

Hizkuntza-teknologiaren abagunea

Euskara jalgi hadi teknologiara, jalgi hadi mundura

Iñaki Alegria Loinaz
Xabier Artola Zubillaga
Arantza Díaz de Ilarraza Sánchez
Kepa Sarasola Gabiola

Ixa Taldea (UPV/EHU)

2017-10-10

Edukiak

Sarrera.....	5
Euskararen egoera hizkuntza-teknologiaren alorrean.....	5
Meta-NET: Euskara aro digitalean	6
Hizkuntza-teknologiak sustatzeko politikak.....	8
Europar Batasuneko sustapen-politikak	8
Tresnak eta baliabideak	9
Europako industria eta ikerkuntza.....	11
Sozioekonomia eta eleaniztasuna: politikak.....	12
Estatu Espainoleko sustapen-politikak	15
Beste sustapen-politika batzuk mundu mailan.....	18
Euskal Autonomia Erkidegoko sustapen-politikak.....	19
Nafarroako Foru Erkidegoko eta Iparraldeko sustapen-politikak.....	19
Euskara HTen arloan sustatzeko proposamen bat	19
Hizkuntza-baliabideen eta -tresnen garapena.....	20
Hizkuntza-tresnen integrazioa eta hodei-konputazioa.....	22
Itzulpen automatikoa	22
Industria eta administrazioko domeinu espezializatueterako proiektu transbertsalak.....	23
Ikerkuntza eta formazio espezializatua	24
Nazioarteratzea.....	24
Ondorio nagusiak	25

Sarrera

Txosten hau idazterakoan bi helburu izan ditugu gogoan: batetik, hizkuntza-teknologiaren arloan Europako beste hizkuntzekin alderatuta euskarak duen presentzia aztertzea; bestetik, Euskal Herrian epe ertainean arlo honetan egin ditzakegun urrats esanguratsu posibleak identifikatzea, abaguneak erakustea. Lana lxa taldeak egin du, EH Bilduren talde parlamentarioaren eskaerari erantzunez.

Erabilitako iturri nagusiak Europa edo estatu mailan azken aldian aurkeztu diren hiru txosten hauek izan dira:

- Hernáez 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/FUD3O>
- EPRS (European Parliamentary Research Service), 2017. *Language equality in the digital age - Towards a Human Language Project. Study*. <https://labur.eus/kx1Pi>
- Ministerio de Turismo, Energía y Agenda Digital, 2015. *Plan de Impulso de las Tecnologías del Lenguaje*. <https://labur.eus/VsB84>

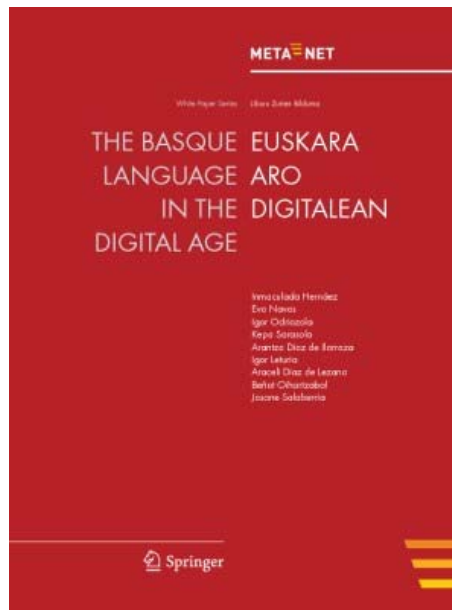
Sarrera honetaz gain, beste lau atal daude txostenean. Lehenengoak hizkuntza-teknologiaren alorrean euskarak bizi duen egoera aztertzen du, META-NETen txostena oinarri hartuta. Ondoren, hizkuntza-teknologiak sustatzeko indarrean dauden politika ofizialak aurkezten dira, goian aipatutako beste bi txostenak oinarri hartuta, Europakoetatik hasi eta Euskal Herri mailakoetaraino etorrira. Hirugarren atalean, hizkuntza-teknologiaren arloan euskara sustatzeko proposamen bat aurkezten da, eta bukaeran ondorio batzuk azaltzen dira labur.

Euskararen egoera hizkuntza-teknologiaren alorrean

Europa mailan hizkuntzaren prozesamenduan koordinazio-lanak egiten dituen META-NET Bikaintasun Sareak 30 “liburu zuri” argitaratu zituen 2012an, Europako hizkuntzak alor horretan zer egoeratan dauden aztertzearen¹. Europako Batzordeak sortutako META-NET sareak gaur egungo hizkuntza-baliabide eta -teknologiei buruzko analisi bat bideratu zuen, eta analisi horren emaitzak bildu nahi izan zituen gero aipatutako bildumako liburuetan. Analisi hori, Europako hizkuntza ofizialentzat ez ezik, Europako beste zenbait nazio- eta eskualde-hizkuntza garrantzitsuentzat ere egin zen. Horietako bat euskararena da, *Euskara aro digitalean* izenburuko alea (Hernáez *et al.*, 2012), hain zuzen. Liburu zuri horien helburua hizkuntza-teknologiei eta beren potentzialari buruzko jakintza sustatzea da, eta hezitzaile, kazetari, politikari eta hizkuntza-komunitateei zuzenduak dira.

META-NETek egindako analisiaren ondorioen artean dago begi-bistakoa dena, alegia, aztertutako hizkuntza guztiak ez daudela egoera berdinean hizkuntza-teknologiei dagokienez, eta kasu askotan ikerketa-hutsune esanguratsuak daudela.

1 <http://www.meta-net.eu/whitepapers/overview>



Euskarari buruzko alearen egile eta koordinatzaile nagusia izan den [Inma Hernaez](#)ek laburpentxo hau egiten zuen:

Oro har, META-NETen ikerketa-lanak ohartarazten du Europako hizkuntza gehienak digitalki iraungitzeko arriskuan daudela, aztertutako 30 hizkuntzetako 21ek duten babes digitala “hutsa” edo, kasurik onenean, “eskasa” baita.

Azterketak agerian uzten du hizkuntza batek ere ez duela “babes bikaina”; ingelesa baino ez da jo “babes ona” duen hizkuntzatzat, eta haren atzetik daude alemana, espainiera, frantsesa, italiara eta nederlandera, “babes ertainaz”. Bestalde euskarak, katalanak eta galegoak, bulgariara, greziera, hungariara eta polonierarekin batera, “babes zatikatua” dute, eta, hortaz arrisku handiko hizkuntzen multzoan ezarri dira horiek ere.

META-NETen arabera, ezinbestekoa da European ekimen koordinatu eta eskala handiko bat sortzea, dauden hutsuneak beteko baditugu eta teknologia ahalik eta hizkuntza gehienetara transferituko badugu. Izan ere, biztanleko karga finantzarioa oso handia da hizkuntza-komunitate txikientzat; hizkuntzen artean teknologia trukatu beharra dago; nabarmena da baliabideen, tresnen eta zerbitzuen elkarreragingarritasun falta; eta bistakoak dira hizkuntza-mugak eta muga politikoak maiz bat ez etortzearen ondorioak.

Esan bezala, euskarari buruzko alean euskarak alorrean egun duen egoera deskribatzen da, eta liburu hori hartu dugu, hain zuzen, txosten honen atal hau osatzeko oinarritzat.

Meta-NET: Euskara aro digitalean

Euskararen egoera, hizkuntza-teknologiaren aldetik, “baikor baina zuhur aztertu beharrekotzat” jotzen da liburu zurian. Ia hiru hamarkadatan zehar egindako lanari esker, “badago HT ikerketa-komunitate bideragarri bat Euskal Herrian, espainiar eta euskal ikerketa-programen bidez bultzatzen dena. Hainbat baliabide eta punta-puntako teknologia ekoitzi eta banatu dira euskararako. Hala ere, garatu diren baliabideen irismena eta tresnen multzoa oso mugatuak dira oraindik ere, gaztelaniarako (eta, noski, ingeleserako) dauden baliabide eta tresnekin alderatuta; beraz, ez dira nahiko, ez kalitateari

dagokionez, ez kantitateari dagokionez, benetako jakintzaren gizarte eleaniztun bat sustatzeko beharrezkoak diren teknologia motak garatzeko.”

Hona hemen, liburu zuritik hartuta, euskararentzako dagoen hizkuntza-teknologiaren egoera deskribatzen duten taula nagusiak:

Hizkuntza-teknologiak (tresna, teknologia eta aplikazioak)

	Kantitatea	Eskuragarritasuna	Kalitatea	Estaldura	Heldutasuna	Iraunkortasuna	Moldagarritasuna
Hizketaren ezagutza	2	1	1	1	4	3	2
Hizketaren sintesia	2	3	4	4	4	3	3
Analisi gramatikala	4	2,5	4	4	4	2,5	2,5
Analisi semantikoa	1	1,5	2	1	1	1	1
Testu-sorkuntza	1	0	0	0	0	0	0
Itzulpen automatikoa	3	5	2	3	3	2	2

Hizkuntza-baliabideak (baliabideak, datuak eta jakintza-baseak)

	Kantitatea	Eskuragarritasuna	Kalitatea	Estaldura	Heldutasuna	Iraunkortasuna	Moldagarritasuna
Testu-corpusak	2	4	3	2	3	4	2,5
Hizketa-corpusak	3	2	3	2	3	3	2
Corpus paraleloak	2	4	2	2	2	2	1
Baliabide lexikalak	4	4	4	5	5	4	3
Gramatikak	2	2	2	2	2	2	2

Zenbakiok (0tik 6rako eskalan) agerian uzten dute non ditugun indarguneak, baina baita non dauden ahulgune edo gabeziarik handienak ere.

Liburuaren ondorioen artean beste taula batzuk ageri dira. Taula horietan aztertutako hizkuntzak multzoka sailkatuta daude, baliabide edo tresna garrantzitsuenei dagokienez dauden egoeraren arabera². Hona hemen taula horietan dagoenaren laburpen gisako bat, euskarari dagokionez:

- Hizketa-prozesamendua: Osagabea.
- Itzulpen automatikoa: Apala/Ez³.
- Testu-analisia: Osagabea.
- Baliabideak: Osagabea.

Datu horien argitan bistan da, beraz, euskararen egoera ez dela denik eta txarrenea, baina ez da ona inola ere: badauzkagu, eduki, hainbat baliabide, tresna eta produktu; baina ez dira aski. Esan behar da taula horietan ingelesa dela egoera *onean* ageri den hizkuntza bakarra, baliabide edo/eta tresna guztiei dagokienez. Alemana, espainiera eta frantsesa, berriz, egoera ertainean agertzen dira kasu gehienetan.

Hizkuntza-teknologiak sustatzeko politikak

Aurreko atalaren oinarritzat hartu den txostenean (Hernández et al., 2012) honako hau irakur daiteke:

Europar, desberdina da, hizkuntza batetik bestera, hizkuntza-teknologiaren eskuragarritasuna eta erabilera. Horren ondorioz, desberdinak behar dute izan, halaber, hizkuntza bakoitzerako hizkuntza-teknologiaren ikerketa eta garapena bultzatzeko behar diren ekimenak.

Kalitate handiko hizkuntza-teknologia garatzea ezinbestekoa eta presazkoa da euskararen kasuan. Ezinbestekoa, hizkuntzaren beraren irautea bermatu nahi bada; eta presazkoa, inondik ere, teknologia hauen bilakabidea eta berritze-abiadura ikusita atzean geratzeko arriskua benetakoa baita. Euskal komunitatearen ikuspegitik, bada, helburu polita –eta egingarria– litzateke epe laburrean –edo ertainean– teknologia horiek guztiak aurreko atalean aipatutako egoera ertain edo on horietara ekartzea.

Teknologia horiek bultzatzea ohikoa da gure inguruko administrazioetan. Laguntzak bi arlotan bana daitezke, lotuta badaude ere: ikerkuntza eta garapena, batetik, eta hizkuntzen babesa, bestetik. Euskararen kasuan biak uztar daitezkeelakoan gaude.

Atal honetan hainbat komunitatetan egindako sustapen-politikak aipatuko dira, Europar Batasunekoak eta Estatu Espainolekoak, besteak beste. Halaber, orain arte EAEn, Nafarroan eta Iparraldean emandako urratsak ere azalduko dira.

Europar Batasuneko sustapen-politikak

Atal honetan azaltzen diren aurreproiektua eta azterketa Europako Parlamentuaren Ikerketa Zerbitzuak aurten egindako txostenean (EPRS, 2017) daude oinarrituta.

² Egoerak honako hauek dira, onenetik txarrenera: *Bikaina / Ona / Ertaina / Osagabea / Apala/Ez* (ez dago halako tresna edo baliabiderik).

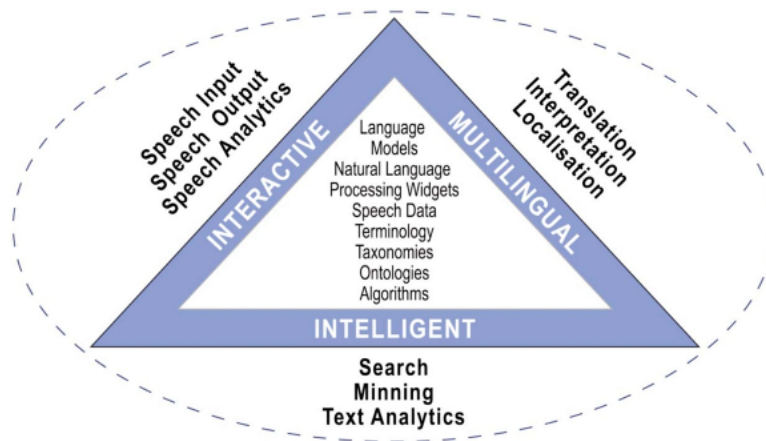
³ Gaur egun, egoera *Osagabea* dela esan liteke.

Aipatutako txostenak anbizio handiko egitasmo bat proposatzen du. Horretarako, orain arteko I+G+b lerroak zein industriaren ezaugarriak aztertzen dira, eta bai egoera sozioekonomikoan eta linguistikoan ere detektatu diren eraginak. Horrez gain, lerro argiak markatzen ditu txostenak. Gomendatzen dituen hainbat politika eta laguntza-lerro euskarara ekarri beharko lirateke lehenbailehen; alde batetik, beharra argia delako, eta, bestetik, proiektua onartuko balitz, orain arte gurean egin ditugun ahaleginengatik posizio onean egongo ginatekeelako modu kalifikatuagoan parte hartzeko.

Tresnak eta baliabideak

Txostenaren 3. atalean egiten da hizkuntza-teknologiaren (HT) sailkapena, honela: **teknologia semantikoak**, **hizketa-teknologia** eta **itzulpen automatikoa**. Horrez gain, **baliabideak** (*resources*) eta **tresnak** (*tools*) bereizten dira, eta hiru ezaugarri azpimarratzen dira: **elkarrekintzazkoa** (*interactive*), **adimentsua** (*intelligent*) eta **eleaniztuna** (*multilingual*). Ondoko eskeman laburbiltzen da teknologia:

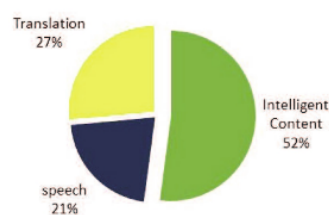
Figure 1: Language Technologies: tools and resources



Source: Taken from (LT-Innovate, 2013)

Merkatuari dagokionez, berriz, honako hauek dira ehunekoak:

Figure 12: European LT vendors by segment



Source: Taken from (LT-Innovate, 2013)

Txosten horren lehen taulan **teknologia semantikoari** dagozkion tresna ezagunenak aipatzen dira. Semantikak testuak ulertzen laguntzen du. Hona hemen arlo honetan aipatzen diren teknologia garrantzitsuenak:

- Analisi morfologikoa eta lematizazioa: lema edo hiztegi-sarrera beraren agerpen desberdinak identifikatzea, dagokien informazio gehigarriarekin. Euskararen kasuan funtsezkoa, bere morfologia aberatsa dela-eta. Funtsezkoa, halaber, bilaketarako.
- *Part-of-speech tagging* edo POS etiketatzea: hitzaren funtzio sintaktikoa ebaztea, testuinguruaren arabera. Euskarazko *zuen* hitz-forma adizkia (*eduki*) edo izenordaina (*zuek*) izan daiteke, eta agerpen bakoitzean erabaki beharko da. Funtsezkoa da ondoren aipatzen diren tresnetarako.
- Ortografia- eta estilo-zuzentzaileak: testuetan egindako akatsak-edo identifikatzea eta zuzentzen laguntzea.
- Izendun entitateen ezagutza, estekatzea eta erlazionatzea: testu batean pertsonak, erakundeak, tokiak, datak eta kopuruak identifikatzea, batetik, eta, ahal bada, datu-base bateko sarrerekin estekatzea (ohikoa da lotura horiek Wikipediarekin egitea), haien arteko erlazioak ezarriz.
- Terminologiaren erauzketa eta sailkapena: aurrekoaren antzekoa baina terminoei begira.
- Korreferentzia-ebazpena: pertsona, erakunde edota toki beraren aipamen desberdinak lotzen dituen teknologia. Adibidez, testu batean pertsona bera aipatzeko lau forma hauek erabil daitezke, besteak beste: *Juan Jose Ibarretxe*, *Ibarretxe*, *Lehendakaria* eta *bera*.
- *Parsing* edo analisi sintaktikoa: testuan perpausen subjektua, objektua, aditza eta bestelako predikatuak identifikatzea, esaldi nagusiak eta menpekoak barne.
- *Semantic rol labelling* edo rol semantikoen esleipena: sintaxi bidez ezagututako elementuei funtzio semantikoa esleitzea. Nork, non, noiz edota nola galderei erantzuteko balio du.
- Adiera-desanbiguazioa: adiera bat baino gehiago duten hitzetan, testuinguruaren arabera dagokiona lortzea. Adibidez, *banku* hitza agertzen denean, *banketxe* edo *aulki* adiera den identifikatzea.
- Diskurtsoaren analisia: testu baten osagai nagusiak (sarrera, hipotesia, arrazoibideak, ondorioak...) bereiztea.
- Testuen sailkapen semantikoa: testuak arloaren edo gaiaren arabera sailkatu ahal izatea (ekonomia, kirola...). Sailkatze hori zehatzagoa ere izan daiteke, eta hizkuntzaren identifikazioa ere garrantzitsua da.

Ikusten denez, tresna zerrenda luzea da; horietako batzuk oso arrakastatsua dira (ortografia-zuzentzaileak, esaterako), eta beste batzuk ikergai dira oraindik (hertsiki hartuta semantikaren arlokoak direnak, bereziki).

Hizketa-teknologiari dagokionez, berriz, hizketaren sorkuntza eta ulermena dira erronka nagusiak. Lortu beharreko kalitatearen eta zarataren eta aldakortasunaren (hiztun kopuruak, motak, dialektoak...) inguruko baldintzak dira aztergai nagusiak. Aurrerapen handiak egin dira *call centre* eta *assistant* direlako teknologiak garatzeko beharretik abiatuta.

Itzulpen automatikoari dagokionez, azkenik, esan behar da erronka handia dela, eta oraindik duen kalitatearekin, oro har, ez dela itzulpen zuzenik lortzen. Edozein kasutan, elkarrengandik gertu dauden hizkuntzetan, itzultzaileen produktibitatea handitzeko balio du, eta aski da, halaber, webeko informazioa ulertu ahal izateko ere. Euskararen kasuan, erronka handia da ikerketaren ikuspuntutik, tipologia aldetik eta hitz-hurrenkera librearengatik inguruko hizkuntzetatik urrun baitago, eta horrek zailtzen du eginkizuna.

Zeinu-hizkuntzen itzulpena aurrekoarekin lotuta dago, baina horrek baditu berezitasun estra batzuk.

Teknologia horiek garatzeko baliabideak prestatu behar dira. Horietako gehienak hizkuntzarekin lotuta daude eta adituek prestatu behar dituzte: hiztegiak, corpusak (gordinak; edo automatikoki ikasi ahal izateko linguistikoki anotatuak), ontologiak, datu-base semantikoak... Aurreko atalean ikusi den legez, ingelesa nagusi da tresnetan zein baliabideetan.

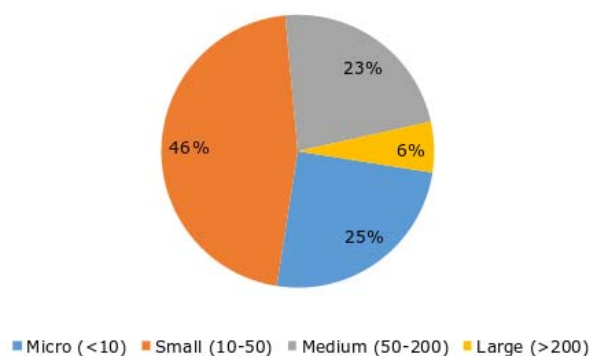
Atal guzti horietan aldaketa garrantzitsu bat gehitu behar zaio: *deep learning* edo ikasketa sakona izeneko tekniken agerpenarekin emaitzak hobetu egiten dira (aurretik aipatutako baliabideak eskuragarri baldin badaude, behar diren neurrian), baina konputazio-ahalmenaren beharra ere handitu da. Horregatik, gaur egun, aurrerapenak lortzeko azpiegitura informatiko ahaltsua (superkonputazioa izena eman ohi zaio) ere garrantzitsua da.

Europako industria eta ikerkuntza

Aipatutako txostenetik, honako ideia nagusi hauek nabarmen daitezke:

- Europan enpresa handi batzuk sortu dira (ingelesaren inguruan, batez ere): SAP, SDL, Systran... Gehienak, hala ere, enpresa txikiak dira.
- Ikerketan badira kasu arrakastatsuak unibertsitatean edo I+G zentroetan, baina alde industrialak egin du kale. Europa lehiakorra da ikerkuntzan, baina enpresak sortzen ez.
- Itzulpen automatikoaren arloan Europak finantzatutako Moses tresnak arrakasta handia izan du.
- META-NET bitartez "jokalariai" saretzen ari dira.
- Azken urteetako proiektuak Horizon 2020-ICT 17 deiaren bitartez bideratu dira.

Ondoko irudian ikus daitekeenez, enpresa txikien pisua oso handia da, eta hori Euskal Herriko industriaren ezaugarriekin bat dator.



Source: (LT-Innovate, 2013)

Bestalde, ikerkuntza eta baliabideen atala txostenaren 3. kapituluaren lantzen da luze eta zabal. Azpimarratzekoa da laburpenean ikerkuntzaren alde egiten den aldarria. Bertan eskala handiko eta epe luzeko I+G+b egitasmo bat proposatzen da, 2030. urtera arte iraungo lukeena.

Ikerkuntzan kontuan hartzekoak dira ildo hauek, besteak beste: adimen artifizialaren paradigma irauli duen ikasketa sakonarena (*deep learning*), hizkuntzatik independenteak diren metodoak,

domeinua/gaia aldaketak ondo jasaten dituzten metodoak eta abar. Horien aurrean, ziurtzat jo daiteke HTen belaunaldi berri baten aurrean gaudela.

3. Refocus and strengthen research in the Human Language Project

The first and probably most crucial research policy is to set up, under the umbrella of the HLP, a large-scale, long-term research and development and innovation (R&D&I) funding programme, in which basic research, applied research and development, as well as innovation and commercialisation, work closely together in order to develop technologies for 'deep natural language understanding' by the year 2030. Fields needed to contribute to this goal are computational linguistics, linguistics, artificial intelligence, language technology, computer science and cognitive science.

The key goal of the research activities of the HLP is deep natural language understanding (including language generation). Current LT approaches are still rather shallow; many research and technology breakthroughs are needed in order to move any closer to the goal of full natural language understanding. Nevertheless, recent breakthroughs in artificial intelligence, and a fresh look at recent results in linguistics, can bring about the needed technology shift for the next generation of LT. With regard to research, key ingredients of the HLP must be artificial intelligence, language technology, machine learning and knowledge technology. Technological goals should be broad coverage, high quality and high precision. It is worth mentioning that the 'Multilingual Value Programme', as suggested by the Strategic Research and Innovation Agenda 2016 (SRIA), can be conceptualised as part of the Human Language Project. In terms of funding sources, a mix can be envisaged between Horizon 2020 (Work Programme 2018-2020) and the next Framework Programme (2021+), as well as national and regional funding sources.

Europak FP7 deialdian finantzaturako proiektuak ere aipatzen dira txosteneko 27. taulan (161. or.). Euskal Herriko unibertsitate eta ikerketa-zentroen parte hartzea ez dago inon bilduta, baina garrantzitsua (izan) dela esan dezakegu. Informazio hori biltzea garrantzitsua litzateke Europako deialdiei begira dugun oinarria finkatzeko.

Sozioekonomia eta eleaniztasuna: politikak

Hizkuntza-aniztasunak Europan aberastasun kulturala dakar. Hizkuntzak bizirik mantendu behar dira, baina erronkak ere hor daude: mugikortasuna, zerbitzu publikoen hedapena... Azpimarratzen den merkatu bakar digitalerako ere erronka da hizkuntza-aniztasuna. Ingelesaren ezagutza ere mugatua da. Teknologiak lagun dezake, inondik ere, horretan guztian.

Erronkaren tamaina adierazteko, txostenaren 3.2 atalean hainbat inplikazio aztertzen dira: sozio-demografia, migrazioak, zerbitzu publikoak, identitatea, e-merkataritza eta abar.

Hizkuntza minorizatuak edo baliabide urrikoak dituzten erronkak ere aipatzen dira. Aurrekoari baliabide eta tresnen eskasia edo falta gehitu behar zaio. Enpresa txikietan ere badu eragina horrek.

Hungariako adibidea aipatzen da:

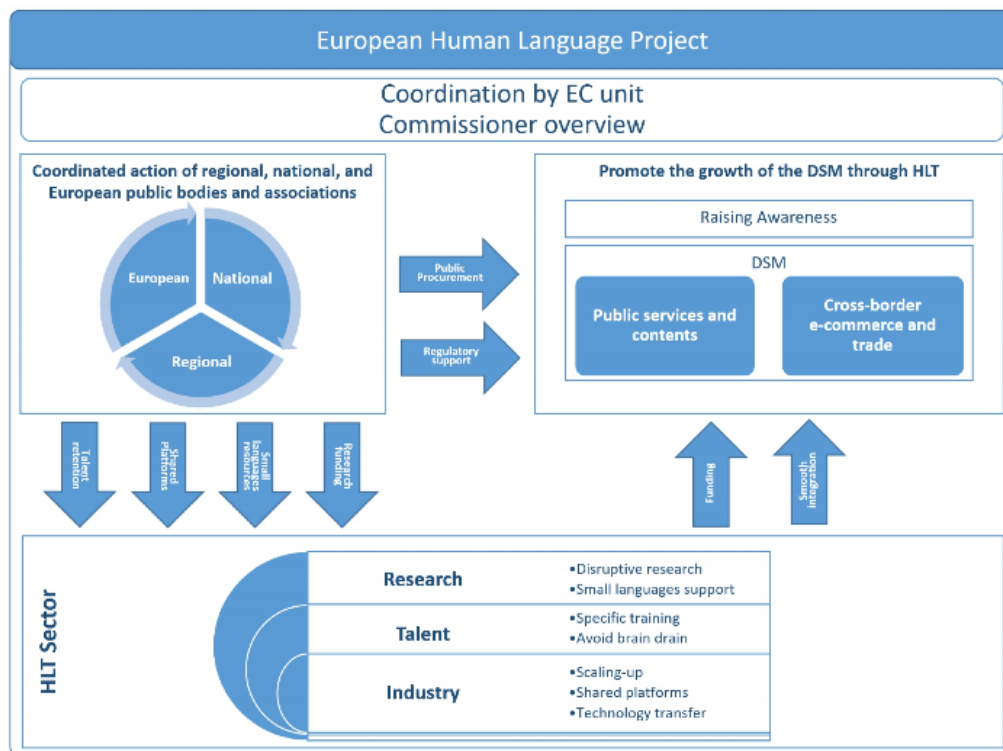
For countries with smaller languages this phenomenon has profound consequences. Take Hungary for example, a country with an isolated language and a low percentage of the population speaking foreign languages. Currently, the people from the rest of the EU shopping on on-line Hungarian websites is negligible. Overcoming language barriers could potentially increase the on-line transactions coming into Hungary from other countries up to 5.9 million, more than twice the current population buying on-line in Hungary (2.2 million).

Hori guztia ezaguna izan arren, HT hauek ez daude berariaz kontsideratuta Europako politiketan. Adibide gisa, gaixoen eskubideak arautzen dituen zuzentarauan ez dira aipatzen hizkuntza-eskubideak.

Horren aurrean politika globalak eta koordinatuak behar direla azpimarratzen da. Txostenaren 4. kapituluaren lantzen dira luze eta zabal. Aipatuak ditugu aurretik ikerkuntzakoak.

Txostenaren 87. orrian dago proposatzen duten *European Human Language Project* diziplinariako proiektuaren eskema (Giza Hizkuntzaren Proiektu Europarra):

Figure 41: The European Human Language Project



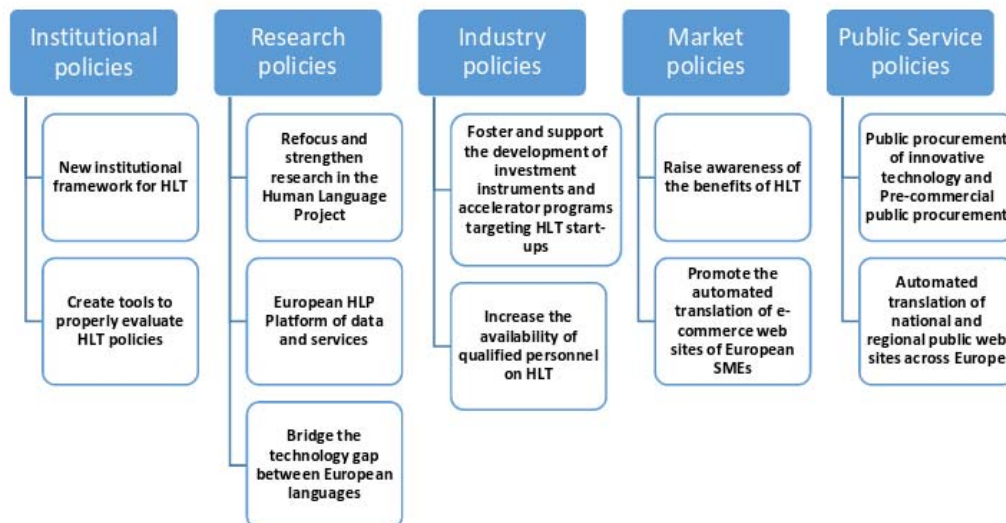
Source: Compiled by the authors

Bete nahi diren hutsunak hauek dira:

- Sozialak: hizkuntzen arteko aldeak, arriskuan dauden hizkuntzak, komunikaziorako mugak.
- Ekonomikoak: administrazioaren informazio eleaniztun mugatua, merkatu digital bakarrerako zailtasunak.
- Sektorekoak: ikerkuntzaren beharra, talentuen galera HTen arloan, enpresen zatikatzea, ekosistemaren galera (AEBekiko).

Egitasmoa aurrera eramateko, honako politika hauek proposatzen dira (txosteneko 88. orritik hartutako irudian zehazten dira, ikus behean):

- Instituzionalak
- Ikerkuntzakoak
- Industrialak
- Merkatuarenak
- Zerbitzu publikoenak



Planteatzen diren politiken artean, honako hauek azpimarratu nahi ditugu gure ingurura ekartzeko duten interesa dela-eta (jatorrizko testuan adibideak daude):

- HTen ikerkuntza areagotzea eta birformulatzea.
- Plataforma komun bat ezartzea.
- Hizkuntzen arteko aldeak murriztea teknologiaren aldetik.
- *Start-up* enpresak sortzea eta azelerazio-programak bultzatzea.
- Adituen formazioa sustatzea.
- Tresnak eta zerbitzuak ezagutaraztea.
- Enpresa txikien webgune komertzialen itzulpen automatikoa bultzatzea.
- Teknologiaren erosketa publikoa.
- Webgune publikoen itzulpena, HTak erabiliz.

Proposamenak ebaluatu egin dira adituen artean; beheko taulak jasotzen ditu emaitzak.

Oraindik zehaztasunetan sartu ez bada ere, Europako Parlamentuak eskatu duen txosten hau irakurrita hainbat ondorio atera daitezke:

- Hizkuntzen etorkizunei begira arloak garrantzi estrategikoa du.
- Hizkuntza-teknologiaren arloa Europarako ardatz bat izango da hurrengo urteetan.
- Ikerketan, berrikuntzan eta transferentzian zentratuko dira indarrak, eta inbertsioak egingo dira.
- Euskara bultzatu nahi bada, eta Europako beste hizkuntzen mailan edo gertu egoteko asmoa baldin badago, urrats garrantzitsuak eman behar dira hemen.
- Beste hizkuntzetan izan den garapena, eta izango dena, euskarazko komunikaziorako oztopo moduan ikus daiteke, jendarte eleaniztun batean bizi garen neurrian; baina aukera bat ere bada, euskara bultzatzeaz gain ikerketa eta garapenerako gune garrantzitsu bat sortzea/indartzea lortzen baldin badugu.

Proposed policies for the European Human Language Project		Criteria										Overall assessment (1 to 100)	Stakeholders involved						
		Costs and benefits	Political feasibility	Feasibility in the European context	Effectiveness	Sustainability	Risks and uncertainties	Coherence with EU objectives	Potential ethical and other impacts	Innovation	Single Market		Economic growth and job creation	Tackling inequalities	EP	EC	National Governments	Regional Governments	Research institutions
Criteria weight (1 to 10)		9	7	6	8	6	4	6	4	8	10	8	10						
Research	Refocus and strength research in HLT	↑	↑	↑	↑	↑	↓	↑	↓	↑	↑	↑	93						
Research	Promote the European LT Platform of data and services	↑	↓	↑	↑	↓	↓	↑	↑	↑	↑	↑	90						
Institutional	Reinforce the role of HLT within the institutional framework of multilingualism related bodies	↑	↑	↓	↑	↑	↑	↓	↓	↑	↓	↑	83						
Public services	Public procurement of innovative technology and Pre-commercial public procurement	↓	↑	↑	↓	↓	↓	↑	↓	↑	↑	↑	82						
Industry	Foster and support the development of investment instruments and accelerator programs targeting HLT start-ups	↓	↑	↓	↑	↓	↓	↓	↓	↑	↑	↑	78						
Research	Bridge the technology gap between European languages	↓	↓	↑	↓	↓	↓	↑	↑	↑	↑	↑	76						
Market	Promote the automated translation of e-commerce web sites of European SMEs	↑	↑	↑	↓	↓	↓	↑	↓	↓	↑	↑	71						
Industry	Increase the availability of qualified personnel on HLT	↓	↓	↑	↓	↓	↓	↑	↓	↑	↑	↑	70						
Market	Raise awareness of the benefits for companies, public bodies, and citizens of the availability of on-line services, contents and products in multiple languages	↓	↓	↓	↓	↓	↓	↓	↓	↑	↑	↑	70						
Public services	Foster the translation of national and regional public webs and documents to other EU languages by using HLT	↓	↓	↑	↓	↓	↓	↓	↓	↑	↑	↑	63						
Institutional	Create tools to properly evaluate HLT policies	↓	↓	↓	↓	↓	↓	↓	↓	↑	↓	↑	57						

Estatu Espainoleko sustapen-politikak

2015eko bukaeran onartutako *Plan de Impulso de las Tecnologías del Lenguaje* delakoa⁴ mugarri izan da arlo honetan. Planaren lan-taldean UPV/EHUko Ixa taldeak hartu zuen parte. Bestek beste, German Rigau planaren koeditorea izanda.

Planaren helburua argia da:

⁴ <https://labur.eus/VsB84>

El Plan de Impulso de las Tecnologías del Lenguaje tiene como objetivo fomentar el desarrollo del procesamiento del lenguaje natural y la traducción automática en lengua española y lenguas cooficiales.

Gauzatze-epea 2016-2020 da, eta 90 milioi € inguru inbertitzea aurreikusten da. Azpimarratzekoa da gaztelania ez diren hizkuntzak ere hartzen dituela kontuan, baina beti dagozkien erkidegoen babesa bilatuz.

Planaren ardatzak hauek dira:

- Babesa azpiegitura linguistikoen garapenari. Aurreko planean aurkeztu diren tresnak eta baliabideen garapena erraztea du helburu, horrela tresna eta baliabide horiek bertako industriarentzat erregai egokia izan daitezen.
- Bultzada hizkuntza-teknologien industriari. Ikerkuntza-alorretik industriara ezagutza transferitzeko asmoa du ardatz honek. Baita sektorearen parte diren enpresen eta erakundeen internazionalizazioa ere.
- Administrazio publikoa hizkuntza-teknologien bultzatzaile. Administrazio Publikoan plataforma bateratuak sortzea proposatzen da ardatz honetan, hizkuntzaren prozesaketarako zein itzulpen automatikorako. Horrez gain, sektore publikoaren informazioa berrerabiltzeak erraz dezake baliabide berrien sorrera.
- Itsasargi-proiektuak hizkuntza-teknologietan. Administrazioak inizatiba hartuko du hainbat sektore estrategikotan: osasuna, turismoa, hezkuntza; inizatiba publiko-privatuaren aukera irekita utziz etorkizunerako. Hori guztia etorkizunerako ikaskizun ere izango da.

FORTALEZAS	DEBILIDADES
<ul style="list-style-type: none"> • Alto nivel investigador en procesamiento de lenguaje natural coordinado por la Sociedad Española para el Procesamiento del Lenguaje Natural. • Buena gobernanza del idioma español (RAE, Asociación de Academias de la Lengua Española en Iberoamérica). • Gran experiencia con el multilingüismo por la existencia de lenguas cooficiales. 	<ul style="list-style-type: none"> • El sector está compuesto por pequeñas y medianas empresas que no alcanzan la capacidad industrial para competir en el mercado internacional o completar la cadena de valor en España. • Dificultad de transferencia de conocimiento del sector investigador a la industria debido principalmente al carácter transversal y multidisciplinar del procesamiento de lenguaje natural.
OPORTUNIDADES	AMENAZAS
<ul style="list-style-type: none"> • Alto potencial de internacionalización del español y de cooperación con Iberoamérica. • Nuevos servicios públicos para ciudadanos y empresas en sectores estratégicos (sanidad, turismo, educación, etc.). • Mercado en fuerte crecimiento asociado a la innovación y al desarrollo. • Potencialidad de la reutilización de la información del Sector Público (RISP) como recursos lingüísticos muy valiosos para la industria y la investigación. 	<ul style="list-style-type: none"> • Pérdida de competitividad económica e industrial de España e Iberoamérica. • Subdesarrollo digital del español. • Extinción digital de las lenguas cooficiales. • Fuga de investigadores y profesionales y deterioro del sector investigador español.

Hiru printzipiotan oinarritzen da:

- Eskala handiko proiektuak identifikatu behar dira, erroan I+G+b izango dutenak.
- Ikuspuntuak eta bitartekoek epe ertainari begira (5-10 urte) izan behar dute.

- Baliabideak kontzentratu egin behar dira efizientzia eta eraginkortasuna handitzeko.

Inguruan ditugun ikerketa-taldeak (UPV/EHU) eta enpresa/unibertsitate aliantzak (Langune) aipatzen dira Madrileko txosten honetan. Baita European aukera berriak irekiko direla ere.

AMIA bat (aukera, mehatxu, indargune eta ahulezien analisia) proposatzen da (ikus goiko irudia). Azpimarratzekoa da bertan ikerkuntzaren eta eleaniztasunaren indargunea.

Plana aurrera eramateko lau ildo nagusi aipatzen dira, eta bakoitzaren barruan urratsak egiten ari dira gaur egun. Hona hemen ildoak eta adibideak:

- Azterketak. Itzulpen automatikorako tresnak edo prestazio handiko plataformak (hizkuntzaren prozesaketa) aztertzen ari dira.
- Hitzarmenak eta enkarguak (*encomiendas*). Plangintzaren bulego teknikoaren lana aurrera eramateko bi enkargu lotu dira.
- Itsasargi-proiektuak. Zaintza teknologikoaren inguruan hainbat proiektu ari dira aurrera eramaten.
- Berrikuntzaren erosketa publikoa. Ildo hau konplexuagoa da, eta beheko helbidean⁵ dago azalduta.

Planaren barruan UPV/EHUko Ixa taldeak hainbat enkargu gauzatu ditu: tresnen inguruko txostena, oinarrizko tresnen plataforma bat eta dibulgazioa *hackathon* baten bidez. *Infoday* batean ere hartu du parte⁶.

Aukerak daude/egon dira Ministerioaren erreferentzia-zentro bat izateko, baina bestelako babesik gabe eta UPV/EHUtik aparteko izaera juridikorik gabe ezinezkoa izan da oraingoz.

Plana martxan jarri da 2016an, eta hainbat neurri ezarri dira:

- Planaren gobernantza, koordinazioa eta jarraipena egiteko. Zuzendaritza-batzordea eta Adituen Batzordea sortu dira.
- Azpiegitura linguistikoen garapena. Hemen kokatzen da interoperabilitatearen definizioa eta oinarrizko baliabideen eta tresnen eskaintza, gaztelaniarako eta beste hizkuntza ofizialetarako. Horrez gain, unibertsitatetik eta ikerketa-zentroetatik egiten den transferentzia hobetzeko beharra eta internazionalizazioa ere aurreikusten dira.
- Administrazio publikoa hizkuntza-industrien motorra: plataforma publiko baten garapena bere ezaugarri guztiekin (eskalagarritasuna, konfidentzialtasuna, domeinu- edo gai-egokitzapena eta abar). Edukien berrerabilpena ere azpimarratzen da (*Reutilización de la Información del Sector Público*, RISP).
- Itsasargi-proiektuak martxan jartzea.

⁵ <http://www.agendadigital.gob.es/tecnologias-lenguaje/compra-publica-innovacion/Paginas/compra-publica.aspx>

⁶ <http://www.agendadigital.gob.es/tecnologias-lenguaje/eventos/Paginas/infoday-abril-2017.aspx>

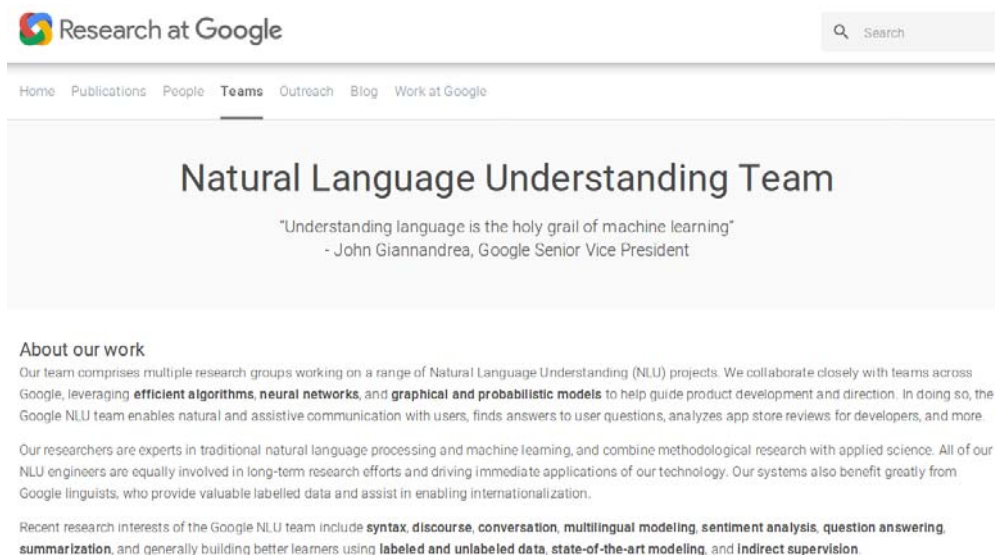


HTak bultzatzeko 90 milioi euro inbertituko ditu Espainiako Gobernuak.

Mundu mailako beste sustapen-politika batzuk

Estatu Batuetan gaia ere agenda ekonomikoa dago. Arloko teknologiaren azken aurrerakuntzetan (adimen artifiziala, ikasketa sakona...) administrazioak, unibertsitateek eta enpresa handiek (Google, Amazon) dituzten programez gain, badaude berezko proiektu interesgarriak. Adibidez: *Low Resource Languages for Emergent Incidents (LORELEI⁷)*. Proiektu hori DARPAk bultzatua eta finantzatua da, eta industriak zein unibertsitateek hartzen dute parte. Beste bat *Deep Exploration and Filtering of Text (DEFT⁸)* proiektua da. Proiektu horietan diru asko inbertitzen da.

Google erraldoia azken 15 urteetan buru-belarri sartu da hizkuntza-teknologiaren arloan, bai ikerketan, eta bai bere bilatzailean (eta beste aplikazio batzuetan) funtzionalitate berriak eskaintzeko ere. Programa bereziak ditu unibertsitateekin ere lankidetzan aritzeko.



⁷ <http://www.darpa.mil/news-events/2015-10-08>

⁸ <https://www.darpa.mil/program/deep-exploration-and-filtering-of-text>

Bestalde, hizkuntza ofizial anitz dituzten estatuetan teknologia aukera gisa ikusi dute, plangintzak egin dituzte eta bai erreferentzia-zentroak sortu ere.

Adibidez, *Indian Technology Development for Indian Languages* (TDIL⁹) programa sortu dute. Proiektu horren webgunean hainbat emaitza eskura daitezke: *toolkit* bat bertako hizkuntzetarako, testu-hizketa bihurgailuak...¹⁰.

Hegoafrikan ere hainbat programa eta gune abiatu dira ildo horretan. Hizkuntza ofizialak hamaika dira eta teknologiaren laguntza baliatu nahi dute. Beste programa batzuen artean duela gutxi *South African Centre for Digital Language Resources* izeneko zentroa sortu dute¹¹.

Euskal Autonomia Erkidegoko sustapen-politikak

Hizkuntza-teknologiak bultzatu izan dira azken urteetan Industria saileko Etorrek eta ElkarTek programen bidez. Berbatek, Ber2Tek, Modela eta ElkarOla proiektuak izan dira lerro horren ondorioa, eta beren emaitzak eman dituzte. Partaideen artean unibertsitateak, ikerketa-zentroak eta enpresa batzuk egon dira. Adibidez, orain dela gutxi bukatu den ElkarOla proiektuan¹² UPV/EHU, Elhuyar, Tecnalia eta Vicomtech izan dira partzuergoaren partaideak.

Zientzia eta Teknologia planetan, hala ere, indarra galdu du hizkuntza-teknologiaren arloak. 2015-2020 planean teknologia hauek ez dira agertzen lehen planoan, eta 4.0 fabrikazioaren zehar-lerro gisa lantzea aurreikusten da batez ere. *Cluster* politiketan ere Langune *precluster* izan zen 2016ra arte, baina dimentsio-arrazoiak direla-eta Eiken *clusterrean* (audiobisualak) integratu behar izan du.

Ez dago erreferentzia-zentrorik eta horrek mugatzen du Europako eta Espainiako programetan parte hartu ahal izatea.

Nafarroako Foru Erkidegoko eta Iparraldeko sustapen-politikak

Nafarroan, guk dakigunez, ez dago lerro estrategikorik arlo honetan, baina tokia eduki lezake indarrean dagoen agenda digitalean¹³.

Iparraldean Elkargoak ere agenda digitala bultzatzeko asmoa du. Bestalde, Baionan dagoen IKER ikerketa-zentroa¹⁴ lankide interesgarria da, eta arloko hainbat proiektutan aritu da lankidetzan lehen aipaturako erakunde batzuekin.

Euskara HTen arloan sustatzeko proposamen bat

Euskara HTen arloan sustatzeko eta garatzeko proposamen honetan, hainbat lan-lerro identifikatu ditugu, bost garapen-ardatzen arabera antolatuz. Gure proposamena oinarrituta dago, neurri handi batean, Ixa Taldeak gaur egun *Secretaria de Estado para la Sociedad de la Información y la Agenda Digital* (SESIAD, Espainiako Ministerio de Energía, Turismo y Agenda Digital) izena duen

⁹ <http://tdil.mit.gov.in/>

¹⁰ <https://labur.eus/aH6Vq>

¹¹ http://www.lrec-conf.org/proceedings/lrec2016/pdf/337_Paper.pdf

¹² <http://www.elkarola.eus/>

¹³ https://www.navarra.es/home_es/Gobierno+de+Navarra/Vicepresidencia+Desarrollo+Economico/Planes.htm

¹⁴ <http://www.iker.cnrs.fr/?lang=eu>

erakundearen eskutik abiatutako *Plan de Impulso de las Tecnologías del Lenguaje* delakorako egindakoan.

Aipatutako plan horretan, honako garapen-ardatz hauek ageri dira:

- Hizkuntza-baliabide eta -tresnen garapena.
- Hizkuntza-tresnen integrazioa eta hodei-konputazioa.
- Itzulpen automatikoa.
- Industria eta administrazioeko domeinu espezializatueterako proiektu transbertsalak.
- Ikerkuntza eta formazio espezializatua.
- Nazioarteratzea.

Horien arabera antolatu dugu, bada, proposamena.

Hizkuntza-baliabide eta -tresnen garapena

Ardatz hau helburu orokorreko prozesatzaile eta baliabide linguistikoak garatzera zuzenduta dago. Garrantzi handikoa da euskararako hizkuntza-baliabideak hobetzea, kopuruan zein kalitatearen aldetik, horri esker garatu ahal izango baitira hizkuntza tratatzeko tresna aurrerratuagoak, testuaren azaleko azterketa ez ezik sakoneko analisia ere gauzatuko dutenak, testu horren esanahia atzeman eta ulertzera hurbilduz.

Honako baliabide mota hauetaz ari gara:

- Hiztegi, terminologia-banku eta ontologia eleaniztunak, non arlo estrategikoetan (medikuntza, justizia...) erabili ohi diren termino espezializatuak bilduko baitira, eta antzeko beste banku eleaniztun batzuekin (UMLS¹⁵, justiziaren edo beste arlo batzuetako thesaurusak...) lotuta egongo direnak. Euskararen garapenerako estrategikoa da hainbat arlotako profesionalak unibertsitatean zein unibertsitatekanpo sortzen ari diren terminologia teknikoak aztertzea, berorren bilakabidea denboran aintzat hartuz. Horretarako, sortu behar dira terminologoari eta profesionalari, oro har, lan horretan laguntza emango dieten tresnak, hainbat arlotan: medikuntza, turismoa, justizia eta abar. Ontologiak direla-eta, azken urteetan halakoak sortzeko edota aberasteko prozedura erdiautomatikoak garatu dira. Ontologiak ezinbesteko baliabidea dira estaldura zabaleko prozesatze semantikoa egingo bada. Wordnet-en (Fellbaum, 1998) erabilerak –prozesatze semantikoa egiten den ia arlo guztietan erabiltzen da– argi eta garbi erakusten du halako baliabideen garrantzia. Europa mailako zenbait proiektu garrantzitsutan, hala nola NewsReader-en¹⁶, frogatuta geratu da ezagutza-ontologiak modu eraginkorrean erabiltzen ahal direla, testutik abiatuta meatzaritza eleaniztuna egiteko.
- Corpus elebakarrak, hainbat mailatan anotatu edo etiketatuak: morfologikoa, sintaktikoa, semantikoa, korreferentziak, rol semantikoak, izendun entitateak, denbora-espresioak, gertaerak...; corpus eleaniztun paralelo edo konparagarriak: en-eu, es-eu, fr-eu... Kontuan hartu behar dira halaber, plan honetan estrategikotzat jotako arloetako corpus espezializatuak: medikuntza, justizia, turismoa, administrazioa... Corpus horiek funtsezko

¹⁵ *Unified Medical Language System*: <https://www.nlm.nih.gov/research/umls/>

¹⁶ <http://www.newsreader-project.eu/results/demos/>

baliabideak dira prozesatzaileak entrenatu eta ebaluatzeko. Corpus orokortzat hartzen dira arlo edo gai jakin bati lotuta ez dauden testu bildumak, edo baita beren barruan hainbat arlotako testuak biltzen dituztenak ere. Garrantzi handikoa da egun ditugun euskarazko corpusen inbentarioa egitea, elebakar zein eleaniztunak aintzat hartuz, eta helburu zehatz batzuk finkatzea beroriek aberasteko, prozesatzaileak modurik egokienean entrenatu ahal izateko aski izango diren testu-masak eratzeko. Hona hemen ingelesezko corpus orokor nabarmenenak: i) Brown corpora: estreinako corpus elektronikoa izan zen, 1961ean sortua, eta milioi bat hitz dauzka; corpus orokorra da, 15 kategoriatako 500 testuk osatua; ii) *British National Corpus* delakoa (BNC): 1994an sortua, 100 milioi hitz (% 90 idatzia, % 10 ahozkoa); XX. mendearen bukaerako Britainia Handiko ingelesaren erakusgarri izan nahi du; iii) ukWaC corpora: corpus orokorra, 2008an sortua, webetik erauzia; Britainia Handiko ingelesezko testuz osatua da, eta 2.000 milioi testu-hitz dauzka (Ferraresi, 2008¹⁷).

- Hizkuntza historikoa eta ez-estandarra. Duela gutxi arte ikerkuntzak hizkuntza estandarren analisi-prozesaketak izan ditu jomuga, baina gero eta beharrezkoagoa da testu historikoak edo/eta euskalkietakoak ere prozesatzeko gai izatea. Konkretuki, maila morfologikoan, beharrezkoa da gai izatea euskalkietako formak eta, oro har, hizkuntza ez-normalizatukoak ezagutzeko: hala nola, txioetakoak (Twitter) eta sare sozialetakoak, eta bai antzinako testuetakoak ere. Paradoxikoki, batzuek zein besteek (testu historiko zein sare sozialetakoek) antzeko tratamendua behar izaten baitute.

Dispositibo mugikorrek eta sare sozialak gure arteko komunikazio moduak eta denborak aldarazten ari dira. Gizarte-hedabideek (*social media*) eta sare sozialek funtsezko zeregina betetzen dute egungo gizartean. Hori dela-eta, aintzat hartu behar da gizarte-hedabide horien ahala baliabideak eraikitzeke orduan. Hedabide horietan hizkuntza anitzetan sortzen den informazio-masa erraldoia, baina, ezin da eskuz prozesatu. Horretarako, informazio hori guztia denbora errealean prozesatu ahal izateko, hizkuntzaren eta komunikazioaren teknologiak behar ditugu, eta horregatik da garrantzitsua gizarte-hedabide eta sare sozialetan dagoen informazio horren guztiaren erabilera aurreikustea, ekarpen berri eta berritzaileak egingo badira. Wikipedia, blogak, foroak, sare sozialak (Twitter, Facebook, LinkedIn, Pinterest, Tuenti...) dira gaur egun gehien erabiltzen diren gizarte-hedabideak. Ezaugarri ezberdinak dituzten arren, denetan argitaratzen eta elkartrukatzen da informazioa erruz. Informazio hori guztia aprobetxatu egin behar da, hizkuntzaren prozesamenduaren (HP) arloan ekarpenak eginez.

Tresnei dagokienez, analisi-tresna orokorrak garatzeari eman behar diogu garrantzia; hala nola: i) analizatzaile morfosintaktikoak, testuak prozesatu behar direnean oinarritzko tresna baitira; ii) analizatzaile sintaktikoak: gaur egungo analizatzaile sintaktikoak hobetu daitezke hainbat teknika konbinatuz, hala nola erregeletan oinarritutakoak, tratamendu estatistikoa egiten dutenak eta sare neuronalen bidezkoak; iii) analizatzaile semantikoak: entitate izendunak ezagutzeko eta sailkatzeko sistemak, adiera-desanbiguatzaileak, hurbiltasun semantikoa kalkulatzeko duten sistemak. Tratamendu semantikoa behar-beharrezkoa da aplikazio aurreratuetan, itzulpen automatikoan edota erantzunak bilatzeko sistema adimendunetan, adibidez. Analisi semantikoaren arloan gaur egun erabiltzen ari diren teknikak askotarikoak dira, ikasketa gainbegiratu edo gainbegiratu gabetik

¹⁷ Ferraresi, A. 2008. "Constructing and evaluating web corpora: ukWaC". *Aston Postgraduate Conference on Corpus Linguistics*. Aston University, Birmingham, UK.

hasi eta sare neuronalak erabiltzen dituztenetarakoak; iv) diskurtsoaren analizatzaileak: testuaren ulerkuntzan harantzago joatea ahalbideratuko diguten tresnak garatu behar dira, korreferentziak ebazteko, denbora- eta espazio-espresioak ezagutzeko, eta diskurtso-unitateak identifikatzeko gai izango direnak.

Proposatzen diren jarduera gehienetarako honako lan-metodologia hau jarraitu beharko litzateke:

- Tresnaren garapena, lehendik dagoen teknologia orokor batean oinarrituz (teknologia hori behar bezain aurreratua dagoen kasuetan, noski).
- Tresnak probatu (testing-a), erabiltzaile-komunitatea ebaluazio kolaboratiboko kanpainen bitartez engaita dadin bultzatuz.
- Tresnak prestatu eta eskura jarri, deskargatzeko modu sinple bat eskainiz.

Hizkuntza-tresnen integrazioa eta hodei-konputazioa

Hainbat formatutan (testua, bideoa...) aurkeztutako informazio-masa handiak prozesatzeak hizkuntza-analisirako tresnak integratuko dituen azpiegitura modular bat eskatzen du. Tresna horien integrazio-arkitektura era askotan molda daiteke, *pipeline* klasikotik hasi eta grafo bidezkoraino, baina beti ere konputazio-ahalmen premia handiak izango ditu.

Gainera, interesgarria izango da tresna horiek hodei-zerbitzu gisa eskaini ahal izatea, erabiltzaileek balia ahal izan ditzaten, azpiegituraren (zerbitzariak, biltegitze-sistemak...) edo ezagutza teknikoaren aldetik kostu handitan sartu gabe. HParen kasuan, zerbitzu horiek arazoak ebaztera zuzenduta daude, edukien tratamenduekin edo itzulpen automatikoarekin zerikusia duten arazoei soluzio bat ematera, esaterako. Hodei-konputazioari esker posible da Internet bidez baliabideak partekatzea eskarien arabera, hau da, zeregin bati esleitutako baliabideak dinamikoki alda daitezke prozesuak eskatutakoaren arabera. Horrela, hodei-konputazioak paradigma-aldaketa bat ekarri dio hizkuntzaren prozesamenduari.

Horregatik oso garrantzitsua da gure baliabide eta tresnak (hizkuntzaren analisia, ulerkuntza, erauzketa, ezagutzea, sinplifikazioa eta abar egiten dutenak) hodeiratzea, kontuan hartuz beti baliabide eta tresna horiek egun indarrean dauden estandarrei jarraituz garatu beharko direla.

Itzulpen automatikoa

Bistan da itzulpen automatikoa eguneroko tresna bihurtzen ari zaigula gaur egun. Amaraunean badira hainbat zerbitzu¹⁸, zeinei esker norbanakoek nahiz enpresek beren webguneen ikusgaitasuna beste hizkuntza batzuetan hobetzen ahal duten, bai eta ezagutzen ez dituzten hizkuntzetan idatzitako informazioaren berri izan ere. Azken 60 urteotan Europa politikoki eta ekonomikoki nolabait egituratu eta bateratu egin da, baina kulturalki eta hizkuntza aldetik askotarikoa da oraindik. Hori dela-eta, Europako hiritarren arteko eguneroko komunikazioak zein negozioen edo politikaren esparrukoak hizkuntzaren mugekin egiten dute topo ezinbestean. Europar Batasuneko (EB) erakundeek ia mila milioi gastatzen dituzte urtero beren eleaniztasun-politikak mantentzeko, hots, dokumentuak hizkuntza batetik bestera itzultzen eta ahoz esandakoak aldi berean interpretatzen.

¹⁸ Adibidez: <http://translate.google.com>, <http://www.deepl.com> eta <http://opentrad.com>

Hizkuntzaren teknologiak eta ikerkuntza linguistikoak lan handia egin dezakete hizkuntza-muga horiek leuntzeko, dispositibo eta aplikazio adimendunekin konbinatuz gero gai baitira hizkuntza bera egiten ez duten pertsonen arteko komunikazioa bideratzeko, negozioak egin ditzaten erraztuz.

Atzerriko hizkuntzak ikastea izan da beti hizkuntza-mugak gainditzeko ohiko modua. Alabaina, teknologiaren laguntzarik gabe, EBko estatu kideetako 23 hizkuntza ofizialak gehi beste 60 hizkuntza ez-ofizial (euskara tarteko) ikasi beharra helmuga gaindiezina da Europako herritarrentzat, hala nola beren ekonomiarentzat, eztabaida politikorako zein zientziaren aurrerapenerako.

Garapenei dagokienez, garrantzitsua da egiten direnak PLATA izeneko plataforman integratzea. Plataforma hori itzulpen automatikoko sistema bat da, EBko *Directorate for Information Technologies and Communications* delakoak garatua, eta herri-administrazioetako web-atarietan integratzeko sortua. Plataforma horren bitartez, integratutako atarien itzulpena lor daiteke, denbora errealean, eta herritarrentzat erabat gardena den moduan. PLATA plataforma *opensource* diren itzulpen automatikoko motorretan (Moses, Apertium) oinarritzen da, zeinen bitartez web-atariak automatikoki itzul baitaitezke, gaztelaniatik abiatuz, hizkuntza koofizialetara eta ingelesera. Gaur egun PLATA baliatzen dute beren edukiak hainbat hizkuntzatarara itzultzeko Espainiako PAe atariak (*Portal de Administración Electrónica*¹⁹), PAGek (*Punto de Acceso General de la Administración*), *Secretaría de Estado de Administraciones Públicas*-ek eta IGAEk (*Intervención General del Estado*).

Euskaratik edo/eta euskarara automatikoki itzultzeko sistemak direla-eta, esan daiteke oinarri ona dugula eta egoera onean gaudela garatutako sistemak hobetuz joateko, beti ere, noski, ikerkuntzako eta erabilera libreko baliabideetako inbertsioari eusten baldin bazaio. Jarraitu behar da esperimendatzen ikasketa automatikoan oinarritutako teknika aurreratuagoetan (sare neuronaletan oinarritutakoetan barne), eta gero eta ezagutza linguistiko sakonagoaz baliatuz. Eta, jakina, egindako eta egingo diren garapenak hainbat hizkuntza-bikoteri aplikatuz (eu-en, eu-es, eu-fr...), eta garatuko diren sistema berriak egungoekin konbinatuz.

Industria eta administrazioko domeinu espezializatueterako proiektu transbertsalak

Administrazio Publikoan zein sektore pribatuan lehentasunezkoa da teknologiak aplikatzeko proiektuak bultzatzea, horrelako proiektuen bitartez egiaztatu ahal izango baita zer onura dakartzaten hizkuntza prozesatzeko teknikek hainbat sektore estrategikotan, eta probatu industriarako aplikazio orokorretan, beste esparru batzuetan berrerabili ahal izango direnak. Horretan “erosketa publiko berritzailea” delakoa funtsezko tresna izan daiteke.

Aipatutako planetan itsasargi-proiektuak interesgarriak izan daitezke, horizontalak eta transbertsalak baitira (hizkuntza-azpiegitura orokorrak, interes linguistikodun sektore publikoko informazioaren berrerabilera, herri-administrazioen hizkuntza prozesatzeko eta itzulpen automatikoko plataformak), eta sektore jakin batzuetan eragiten baitute, balio-kate osoa barne hartzen duten produktu eta zerbitzu bukatuak sortuz. Horietan erabil liteke aipatutako erosketa publiko berritzailea.

¹⁹ <http://administracionelectronica.gob.es>

Ikerkuntza eta formazio espezializatua

Euskal Herrian badira ibilbide luzeko ikerketa-talde espezializatuak arloan, unibertsitatean nahiz ikerketa-zentroetan. Garrantzi handikoa da talde horietan ikerkuntza sustatzea, hizkuntzaren prozesamenduaren arloan puntako diren tekniken berri egunean-egunean izatearren. Arloa oso zabala da, komunikazioaren fase guztiak barne hartzen baititu, esparru eta domeinu guztietan, adituen premia handia du eta etengabe aldatuz doa.

Aditu horiek hezi ahal izateko da behar-beharrezkoa master espezializatuak eta doktorego-programak bultzatzea, ikasle kopurua handitzeko eta nazioarteko lankidetzak handitzeko. Dagoeneko badago muinean kokatzen den master bat eta doktorego-programa bat²⁰, Erasmus-Mundus²¹ bermea duena gainera, baina beka- eta nazioarteratze-politika indartsuago bat beharko litzateke.

Nazioarteratzea

Ikerkuntza-sektoretik industriaranzko ezagutza-transferentzia sustatu behar da. Tinko bultzatu behar da hizkuntzaren teknologien inguruan talentuak sortzea eta erakartzea, unibertsitateetan dagoen ezagutza gorde eta sektoreko enpresetara edo berrikuntza sustatzeko programen bidez sortutako enpresa berrietara transferituz.

Euskal Herrian bada merkaturanzko hurbiltze hori modu eraginkor batean egin ahal izateko *know-how*-a. Gainera, une honetan badugu merkatuari aplikazio berriak eskaintzeko moduko teknologia, enpresei beren produktuetan integratuz gero salmenta konpetitiboak egiteko aukera emango liekeena.

Beste herrialde batzuetara hedatzeko, berriz, bi estrategia aipatu nahi genituzke:

A) Garapena egiten duten enpresentzako tresnen sorkuntza: i) interaktiboki komunikatzen diren bezero birtualen laguntzaileak; ii) ingurune industrialetarako entrenamendu-sistemak; iii) zerbitzu-enpresak, zeinen zerbitzuok balio erantsia jasoko luketen komunikazioa lengoia natural idatzian edo/eta ahozkoan egingo balitz; iv) testuzko materiala argitaratzera dedikatzen diren enpresak, bai komunikabideetan, argitalpenaren munduan, itzulpengintzan eta abar.

B) Sare sozialetako hizkuntza analitzatzeko tresnak sortzea, sareko informazio-fluxua maneiatzea posible egiteko. Sare sozialen erabilera gero eta handiagoak hatz-marka digital bat uzten du, geure ohiturak, interesak, iritziak, gustuak eta abar bilduz, baina hori dena analitzatzea erraldoi-lana da erreminta egokiak eskura izan ezean. Behar horrek beste bultzakada bat eman dio “datuen zientziari”.

Nazioarteratze-eredu egokiak definitu beharra dago, honako helburu hauekin: i) identifikatu zein industriarengana iritsi nahi den; ii) teknologia hauen beharra izan dezaketen beste industria batzuk bilatu; iii) integra daitezkeen aplikazio berrien gainean ikertu; iv) bazkide teknologikoak bilatu, bertan, estatuan eta nazioartean; v) teknologia hau esportatu nahi den herrialdeetako agenteekiko lankidetzak-hitzarmenak lortu.

²⁰ <http://ixa.eus/master/>

²¹ https://ec.europa.eu/programmes/erasmus-plus/opportunities-for-individuals/students/erasmus-mundus-joint-master-degrees_en

Ondorio nagusiak

Aurreko atalak sintetizatuz, hauek dira gure ustez ondorio nagusiak:

- HTen arloak garrantzi estrategikoa du hizkuntzen etorkizunari begira. Euskarak ere hor egon behar du.
- Hizkuntza-teknologiaren arloa ardatz bat izango da Europan hurrengo urteetan, eta inbertsioak egingo dira.
- Estatu Espainolean bada dagoeneko plan bat, diruz ondo hornitua. Euskal Herrian pareko plan bat beharko litzateke, Europakoarekin eta Espainiakorekin bateragarri izango litzatekeena eta euskarari garrantzi berezia emango liokeena.
- Erronka handia da, baina egokiera onean gaude erronkari heltzeko, aurretik egindako lanean eta egun ditugun agente aktiboetan (unibertsitateak, ikerkuntza-zentroak, enpresa-elkarteak...) oinarrituta.

Neurri zehatzetara eramanda, hauek dira proposatzen direnak:

- Erreferentzia-zentro bat sortzea UPV/EHUrekin eta Langune/Eiken-ekin lankidetzan. Ikerketaren eta trebakuntzan aritzeaz gain, berrikuntza eta transferentzian ere arituko da lankidetzan, bai beste ikerketa-zentroekin, baita enpresa eta erakundeekin ere. Nafarroarekin eta Iparraldearekin lankidetzan aritzeko malgutasun nahikoa beharko luke aipatutako zentroak.
- Ikerketari dagokionez, Eusko Jaurlaritzaren hurrengo zientzia-planean arlo hau lerro nagusien artean sartzea, arloko berebiziko katalizatzaile bihurtu dadin. Bestalde, arloak Europan hartuko duen dimentsioa kontuan hartuta, aukerak daude Europako deialdietan arrakastaz parte hartzeko; noski, inbertsio egokiak eginez gero aukera horiek askoz handiagoak lirateke. Horretarako, aipatutako erreferentzia-zentroak berebiziko garrantzia edukiko luke.
- Ikerketa garrantzitsua izanda ere, erronka nagusietako bat transferentzia egitera eta balio ekonomiko sortzera iristea da. Kontuan hartuz inguruko enpresak txikiak direla eta balizko lehiakideak oso handiak (Google, Oracle, SDI translations...), erosketa publiko berritzailea tresna garrantzitsu bat izan daiteke hemengo enpresak bultzatzeko. Zenbait itsasargi-proiektu indartsu abiatzea izan daiteke arlo honetarako gakoa. Ministerioko proiektuetan horrelaxe egiten dute.