

Euskara (batua) ingurune digitalean: bidean *ikasiXa* eta etorkizuneko erronkak

I. Aduriz*, I. Alegria, I. Aldezabal, X. Artola,
A. Díaz de Ilarraza, N. Ezeiza, K. Sarasola, R. Urizar

Ixa Taldea – UPV/EHU (*Bartzelonako Unibertsitatea)

1 Sarrera gisa

Euskara batuaren etorkizuneko erronkak ingurune digitalean: horra Euskaltzaindiak Ixa Taldekooi¹ paratu zigun gaia, Arantzazun 2018ko urrian egitekoa zen biltzarrera gonbidatu gintuelarik. Bada aurkezpen hartan jorratutakoak dira, hain zuzen ere, artikulu honetara ekarri ditugunak.

Gure ustez, batuaren etorkizuna euskararen etorkizuna da, ez dago bien artean bereizterik; horregatik ipini dugu *batua* parentesi artean, artikulua izenburuan. Izenburuaren azpikoak, berriz, argi uzten du nondik ekin diogun, etorkizuneko erronken inguruan hausnartzen hasi garenean: gure eskarmentutik, Ixa Taldearen 30 urte luzeko ibilbideak emandako eskarmentutik jarri gara aurrera begira.

Teknologia, oro har, baina –artikulua gaiari helduz– hizkuntza-teknologia bereziki, gure bizitzetan sartu da oharkabean: hor ditugu ordenagailuak, Internet, sare sozialak, sakelako telefonoak...; hor ditugu, beste maila batean, Xuxen, EuskalBar, AhoTTS, Lexikoaren Behatokia, Elebila, Matxin, Modela bezalako tresnak eta baliabideak... Ordenagailuan ari garela, amaraunean gabiltzanean, telefono mugikor bat eskuetan dugunean, hizkuntzak eta teknologiak bat egiten duten bidegurutze horretan gabiltza askotan: aukera dugu aginduak ahoz emateko; baliabideak ditugu testuak automatikoki itzultzeko; ordenagailu-programa batek idazten ari garena zuzentzen digu unean-unean; eskura ditugu hiztegiak mahaitik altxatu ere gabe... Horietan guztietan hizkuntza-teknologiak baliatzen ari gara, gutxi edo asko.

Xuxen aipatu dugu, ezinbestekoa bihurtu den lantresna, ordenagailu bidez euskaraz idazten duenarentzat. Eta, euskara batua dugunez hizpide, zilegi bekigu esatea jakin nahiko genukeela zer-nolako eta zenbaterainoko eragina izan duen Xuxenek, azken mende laurdenean, Euskaltzaindiaren arauak eta lexiko estandarra gizarteratzerakoan, esate baterako; ez da erraza hori jakitea, ez, baina ikergai polita litzateke. Eta beste tresna edo aplikazio erabilgarri batzuk ere aipatu nahiko genituzke: zenbat hauts biltzen ari ote dira paperezko hiztegiak azkenaldian, EuskalBar hor dugula, hain eskura gure ordenagailuan; edo Elhuyarren hiztegiak gure mugikorretan! Corpusak direla-eta, berriz, esan beharrik ez dago Lexikoaren Behatokia edo antzekoak tresna paregabeak direla hizkuntzaren profesionalentzat, hitzen baten edo egitura gramatikalen baten erabileraren berri jakin nahi dutenean.

Kontua da teknologia hau oso-oso arin ari dela garatzen, eta, gainera, esan daiteke heldu samarra dagoela jada hainbat arlotan. Oinarriak (baliabideak, formalizazioa, tresnak) finkatuta daude, eta aplikazio batzuk arrakastatsuak izan dira zenbait domeinu eta testuingurutan. Gutxi batzuk aipatzearen: medikuntza eta administrazioa, esaterako; telefonia mugikorra, zeina hautagai berebizikoa den teknologia hauen ezarpenerako; eta itzulpen automatikoa, non emaitza nabarmen hobeak lortzen ari baitira azkenaldian.

Artikulua egitura honako hau da. Hasteko, sarrera gisako honen ondoren, hizkuntzaren biziraupena jokoan ikusten dugula-eta hausnarketa labur bat egingo dugu, 2. atalean; gero, hizkuntza-teknologiaren gaia zertan den azalduko dugu (3. atala), euskararen tratamendu automatikoaren arloan non gauden ikusiz (4), eta ikerkuntza

¹ <https://ixa.eus>

(5) eta formazioa (6) bereziki aipatuz. Zazpigarren atalean, berriz, Euskaltzaindia, ikertzaileok, industria eta erakundeak, euskararen teknologiei dagokienez, zertan ari diren –eta garen– azalduko dugu. Zortzigarrenean, aurrera begirako ohar batzuk egingo ditugu, honako atal hauetan gure ustez nondik jo beharko litzatekeen iradokiz: corpusgintza, arauak eta estandarra, hiztegi-gintza eta terminologia, itzulpen automatikoa eta itzulpen-erako laguntzak, eta humanitate digitalak. Eta bukatzeko (9), ondorioak zerrendatu eta proposamen gutxi batzuk egiten ausartuko gara.

2 Hizkuntzaren biziraupena jokoan

Zer gertatzen da gurea bezalako hizkuntza “txikiek” mundu digitalean? Ba hortxe dugu islandiera, esate baterako: atzean estatu baten indarra duen hizkuntza izanagatik, itoan dagoela-eta kezka zabaldu da, ingelesaren ozeanoko ur handietan itolarri hilgarrian ikusten omen dute². Hango gazteak, teknologiari dagozkion kontuetan, ingelesera jotzen du, antza; arrazoi asko egongo dira horretarako, baina, dudarik ez, askotan beste aukerarik ez duelako izango da, hots, bere garaian islandieraz jardun ahal izateko baliabideak eta tresnak garatu ez direlako.

Hizkuntza biziko bada, teknologiarekin batera bizi beharko du. Esan dezakegu ideia hori dagoela Ixa Taldearen jatorrian, eta azken hiru hamarkadetan teknologia horren azpiegitura euskararako prestatzen aritu gara –datu-base lexikala sortu, hizkuntzaren formalizazioa egin, testu-corpusak eratu, analizatzaile morfologikoa eta sintaktikoa garatu, semantika lexikala landu...–, gure partez kanpoko beste inork ez zuela egingo jakitun.

Eta gure ustez garbi zegoen talde gisa jaio ginenetik, eta inoiz baino argiago dugu orain: hizkuntzaren biziraupenaz ari gara, hori da jokoan dagoena. Tren hori, eta beste batzuk, noski, baina abiada bizian doan tren hori pasatzen utziz gero, nekez egingo dugu aurrera: erronka da, eta geltokian geratzeko arriskua handi samarra dela esango genuke.

3 Hizkuntzaren tratamendu automatikoa

Hizkuntzaren tratamendu automatikoak hizkuntza bitarteko informatikoz prozesatzea du helburu. Eta hizkuntza diogunean ahazkoaz zein idatzizkoaz ari gara, baina baita hizkuntza formalaz zein informalaz, estandarraz nahiz ez-estandarraz.

Ixa Taldean ere, orain arte, hizkuntza idatzia prozesatzea izan dugu jomuga, eta idatziaren barruan estandarra batik bat. Baina ahazkoa ere –hizketa– hor dago, eta berorren tratamendua gero eta garrantzi handiagoa hartzen ari da. Izan ere, hizkuntza, lehenik eta behin, ahazkoa da, eta teknologiak ahazkoa ere bere baitan hartzen du. Beraz, horretarako ere prestatuta egon behar dugu, zalantzarik gabe. Testuinguru honetan, aipa dezagun UPV/EHUko AhoLab laborategia³, urteak daramatzana arlo horretan lanean.

Idatzizkora itzuliz, formala eta informala aipatu ditugu, estandarra eta ez-estandarra. Hizkuntza informal ez-estandarra erruz erabiltzen da gaur egun Whatsapp-az ari garenean, esaterako, edo sare sozialetan, foroetan eta abar. Esan daiteke badela hor beste erregistro bat, idatziaren eta ahazkoaren artekoa dena, nolabait, eta munta handia hartzen ari dena, erabilerari dagokionez bederen; gazteen artean-eta, esate baterako. Bistan da hizkuntza hori ere prozesatzeko gai izan beharko dugula aurrerantzean.

² <https://labur.eus/ITXu>

³ <https://labur.eus/Aams2>

4 Euskara aro digitalean: META-NET txostena

Zein da gure hizkuntzaren egoera aro digitalean? Galdera horri erantzun bat eman nahi izan zion duela urte batzuk argitaratu zen *Euskara Aro Digitalean – The Basque Language in the Digital Age* (Hernández *et al.*, 2012) txostenak. META-NET Europako informazio-gizarte eleaniztunaren testuinguruan erreferentziazkoa den bikaintasun-sarea da, eta 30 liburu zuri argitaratu zituen, beste horrenbeste hizkuntzaren egoera aztertuz. Hizkuntzek agertoki digitalean duten presentzia, eta etorkizunari begira dituzten aukera-arriskuak aztertu ziren liburu horietan. Euskara zen hizkuntza horietako bat⁴.

Liburu zuri horietan jaso ziren emaitzak honela laburbil daitezke:

- Alde batetik hizkuntza bat dago, ingelesa, eta bestetik aztertu ziren beste hizkuntza guztiak. Inondik ere, ingelesak agertoki digitalean oraintxe duen egoera nabarmena da, eta beste ezein hizkuntzarena ezin da harenarekin konparatu.
- Aztertutako 30 hizkuntzetatik, 21ek duten babes digitala “hutsa” edo, kasurik onenean, “eskasa” da.
- Aipatu ez diren hizkuntzak (8) alemana, frantsesa, espainola eta abar dira, tarteko egoera batean dauden hizkuntza *indartsuak*, alegia.
- Euskara babes digital eskasa duten 21 hizkuntza horien artean sailkatuta geratu zen, azkenekoetakoa ez bazen ere. Hona hemen, atalez atal, euskararen egoera nola kalifikatu zuten txostena idatzi zuten adituek:
 - Hizketa-prozesamendua (ahozkoa): osatugabea.
 - Itzulpen automatikoa: apala / ez⁵.
 - Testu-analisia: osatugabea.
 - Baliabideak: osatugabea.

Txostenean, euskararen egoera “baikor”tzat jo zen, baina “zuhur aztertu beharrekotzat”.

Honen aurrean, zer esan? Ez ginen *azken putz* geratu, egia da, 21 horien artean azaldu ziren gurea baino egoera okerragoan zeuden hizkuntzak. Edonola ere, ez dugu uste orain, txosten hartatik sei urte baino gehiago joan direnean, besteekiko egoera hobean gaudenik: ingelesak, jakina, baina baita beste hizkuntza handi horiek ere, aurrera egin dute, eta iruditzen zaigu arrakala handiagotuz doala. Hizkuntza batek agertoki digitalean presente egon nahi badu, hizkuntza batek etorkizun digitalean egunero eta normal erabilia izan nahi badu, beste hizkuntza handi horien pare egon behar du, baliabide, tresna eta teknologia hauei, oro har, dagokienez; bestela, eta hau da gure iritzia, jai du.

5 Ikerkuntza: sustatu, lagundu eta bultzatu

Egoera den bezalakoa izanik, uste dugu lanean jarraitzea beste biderik ez dagoela, eta lan hori, gure hizkuntzari dagokionez behintzat, geuk –euskaldunok– egin beharko dugula.

Ikerkuntzaren garrantzia itzela da. Diziplinarteko taldeak behar dira, non informatikariek eta ingeniariak filologo eta hizkuntzalariekin batera jardungo duten. Oinarrizko ikerkuntza behar-beharrezkoa da: hizkuntzaren formalizazioa hobetu egin behar da, analisi- eta sintesi-tresnak garatu behar dira, eta aplikazioek ezinbesteko dituzten baliabideak –lexikoak, termino-bankuak, corpusak, gramatikak eta abar– sortu eta elikatu behar dira.

⁴ Euskarari buruzko alearen idazketan taldeko bi kidek hartu zuten parte (ikus erreferentzia).

⁵ Gaur egun “osatugabetzat” jo genezake.

Halaber, azken erabiltzailea helburu izango duten aplikazioak ere garatu behar ditugu, hala nola hizkuntzaren *input*-ean zein *output*-ean laguntza emango diguten aplikazioak, itzulpenerako laguntzak, hizkuntza idatzian nahiz ahokoan arituko diren era guztietako aplikazioak. Hona hemen adibide batzuk: laguntzaile birtual adimendunak⁶, azpigituluaren sorkuntza automatikoa, EITBren *Nahieran* zerbitzuko bideo batean hitz bat edo pasarte jakin bat bilatzeko laguntza, sare sozialetan emandako iritziak polaritatearen arabera arakatzeko aukerak...

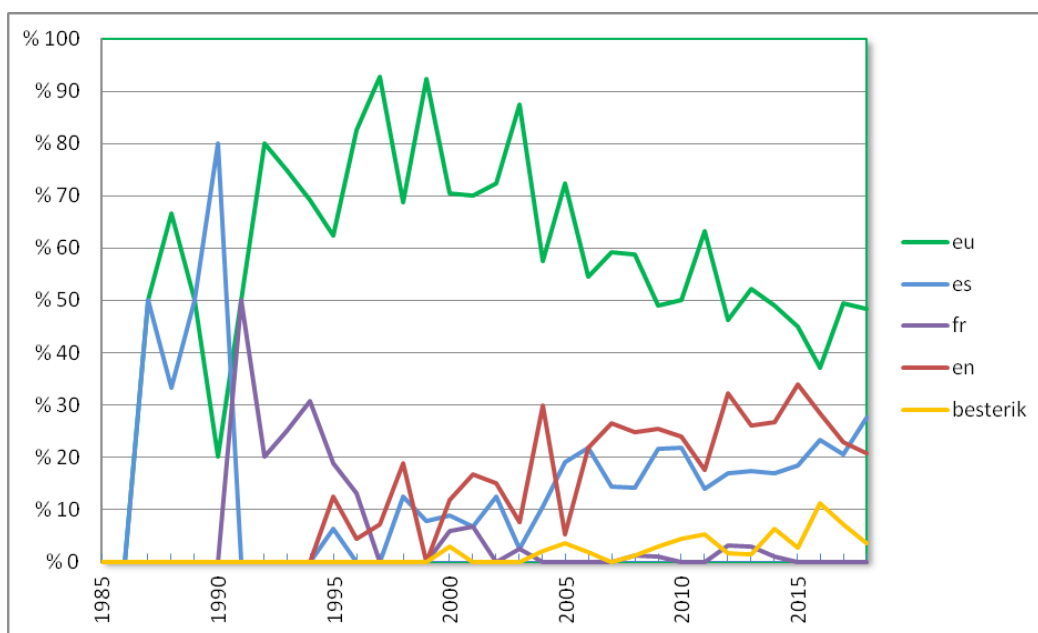
5.1 Zenbait arazo

Ikerkuntzaz ari garela, arazo batzuk aipatu nahi genituzke, aintzat hartu beharrekoak iruditzen baitzaizkigu.

Batetik, kontuan izan behar da hizkuntza "txikiok" baliabide gutxiago dituztela, oro har, eta horrek ikerkuntza zaildu egiten duela. Gaur egun, metodo estatistiko eta neuronalek garrantzi handia hartu dute gure ikerrarloan, eta metodologia horiek aplikatu ahal izateko oso corpus handiak, testu-masa erraldoiak behar dira. Ez da inor harrituko esaten badugu ingeleserako hainbat eta hainbat baliabide ditugula eskura, doan eta publiko, eta esperimenduak egitea berehalakoa dela; esperimendu horiek berak euskararen gainean egin nahi izanez gero, ordea, aurrena baliabideak bildu eta prestatu behar dira, eta hori lan nekeza eta garestia da⁷.

Esan daiteke neurri batean teknologia komuna dela hizkuntza guztientzat, baina hizkuntza bakoitzak bere ezaugarriak ditu, eta formalizazio eta baliabide bereziak behar ditu. Eta hor dago kokka!, hortxe topatzen dugu (lehen) traba handi bat, aurrerapen teknologiko gehienak ingelesa helburu egiten baitira.

Arloko ikertzaileok ondo dakigu gurea ez beste hizkuntzen gainean egindako ikerketa-ahaleginak ez direla alferrik. Batetik, esan behar da arlo honetako ikerkuntzak orokorretik ere baduela, eta hizkuntzaren prozesamenduan oro har egindako aurrerapenak hizkuntza guztientzat direla baliagarri; bestetik, argi dago jarduteak beti balio duela ikasteko, eta beste hizkuntza baten gainean ikertuz ikasitakoa balia daitekeela gero euskararen aurrerabidean lagunduko duten beste ikerlan batzuk egiteko. Baina azken horren ifrentzia ere bistatik kendu gabe, jakina: beste hizkuntza batzuetan egindako inbertsioak gurean ez egindako inbertsioak dira, eta inbertsio horien etekin gehienak beste hizkuntza horrek jasotzen ditu.



6 Apple-ren *Siri* edo Microsoft-en *Cortana*ren gisakoak.

7 Hasteko, esan behar da batzuetan baliabide horiek edo ez direla existitzen edo ez daudela publikoki eskuragarri.

1. irudia: Ixa Taldearen argitalpenetan (1.071 artikulua, 1987-2017) ikergai diren
hizkuntzak: ehunekoak, urtez urte.

Bestalde, esan beharra dago unibertsitate “ofizialak” –eta gu hor gaude–, ez duela errazten euskararen gaineko ikerketa. Hor ditugu unibertsitatean indarrean dagoen meritu-sistema, argitalpenen “kalitatea” neurtzeko parametroak eta abar. Hizkuntza “exotikooi” buruzko ikerketa mundu akademikoan “saltzea” zailagoa da, hau da, arlo honetan argitaratzea, proiektuak eskatzea eta abar betiere zailagoa da, jokoan daudenak oinarrizko ikerkuntza eta hizkuntza “txiki” bat direnean.

Ixa Taldean egindako argitalpenetan argi ikusten da joera bat, eta horixe erabiliko dugu esan berri duguna ilustratzeko. Goian dagoen 1. irudiko grafikoak 1987tik 2017ra bitarteko gure argitalpenetan zein hizkuntza izan ditugun ikergai islatzen du. Irudi hori egiteko, hiru hamarkada horietako milatik gora argitalpen aztertu ditugu, eta uste dugu hor argi geratzen dela gure artean orain arte susmo sendoa baino ez zena: euskara ikergai duten argitalpenen kopurua, proportzioan, behera doa (erreparatu marra berdeari grafikoan), eta beste hizkuntza batzuen gainean ikertzen duguna, berriz, gero eta gehiago da (marra gorria, adibidez, ingelesa da). Joera horren azpian daudekeen arrazoiak asko izan daitezke, dudarik ez, baina seguru gaude aurreko paragrafoan aipatu ditugun pisua oso handia dela⁸.

Honen guztiaren aurrean, zer egin daiteke aipatu arazo horiei aurre egiteko, zer egin daiteke Euskal Herrian hizkuntzaren prozesamenduaren arloan egiten den ikerketak nahi dugun emaitza izan dezan, hau da, euskara prestatu –gaitu– dezan, ingurune digitalean dagoeneko dituen eta etorkizunean izango dituen erronkei erantzuteko? Galdera honi denon artean erantzun beharko diogu, baina guk ezinbestekoak iruditzen zaizkigun hiru kontu iradoki nahiko genituzke bederen: 1) nekez egingo dugu aurrera herri-administrazioen esku-hartzerik gabe; 2) denon artean sortzen ditugun baliabideak eta tresnak denon eskura jarri behar ditugu, publikoak izan behar dute: gutxi gara, eta egindakoa nork beretzat gordetzen duen bitartean alferrik xahutzen ariko gara hizkuntzaren onerako izan litekeen altxorra; eta azkenik 3), eta hau esatea boladan ez badago ere, uste dugu lan militantea ezinbestekoa izango dela luzaroan ere.

6 Formazioa: graduak, masterrak, doktoregoa

Ikerkuntzaren oinarria formazioa da, eta arlo honek formazio espezializatua eta diziplinartekoa eskatzen du.

Hizkuntzalaritza edo filologia irakasten den lekuetan, informatikak berezkoa duen buru-antolakuntza ez da batere lantzen; eta informatika-ikasketetan, berriz, ez zaio hizkuntzari duen garrantzia ematen. Baina diziplina horien elkarrekin beharrezkoa da hizkuntzaren teknologiotan aurrera egingo bada. Diziplinartekotasunak, bada, behar duen lekua hartu behar du hizkuntza- eta informatika-gaiak jorratzen diren fakultate, zentro, gradu, graduondo eta abarretan –Informatika Ingeniaritza, Euskal Hizkuntzalaritza eta Filologia, Itzulpenlarietza eta Interpretaziorik, Hizkuntzalaritza, Telekomunikazioa...

Gure aldetik honetan aritu gara urte hauetan unibertsitatean, eta aipagarriak iruditzen zaizkigu UPV/EHUn gauzatutako doktorego-programak eta masterrak: 56 doktorego-tesi aurkeztu dira arlo honetan, eta 126 titulatu atera dira⁹ graduondoko titulazio eta masterretatik. Hala ere, uste dugu hori dena ez dela aski beharrak betetzeko, gehiago egin behar genukeela; ez da aski, eta hala ere, master horietara datozen ikasle euskaldunen kopurua ez da handitzen ari, joera behintzat ez da goranzkoa.

8 Itzulpen automatikoaren gaineko ikerketak pisu handia hartu du taldean azken urteetan, eta, noski, gai horri buruzko argitalpenetan ikergai diren hizkuntzak bi edo gehiago izan ohi dira. Horrek azal dezake, neurri batean behintzat, euskararen gaineko ikerketa proportzionalki pisua galduz joatea azken urteetan; hala ere, ziur gaude ez dela hori beheranzko joera nabarmen horren arrazoi nagusia.

9 Datuok 2017koak dira.

7 Ekimenak

7.1 Zertan ari gara?

Horren guztiaren aurrean, zertan ari gara, zertan ari dira herri honetako agenteak euskararen teknologiei dagokienez?

Euskaltzaindian, urteak dira hizkuntzaren teknologiarekin zerikusia duten lanak egiten direla: hor dugu azken hamarkadan sortu eta elikatzen ari den Lexikoaren Behatokia, 60 milioi hitzeko corpus etiketatua, zeinetan Elhuyarrek, UZEIk eta Ixa Taldeak diharduten elkarlanean; edo, oraintxe bertan egiten ari den beste proiektu bat aipatzearren, hor dugu *Orotariko Euskal Hiztegia* (OEH) corpus kontsultagarri bihurtzeko proiektua, Euskaltzaindiaren zuzendaritzapean eta UZEIren eta Ixa Taldearen parte-hartzearekin garatzen ari dena; hiztegiaren egitura konplexua esplizitu egiteko helburua duen proiektu horrek ahalegin handia eskatu du, eta agerian utzi du teknologia berriak hiztegi gintzan barneratzearen premia.

Esan behar da, halere, zenbait hiztegi –Euskaltzaindiarena, kasu, lehen *Hiztegi Batua* (HB) zena– gaur egungo metodologiekin lantzen ari direla ikuspegi teknikitik; OEHk ere izango du, epe labur edo ertainean, eta arestian aipatu dugun proiektuaren emaitzei esker, azpiegitura bat, aurrerantzean edizio-lana modu egokian egiteko aukerak eskainiko dizkiona.

Oro ez da urre, baina. Euskaltzaindiak zenbait urtetan indarrean izan duen e-AEB zerbitzua (Arauen Banaketa Elektronikoa) etenda dagoela dirudi, eta ez da albiste ona. Zerbitzu horren bidez, besteak beste hiztegia eta arauak banatzen ziren elektronikoki, informazio hori automatikoki prozesatzeko erraztasunak eskainiz. Esaterako, *Hiztegi Batua* eguneratzen zelarik, PDF bertsioaz gain XML fitxategi bat ere argitaratzen zuen Euskaltzaindiak (webgunearen bitartez); esan beharrik ez dago, hiztegia eguneratuta mantentzeko lanak XML bertsio hori automatikoki ustiatuz egitea da egokiena eta normalena¹⁰. Gaur egun, berriz, atzerapauso bat eman da, eta PDF bertsioa baino ez da eskaintzen: PDF formatua ez da egokia automatikoki prozesatzeko, eta eskuzko lana egitera behartzen gaitu, denbora eta ahaleginak xahutuz.-

Ixa Taldean, berriz, eta AhoLabekin lankidetzan, HiTZ izeneko Hizkuntza Teknologiko Zentroa sortu berri dugu unibertsitate barruan, 2018ko udazkenean. Zentroaren helburuak honako hauek dira: arloko ikerkuntza eta formazioa sustatu eta bideratzea, eta teknologia-berrikuntzan eta -transferentzian eragitea. Urrats honekin, gure ikerketa-taldeei bestelako gorpuzkera bat eman nahi genieke, zentro bati beste aukera batzuk zabaltzen baitzaizkio proiektuetan-eta parte hartzeko entitate gisa, adibidez.

Gure artean badira arloan lanean ari diren beste talde, enpresa, erakunde eta zentro teknologiko batzuk ere. Besteak beste, hor dira Elhuyar, UZEI, Ametzagaiña Taldea, Vicomtech, Tecnalia, Mondragon Lingua eta beste batzuk, zeinek hizkuntzaren tratamendurako beren atalak dituzten; horietako askorekin lankidetzan aritu izan gara edo ari gara egun zenbait proiektutan.

Industria, berriz, zertan ari da? Hemen Langune¹¹ aipatu behar da, Hizkuntzen Industriaren alorreko Euskal Herriko enpresen elkarte, 2010ean sortua eta egun 35 enpresatik gora biltzen dituen. Azkenaldian, bere kasa aurrera egiteko zailtasunak direla-eta, EIKENen¹², Euskal Ikus-Entzunezkoen Industrien Klusterrean, sartu behar izan dute. Baina, ekimen horietaz gain, benetako kezkarik ba ote da hemengo enpresa askotan euskararen alde egiteko, beren aplikazio eta zerbitzuetan euskara beste hizkuntzen pare jartzeko? Noiz

¹⁰ Gure taldean, adibidez, inplementatuta ditugu zenbait prozedura automatiko, zeinak egikaritzuz Euskararen Datu-Base Lexikala (EDBL) eguneratzen baitugu (*baikenuen*, esan beharko dugu, tamalez), Euskaltzaindiak HBren bertsio eguneratua argitaratzen zuen aldiro. EDBL da Xuxenen hiztegia mantentzeko erabiltzen dugun tresna, baina ez hori bakarrik: EDBL da Ixa Taldeak euskara automatikoki prozesatzeko egiten dituen lan guztien oinarri eta euskarri lexikala.

¹¹ <https://labur.eus/IVWGY>

¹² <https://labur.eus/vrilx>

ohartuko dira, esaterako, merkataritza elektronikoaren bidez onura handiak lor daitezkeela bezeroaren hizkuntzan jardunez gero? Hor uzten ditugu galdera horiek.

Edonola ere, eta bukatzeko, esan behar da Euskal Herrian hizkuntzaren teknologien arloan ari garenok oso komunitate aktiboa osatzen dugula –hizkuntzaren *tamaina* kontuan hartuta, noski–, eta komunitate hori erreferentziazkoa dela, gainera, beste hizkuntza minorizatu batzuetako ikertzaileentzat (Aduriz *et al.*, 2017; Agerri *et al.*, 2017; Alegria & Sarasola, 2017).

7.2 Zertan ari dira erakundeak?

Argi esango dugu: Europan eta Espainian, behintzat, mugitzen ari dira. Euskal Herri mailan ez dugu, erakundeek dagokienez, halako sentsibilitaterik sumatzen gai honen inguruan, eta larria iruditzen zaigu.

Europako Parlamentuak gehiengo handiz onartu berri du txosten garrantzitsu bat, *Language equality in the digital age* izenburukoa (EPRS, 2017). Txostena onartzeko ebazpenean¹³, Parlamentuak Batzordeari eta estatu kideei eskatzen die merkatu digitalean eleaniztasuna erraztuko duten estrategia eta politikak garatzeko; defini dezatela Europako hizkuntza guztiek, heriotza digitalari ihes egiteko, izan beharko lituzketen gutxienezko baliabideak zein diren (datu multzo, lexikoi, grabazio, itzulpen-memoria, corpus anotatu eta abar). Azpimarratzen du, halaber, Europar Batasunak, estatu kideek, unibertsitateek eta bestelako erakundeek hizkuntza guztiak, baita gutxien mintzatzen direnak ere, ingurune digitalean babesteko eta zaintzeko duten betebeharra, guztientzako datu-base eta itzulpen-teknologiak garatu behar direla esanez. Bestalde, dei egiten die ikertzaileei eta industriari helburu komun bati begira koordina daitezen, hizkuntzen arteko itzulpenarako aukera digitalak emendatuz eta aurrerapen teknologikorako behar diren datuen atzibide libreak bermatuz.

Txostenak txosten eta ebazpenak ebazpen, ikusteko dago politikak gauzatzeko garaian zer gertatuko den; baina ezin da ukatu pauso garrantzitsu bat dela egoeraz jabetzea bera, eta neurri zehatzak proposatu eta hartzera bultzatzeko erabakia.

Espainiako Gobernuak, berriz, *Plan de Impulso de las Tecnologías del Lenguaje* delakoa du martxan (Ministerio de Turismo, Energía y Agenda Digital, 2015): 90 milioi euro, 2015-20 epealdirako. Plan horren helburua hizkuntzaren prozesamendua garatzea eta sustatzea da, eta aurreikusten du Estatuko hizkuntza koofizialetarako lan egitea ere (itzulpen automatikoa eta elkarrizketa-sistemak barne).

Bukatzeko, eta oso labur, Erkidegoko hiru erakunde aipatu nahi genituzke.

Hasteko, Eusko Legebiltzarra. 2018ko apirilaren 16an, taldeko bi kidek, Ixaren izenean, agerraldia egin genuen, Kultura, Euskara eta Kirol Batzordearen aurrean¹⁴, eta gaur hemen deskribatzen ari garen egoera eta kezka azaltzen saiatu ginen: hitz onak jaso genituen, baina oraingoz ez dute fruiturik eman; hala ere, eta deia Legebiltzarretik etorri zenez, parlamentari batzuegan dagoen kezka baten isla bada behintzat, eta ongi etorria izan bedi.

Eusko Jaurlaritza dela-eta, berriz, esan behar da badituela zenbait laguntza-programa (Etorrek, Elkartek...), eta 2020ko Agenda Digitalak ere zabaltzen dituela ate batzuk arloko ikerketa bideratzeko. Hala ere, euskararekin zerikusia duten kontu guztiak bezala, hizkuntzaren teknologien arlokoak ere gurean kulturaren ildotik bideratzen dira ia beti, ez industriarenetik. Eta hori ez da egokia, gure ustez: hizkuntzaren industriaz ari gara, eta ingurura erreparatu besterik ez dago ikusteko auzoan zertan ari diren –aipatu berri dugun *plan de impulso* hori ere, esaterako, Espainiako Industria Ministerioaren ardura da–, bistan baita aurrera egiteko jarrerak eta neurriek bestelakoak izan behar dutela.

Azkenik, Euskararen Aholku Batzordearen txosten bat (EAB, 2016) ekarri nahi genuke hona, zeinaren bidez Batzordeak egoeraren azterketa egin eta gomendio batzuk eman baitzituen. Lehentasun handienekoen artean,

13 Estrasburgo, 2018-09-11: <https://labur.eus/jiz1uf>

14 <https://labur.eus/UYbo4>

hauxe zen 12.a: “Hizkuntza-teknologiak garatzen lagundu eta berariazko plana garatu”. Bi urte joan dira harrezkero, eta egin berri duten diagnostikoak¹⁵ agerian dagoena egiaztatu baino ez du egin ahal izango, hots, gomendioak betetzeke jarraitzen duela, “berariazko planik” behintzat ez baita garatu.

Uste dugu puntu honetan gaudela, hain zuzen. Teknologia horien garapena bizkortu egin behar da, eta premia handikoa da administrazioak lehenbailehen ohartaraztea egoeraren larritasunaz. Eta ez da aski, jakina, gauzak egin behar direla esatearekin: duela mende erdi batuarekin egin zuten bezala, honetan ere sinetsi egin behar da, eta ekin!

8 Nondik jo: aurrera begirako ohar batzuk

8.1 Corpusgintza

Corpusgintza funtsezkoa da hizkuntzaren prozesamenduaren gaineko ikerkuntzaren euskarri gisa. Gaur egungo ikerketa corpusetan oinarritzen da neurri handi batean –eta areago azkenaldian, ikasketa sakona (*deep learning*) eta haren inguruko metodologia eta teknikak nagusitzen hasi direnetik–. Testu-corpusak behar ditugu, beraz, era eta tamaina guztietakoak, oportunistak zein orekatuak, eta behar bezala etiketatu edo anotatuak. Aipatu dugun *Plan de Impulso de las Tecnologías del Lenguaje* espainiarrean ere azpimarratzen da corpusak eta gainerako oinarritzko baliabideak sortzeko premia.

Uste dugu, beraz, garaia dela Euskararen Erreferentziako Corpora sortzeko, eta, gure ustez, Lexikoaren Behatokia abiapuntu egokia litzateke horretarako: euskararen corpus handi eta orekatu bat behar dugu, informazio linguistiko aberatsez hornitua hainbat mailatan. Ahozko atala ahaztu gabe hemen ere, ezinbestekoa baita ahozkoaren teknologia garatzen jarraitzeko.

Badira hainbat ekimen corpusgintza-bide honetan. UPV/EHUK, Elhuyarrek, Euskaltzaindiak, Ixa Taldeak... nork bere corpusak ditu, eta nolabait bildu eta bateratu egin beharko genituzke. Gainera, eta ez gara aspertuko esaten, oso garrantzitsua dela uste baitugu: eduki horiek libreak izan behar dute, ikerketarako trabarik gabekoak.

8.2 Araugintza eta estandarra

Araugintzari dagokionez, berriz, erreferentziako lanen biltoki bateratu bat sortzea proposatu nahi genuke, informazio-sistema online integratu bat, bilaketa eta kontsultak erraztuko lituzkeena: Euskaltzaindiaren arauak, hiztegi estandarra, terminologia espezializatua, estilo-liburuak, han eta hemengo foroak (ItzuL, esaterako), leku-, pertsona- eta abarren izen zerrendak... Informazio asko dago han eta hemen, eta ona litzateke hori dena bildu eta integratzea. Arantzazun 2018ko urrian egin zen biltzarrean aurkeztu zen *Euskara Batuaren Eskuliburu*¹⁶ aurrerapauso handi bat da, noski, bide horretan.

Idazketarako laguntza integratuak ere behar ditugu, hainbat software-ingurunetan erabili ahal izateko modukoak: laguntza ortografiko-gramatikala eskainiko duten aplikazioak, esaterako; edo testuingurua aintzat hartuko duten zuzentzaileak, hizkera espezializatueterako egokituak eta estilo kontuak ere jorratuko dituztenak; hiztegi-laguntzak eta abar.

Eta honetan ere eduki libreak behar ditugu, trabarik gabe eskuratzeko moduan eta automatikoki prozesatzeko egokiak diren formatuetan ere argitaratuak.

¹⁵ Artikulu hau idazten ari garela, ez da oraindik publiko egin.

¹⁶ <https://labur.eus/uZVBZ>

8.3 Hiztegi gintza, terminologia

Lexikoaren eta terminologiaren aldetik, baliabide eguneratuak behar ditugu, esportaziorako modu estandarrak eskainiko dituztenak eta teknologia aurreratuak hornituak. Bestalde, azpimarratu nahi genuke hiztegiok lantzeko eta argitaratzeko edizio-ingurune integratu baten premia¹⁷.

Adituen eta hizkuntzalarien arteko lankidetzak bideratu ahal izateko, berriz, ezinbestekoak dira terminologia lantzeko tresnak ere. Euskara Errektoreordetzaren babespean 2008. urtetik aurrera abian den Terminologia Sareak Ehunduz programa (TSE)¹⁸ aipatu nahi genuke hemen, gauzak egiteko modu egokia dela iruditzen baitzaigu. Programa horri esker garatu da GARATERM lan-ingurunea, izen bereko corpusa eta terminologia elkarlanean lantzeko TZOS sistema (Terminologia Zerbitzurako Online Sistema) biltzen dituen¹⁹.

Azkenik, irakurketarako laguntzak ere aipatu nahi genituzke atal honetan: e-liburuetan (telefono mugikor, tableta eta ordenagailuetan irakurtzen ahal ditugun liburu digitaletan) integratutako hiztegiak behar ditugu, irakurketarako laguntza paregabea eskaintzen baitute. Bitxia da, baina teknologikoki zailtasun handirik ez duen honetan dagoeneko atzean geratu gara, beste hizkuntza batzuekin konparatuta; eta dagoen arazo bakarra, antza, mota honetako produktuak garatzeko finantziario falta da.

8.4 Itzulpen automatikoa, itzulpen erako laguntzak

Adimen artifizialaren historian, aplikazio "izarra" izan da beti itzulpen automatikoa, gizakion arteko komunikaziorako ezinbestekoa den itzulpen gintza arindu eta merkatzeko izan dezakeen ahalmenagatik, inondik ere.

Egia da itzulpen automatikoaren historia ez dela arrakasta ikusgarri osatutako historia bat, baina, azken hamarkadetan, ikuspuntu ekonomiko batetik interesekoa den zerbait izatera iritsia da, eta esango genuke oso perspektiba onak zabaltzen ari zaizkiola, problemari aurre egiteko dauden hurbilpen berriei esker, batik bat; eszeptikoenak ere konbertitzaileentzako perspektibak, gure iritzi.

Horregatik, uste dugu haren erabilpena zabaldu eta orokortu egin beharko litzatekeela gurea bezalako gizarte batean, hainbat foru eta esparrutan; izan ere, itzulpenik ezan komunikazioa gaztelaniara edo frantsesera lerratu ohi den egoera askotan, itzulpena automatikoki eta berehala egiteko aukera –ahozko funtzionalitatea barne– izateak lerratze horiek saihesteko bideak zabalduko litzatekeela pentsatzen dugu.

Azkenik, esan dezagun Euskal Herrian badela itzulpen automatikoaren ikerketan puntan ari den talderik, ez dagoela kanpora joan beharrik teknologia horren bila. Jo beza, horren berri izan nahi duenak, Modela proiektuaren webgunera²⁰, esaterako, eta bertan ikusiko du zertan diren kontuak orain.

8.5 Humanitate digitalak

Bukatzeko, humanitate digitalen olatua ere aipatu nahi dugu, indarrez dator eta. Hizkuntzaren prozesamendua giza zientzietara zabaltzen ari da. Hor ditugu artxibo digitalak, liburutegiak, testu-meatzaritza eta -analisiak, testu historikoen tratamendua, hizkuntzaren irakaskuntza eta abar, zeinetan perspektiba berriak zabaltzen ari diren ahotsa eta testuak automatikoki tratatzeko baliabide eta aukerei esker.

17 *Orotariko Euskal Hiztegia*-ren kasurako, Euskaltzaindiak badu martxan proiektu bat, zeinetan orain arte erabili den datu-basea eta metodologia ordezkatzeko dituen ingurune bat prestatzen ari baitira, egungo teknologiaz hornitua.

18 <https://labur.eus/tNyFB>

19 <https://labur.eus/BaEYd>

20 <https://labur.eus/WMgrb>

Arlo honetan mugimendu handia dago azkenaldian. CLARIN sarea (*European Research Infrastructure for Language Resources and Technology*)²¹ aipatuko dugu lehenik, Europa mailako azpiegitura bat, teknologia – zerbiztuak eta tresnak– eskaintzen duena humanitateetan eta giza zientzietan erabiltzeko.

Orobat, aipatzekoa da hizkuntzaren beraren irakaskuntzan teknologiak dagoeneko duen eta aurrerantzean seguru izango duen gero eta baliagarritasun handiagoa. Humanitate digitalak eta hizkuntzaren teknologia rol garrantzitsua jokatzeko ari dira dagoeneko hizkuntza i(ra)kaste den inguruneetan: hizkuntza i(ra)kaste osagarri digitalak baliagarriak suerta daitezke klase barruan zein kanpoan –metodo eta metodologia egokiekin erabiliz gero, jakina–, eta esan daiteke askotan i(ra)kaste metodo tradizionalak eraldatzen ari direla²².

9 Ondorioak eta zenbait proposamen

Ondorio gisa, honako hauek azpimarratu nahi genituzke:

- Bat eta garrantzitsuenak: hizkuntza-teknologiaren arloak garrantzi estrategikoa du hizkuntzaren etorkizunari begira.
- Bi: hizkuntza-teknologia ardatz bat izango da Europan datozen urteetan, eta inbertsioak egingo dira; Espainian bada dagoeneko plan bat, diruz ondo hornitua, eta Euskal Herrian ere pareko zenbait beharko genuke berandu baino lehen.
- Eta hiru: trena galduko ez badugu, ikerkuntza eta trebakuntza espezializatuak bultzatu behar dira, egiten ari garena ez baita aski.

Uste dugu, halere, egokiera onean gaudela erronkari heltzeko, orain arte egindakoa oinarri hartuta; baina sinetsi egin behar da eta aurrera egin.

Eta, bukatzeko, zenbait proposamen:

- Pentsatzen dugu beharrezkoa dela erreferentzia-zentro bat sortzea, unibertsitatearekin eta Langune/Eiken-ekin lankidetzan: HITZ, lehen aipatutako Hizkuntza Teknologiako Zentroa, izan daiteke, gure ustez, erreferentzia-zentro nagusi horri begirako lehen pausoa.
- Bestalde, lortu beharko genuke arlo hau Eusko Jaurlaritzaren hurrengo zientzia-planaren lerro nagusien artean sartzea, estrategikotzat hartzea; ez dadila izan, baina, txosten batean idatzita uztea bakarrik: ekintzak –eta dirua– behar dira, zenbait itsasargi-proiektu definitu eta garatzeko, eta erosketak publiko berritzailea sustatzen hasteko.
- Azkenik, uste dugu Euskaltzaindiak ere bere lekutik eragile izan beharko lukeela honetan guztian, eta duen erreferentzialtasuna baliatu administrazioen aurrean, gauzak bizkortu daitezela.

10 Erreferentziak

Aduriz I., Alegria I., Arregi O., Diaz de Ilarraza A. eta Sarasola K., 2017. *Hizkuntza-teknologia "datu handien" garaian: programa bilatzaileak, itzultzaileak...* Senez, 48, 191-200. ISSN: 1132-2152.

Rodrigo Agerri, Aitzol Astigarraga, Iñaki Alegria, Itziar Cortes, Arantza Diaz de Ilarraza, Igor Leturia eta Kepa Sarasola, 2017. Language Technology projects in the frame of the European Capital of Culture 2016. META-Forum, 2017. Brusela.

²¹ <https://www.clarin.eu/>

²² <https://labur.eus/MIUuS>

Alegria I. eta Sarasola K., 2017. Language technology for language communities: An overview based on our experience. Int. Conference on Communities in Control: Learning tools and strategies for multilingual endangered language communities (CinC 2017). Alcanena, Portugal.

EPRS (European Parliamentary Research Service), 2017. *Language equality in the digital age - Towards a Human Language Project*. Study. <https://labur.eus/kx1Pi>

Euskararen Aholku Batzordea (IKT Batzorde-atala), 2016. *Euskarazko IKTak. Gomendioak herri-aginteentzat*. Hizkuntza Politikarako Sailburuordetza. <https://labur.eus/0LLG>

Hernández I., Navas E., Odriozola I., Sarasola K., Díaz de Ilarraza A., Leturia I., Díaz de Lezana A., Oihartzabal B. eta Salaberria J., 2012. *Euskara Aro Digitalean - The Basque Language in the Digital Age*. META-NET White Paper Series, Georg Rehm eta Hans Uszkoreit (arg.). <https://labur.eus/FUD30>

Ministerio de Turismo, Energia y Agenda Digital, 2015. Plan de Impulso de las Tecnologías del Lenguaje. <https://labur.eus/VsB84>