkuntzaren prozesamenduan dihardutenean komunitatea, handia eta kohesionatua dena.
- Badira, halaber, eskarmentu eta kalitate handiko enpresak edo agente teknologikoak.

Zailtasunak edo aurkako alderdiak ere ondo identifikatuta daude:

- Ikerkuntzatik industriarako transferentzia ez da erraza. Hizkuntza-teknologiaren ahala ez da, gainera, ondo ezagutzen eremu industrialean.

- Talentu-ihesaren arriskua handia da. Hemen bertan baldintza erakargarriak sortu ezean, zaila da kanpoko deiarri entzungor egitea, are gehiago arloko korporazio edo enpresa erraldoiak hain boteretsuak eta aberatsak izanda.
- Korporazio horiek merkatu-legearen arabera funtzionatzen dute. Jokoa arau horiekin euskarak eta hizkuntza gutxituek zail dute beren garapena.

Oro har, aldeko eta aurkako faktoreak baloratuta, esan dezakegu zimendu sendoak ditugula etorkizuneko lanari heltzeko.

## 3. ETORKIZUNERAKO PLANGINTZA

Esana dugu Europak ardatz gisa hartu duela hizkuntza-teknologiaren arloa etorkizuneko inbertsioetarako. Hala dago jasota Europako Parlamentuak 2018an onetsitako “*Language equality in the digital age*” txostenean<sup>2</sup>. Dokumentu horretan, aro digitalean eleaniztasuna bultzatzea eskatzen zaie estatu kideei, eta Europako hizkuntza guztiek eduki beharko lituzketen baliabide linguistiko gutxienekoak identifikatzea –datu-multzoak, lexikoak, ahots-grabazioak, itzulpen-memoriak, corpus anotatuak, eduki entziklopedikoak, eta abar–. Parlamentuaren ebazpen horretatik eratorri da “*Developing a strategic research, innovation and implementation agenda and a roadmap for achieving full digital language equality in Europe by 2030*” deialdia. “*European Language Equality (ELE)*”<sup>3</sup> izeneko proiektua aukeratu da agenda eta bide-orri hori diseinatzeko, eta proiektu hori garatzen ari den partzuergoko bost kideetako bat da UPV/EHUko HiTZ zentroa.

Europak estatu kideei eskatzen die engaia daitezen hizkuntzen garapen digitalean. Esate baterako, Espainiak badu bere “*Plan Nacional de Tecnologías del Lenguaje*”,

2. [https://www.europarl.europa.eu/thinktank/es/document.html?reference=EPRS\\_STU%282017%29598621](https://www.europarl.europa.eu/thinktank/es/document.html?reference=EPRS_STU%282017%29598621)

3. <https://ecspm.org/multilingualism-and-digital-technology/european-language-equality/>



lehenago ondutako “*Plan de Impulso de las Tecnologías del Lenguaje*” zeritzonetik eratorria<sup>4</sup>, eta, maila orokorragoan, badu “*Estrategia Nacional de Inteligencia Artificial*” izeneko plana ere.

Euskal Autonomia Erkidegoan, Eusko Jaurilaritzaren Hizkuntza Politikarako Sailburuordetzak idatzia du “Hizkuntza-teknologiak sustatzeko planaren txosten teknikoa”. Ikuskitzen dago zer bilakaera izango duen egitasmo horrek, baina premiazkoa da, ez bairik gabe, plan estrategikoa bat, helburuak eta lehentasunak ondo zehazteko, aktoreen rola ondo definitzeko, politika publikoen nolakotasunak argi izateko, eta modu koordinatuan jarduteko.

Kataluniako Generalitateko Politika Digitalen eta Administrazio Publikoaren sailak eta Bartzelonako Superkonputazio-Zentroak (BSC) ondu duten AINA proiektua<sup>5</sup> eredu egokia izan daiteke guretzat. Proiektu horretan, oso ondo zehaztu da jomuga: adimen artifizialaren laguntza, katalanaren biziraupena eta lehiakortasuna bermatzea mundu digitalean. Helburuak ez ezik, bitarteko materialak eta ekonomikoak finkatzen dira, eta katalanak beharrezko dituen baliabide teknologiko eta linguistikoak identifikatzen dira. Gaurkotasan handiko baliabideak eta tresnak aipatzen dira: corpusak, hizkuntza-ereduak, itzultzaileak eta ahots-ezagutzailak. Hasieratik esaten da, gainera, baliabide horiek komunitate zientifikoaren eta enpresarialaren esku utziko direla modu irekian, horien erabilera eta zabalkundea errazteko. Dagoeneko emaitzak agertzen hasi dira<sup>6</sup>.

### 3.1 AURRERA EGITEKO GAKOAK

Katalanari eskaini zaion moduko plan bat behar dugu guk ere lehenbailehen, euskararen biziraupen digitala bermatzeko. Estrategia adostua behar dugu, gure erakunde publikoetatik sustatua, irekitasuna zainduko duena, eta sektoreko agenteen lan koordinatua bultzatuko duena.

Esan dugu hiru zutabe dituela adimen artifizialaren eta hizkuntza-teknologiaren susperraldiak. Batetik, datu masiboen eskuragarritasuna; bigarrenik, datu horiek ustiatzeko ikasketa sakonaren edo sare neuronalen errendimendu aparta; hirugarrenik, konputazio-ahalmena, datu eta eredu horiek eskatzen dituzten exekuzio astunak eta garestiak ahalbidetzeko. Hiruretan dugu zer hobetua, baina bereziki eragin beharra dago konputazio-ahalmenean eta datuen eskuragarritasunean.

Gestio publikoko konputazio-zentro bat sortzea apustu estrategikoa da. Asko dira Euskal Herritik kanpo exekutitzen diren prozesuak, hemen behar bezalako azpiegiturarik ez dagoelako. Eta kanpoan exekutatzeko esan nahi du besteen esku uzten ditugula gure datuak, eta ez daukagula geure prozesuen gaineko kontrol erabatekorik. Administrazioaren eskutik sortu beharko litzateke konputazio-zentro publiko hori.


Bestetik, oso garrantzitsua da diru publikoak lagunduta sortutako baliabideak, ahotsezkoak nahiz idatziak, komunitate zientifikoaren eta –baldintza egokiak jarrita– bestelako agenteen esku uztea, modu gardenean eta mesede-traturik gabe. Datuek halako balioa dutenean, eskuragarritasuna eta irekitasuna ezinbestean zaindu behar dira sostengu publikoa izan duten corpus, baliabide eta tresnez ari garela. Partekatzeak kultura ez dago aski zabalduta gurean, baina aurre egin behar zaio joera itxiei, ezin baita onartu baliabide publikoen erabilera pribatiborik. Euskarari ezin zaio oztopo hori jarri.

Burujabetza teknologikoaz eta baliabideen gestio publikoaz ari gara. Hizkuntza-teknologiaren arloan, mundu mailako enpresa erraldoien indarra eta eragina hain handia izaki, pentsaezina da erabateko burujabetza, baina ezin dugu erabateko menpekotasunean erori. Zenbat eta baliabide propio gehiago ditugun, orduan eta errazago hartuko gaituzte konputazio enpresa handiek ere.

4. <https://www.plantl.gob.es/>

5. <https://politiquesdigitals.gencat.cat/ca/details/ActivitatAgenda/20201210-projecte-Aina>

6. <https://www.linkedin.com/content-guest/article/aina-neix-amb-for%C3%A7a-marta-villegas>



Azpiegitura propioak sortzearen ildoan, duela gutxi sortu da HiTZ: Hizkuntza Teknologia Euskal Zentroa (hitz.eus). UPV/EHU unibertsitate publikoko zentroa da HiTZ, Ixa eta Aholab ikerkuntza-taldeak biltzen dituena. Hizkuntza-teknologiaren erreferentziako zentroa izateko bokazioa du HiTZek. Arloko ikerketa, prestakuntza, transferentzia teknologikoa eta berrikuntza sustatzea du helburu, eta rol garrantzitsua joko dezake testuinguru honetan.

Gestio-eredu eta azpiegitura egokien bitartez, hizkuntza-teknologietan ari diren euskal ikertzaileek egitasmo motibagarriak dituzte. Labur-labur bada ere, hona hemen horietako batzuk:

- Elkarriketa-sistemak.  
Txatbotek eta laguntzaile adimendunek presentzia handia dute gure bizimoduan. Ez dute euskaraz egiten, ordea.
- Alborapenak edo erabilera desegokiak kontrolatzeko beharra.  
Ikuspegi etiko, sozial eta soziolinguistikotik garrantzitsua da tresna automatikoen eragina kontrolatzea. Paradigma neuronalak dagoenetik ikastea ekarri du. Hori sekulako ahalmena da. Baina, ikasi, onetik bezala ikasten da erdipurdikotik. Eta, zer ikasi, huraxe errepikatzen denez, tresna linguistiko neuronalek ohitura okerrak biderkatzeko ahalmena ere badute. Hori zaindu beharreko kontua da, adibidez, hizkuntzaren normalizazio linguistikoari begira, edo bestelako gako etiko edo sozialei begira. Genero desoreka da maiz aipatzen den gaia, eta gaur egun oso presente dago gai hau ikertzaileen artean.
- Hizkuntzaren aldaera, modu eta erregistro desberdinei heltzeko aukera.  
Oraintsu arte urrun ikusten bazen ere, eredu neuronalak askotariko hizkuntza-erregistro eta aldaerak lantzeko aukera ere ekarri du. Adibidez, landu daitezke zeinu-hizkuntza, edo aniztasun funtzionalari begirako aukera desberdinak. Eta landu daitezke modalitateak, testua, ahotsa, irudia eta abarrak konbinatuta.

- Aldiberekotasuna.

Aplikazioen erabilgarritasunari begira, aldiberekotasuna oso garrantzitsua da. Ondo ez ezik, azkar, eta ahal bada,aldi berean jokatzea eskatzen zaie aplikazioei. Ez da samurra erantzun-denbora hain azkarrak lortzea, baina lortuz gero, aukera funtzional interesgarriak zabaltzen dira.

- Erabilera –alderdi soziolinguistikoa–.  
Teknologia egokia izatearekin bat, asmatu behar da testuinguru errealean modu egokian integratzen. Horretarako, kontuan hartu behar dira erabiltzaileen inguruneak eta lan-baldintzak, eta horiei begira diseinatu aplikazio linguistikoak. Ez da ahaztu behar alderdi soziopolitikoak ere. Hizkuntza gutxitua izanda, kontua da ea nola baliatu daitezkeen hizkuntza-teknologiak euskararen estatusa indartzeko. Adibidez, oso nabaria da izaera elebiduna (edo eleaniztuna) duten aplikazioetan euskara lehenetsi beharko litzatekeela, alegia, euskarak sortze-hizkuntzaren rola jokatu beharko lukeela.

#### 4. ONDORIOAK

Artikulu honetan hizkuntza-teknologiak gaur egun duten garrantzia azpimarratu da. Euskarak mundu digitalean iraungo badu, guretik bertatik pentsatu, sortu eta garatu beharreko plan estrategiko baten premian jarri dugu begia, eta azaldu da zergatik euskal administrazioek euskararen garapen teknologikoa arlo estrategikotzat hartu beharko luketen.

Abagune egokian gaude, ibilbide sendoa egin dugulako orain arte, eta oinarri ona dugulako aurrera egiteko ere. Testuinguru orokorrak, Europatik datorren joera aintzat hartuta, unea dela erakusten du. Kataluniak, adibidez, erabakitasun handiz ekin dio ibilbide horri.

Etorkizunerako bideak kudeaketa publiko, ireki eta gardena eskatzen duela defendatu da, sektoreko agenteek, nork bereari eutsiz, plan horretan zer rol jokatuko duen garbi zedarrituta.

