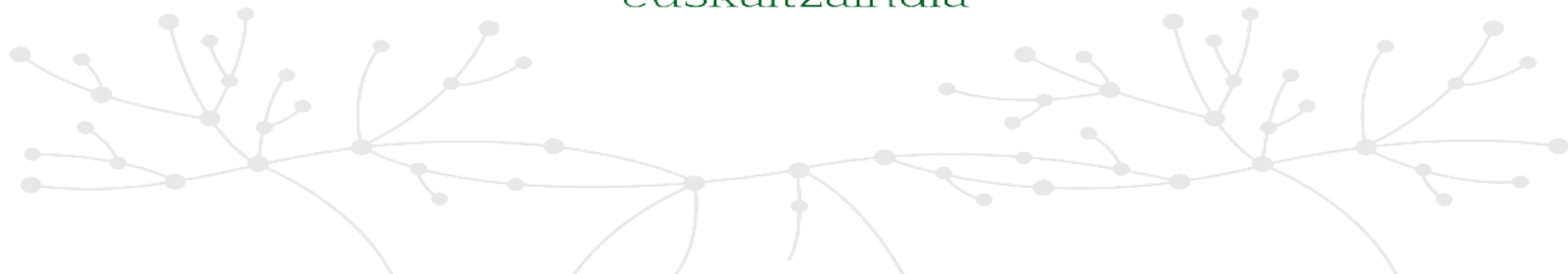




100

EUSKARA EHUNTZEN

euskaltzaindia



ARANTZAZUTIK MUNDU ZABALERA (1968-2018)
EUSKALTZAINDIAREN XVII. BILTZARRA

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



Aurkezpenaren nondik norakoa

- Hizkuntza eta teknologia: hausnarketa
- Hizkuntzaren tratamendu automatikoa: zertaz ari gara?, non gaude?
 - Ikerkuntza eta formazioa: sustatu beharra, arazoak, joerak
 - Ekimenak: zertan ari gara?, zertan ari dira erakundeak?
- Nondik jo:
 - corpusgintza
 - arauak, estandarra
 - hiztegi-gintza, terminologia
 - itzulpen automatikoa, itzulpenerako laguntzak
 - humanitate digitalak
- Ondorioak eta proposamenak

Hizkuntza eta teknologia

- ◎ Hizkuntza-teknologia gure bizitzetan sartu da oharkabeen: ordenagailuak, Internet, sare sozialak, sakelako telefonoak...
 - ahozko aginduak, itzulpen automatikoa, hizkuntza-zuzenketa, hiztegi-laguntzak, edukien bilaketa amaraunean...
 - Xuxen, EuskalBar, AhoTTS, Lexikoaren Behatokia, Elebila, Matxin, Modela...

Teknologia hori...

- oso azkar doa
- heldu samarra dago hainbat arlotan:
 - oinarriak (baliabideak, formalizazioa, tresnak) finkatuta
 - aplikazio batzuk arrakastatsu zenbait domeinu eta testuingurutan: medikuntza, administrazioa...; telefonia mugikorra...
 - itzulpen automatikoan ere emaitza nabarmen hobeak azkenaldian
- komuna izan arren, **hizkuntza bakoitza mundu bat da, formalizazio, tresna eta baliabide bereziak behar dituena**
 - Ixa Taldearen jatorrian ideia hori dago eta, besteak beste, oinarriak finkatzen aritu gara: datu-base lexikala, hizkuntzaren formalizazioa, analizatzaile morfologikoa eta sintaktikoa, semantika lexikala...

Hizkuntzaren biziraupena jokoan

- ⦿ islandiera, esaterako, “iraungitze digitalaren” arrisku bizian ikusten dute:
 - ingelesa nagusitzen ari da erabat teknologiaren erabilerari dagokionean

Icelandic language battles threat of 'digital extinction'

Iceland's mother tongue and cultural identity is drowning in an online ocean of English

- ⦿ hizkuntzaren erabilera –eta biziraupena– da jokoan dagoena: etorkizuna hemen dago eta orain da!
- ⦿ aurrera aterako bada, hizkuntza batek ezin du *tren* hori galdu
- ⦿ aurrean dugun erronka da, **baina atzean geratzeko arriskua handia da!**

Hizkuntzaren tratamendu automatikoa

◉ Idatzia

- formala
- hizkuntza informal, ez-estandarra: Whatsapp, sare sozialak (Facebook, Twitter...), foroak...
 - gazteen hizkuntza-erabilera jomugan!

◉ Ahozkoa, hizketa

- gero eta garrantzi handiagoa hartzen ari da

Euskara aro digitalean: META-NET txostena

- *Euskara Aro Digitalean – The Basque Language in the Digital Age (META-NET, 2012)*

Hernáez I., Navas E., Odriozola I., Sarasola K., Díaz de Ilarraza A., Leturia I., Díaz de Lezana A., Oihartzabal B. eta Salaberria J., 2012. META-NET White Paper Series.

- <https://labur.eus/FUD30>



META-NET txostenaren ondorio nagusiak

- ◉ Alde batetik ingelesa, eta bestetik beste guztiak
- ◉ Aztertutako 30 hizkuntzetatik, 21ek duten babes digitala “hutsa” edo, kasurik onenean, “eskasa” da
- ◉ Euskara 21 hizkuntza horien artean dago, ez azkenekoetakoa:
 - Hizketa-prozesamendua (ahozkoa): osatugabea
 - Itzulpen automatikoa: apala / ez (gaur egun: osatugabea)
 - Testu-analisia: osatugabea
 - Baliabideak: osatugabea
- ◉ Egoera baikortzat jotzen da, baina zuhur aztertu beharrekotzat

Ikerkuntza: sustatu, lagundu eta bultzatu beharra dago

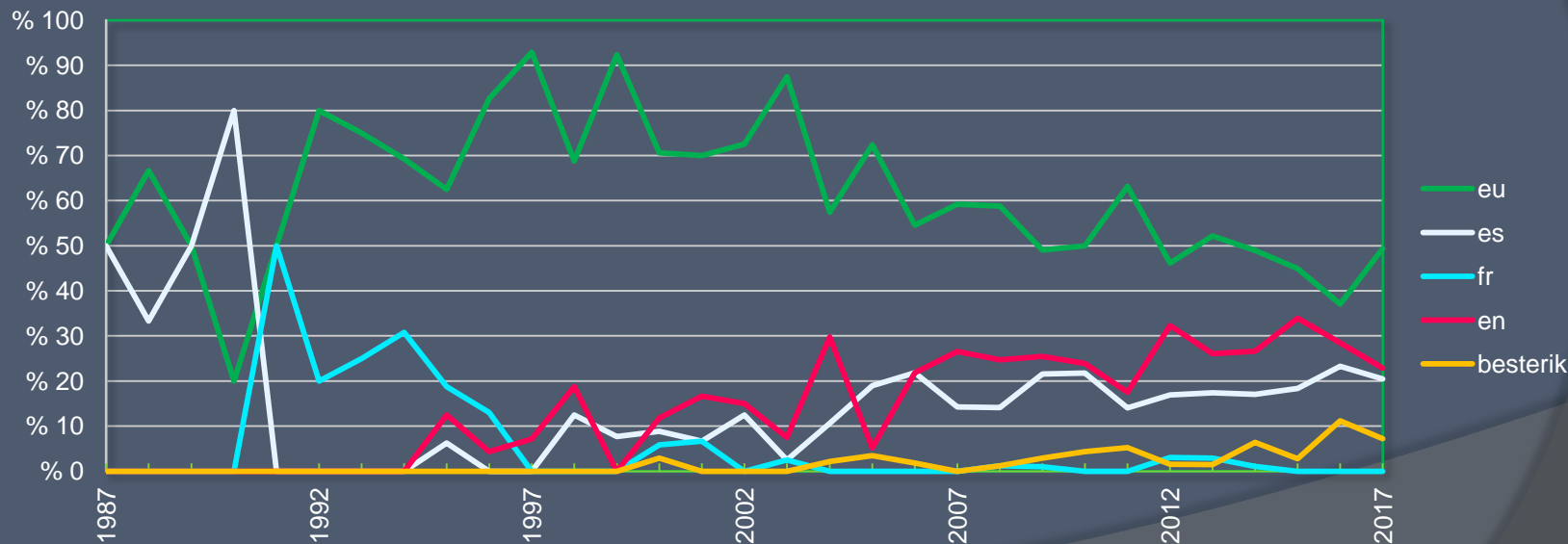
- ◎ **Diziplinartekoa**: informatikariak, hizkuntzalariak, ingeniariak, filologoak...
- ◎ **Oinarrizkoa**
 - hizkuntzaren formalizazioa
 - analisi- eta sintesi-tresnak
 - baliabideak: lexikoak, gramatikak...
 - ...
- ◎ **Aplikatua**
 - hizkuntzaren input zein output-ean laguntzeko aplikazioak
 - itzulpen-laguntzak
 - idatzian zein ahozkoan
 - ...

Ikerkuntza: arazoak

- Hizkuntza “txikiek” baliabide gutxiago; horrek ikerkuntza zailtzen du
 - metodo estatistiko eta neuronalek garrantzi handia hartu dute, corpus handiak behar dira
 - hainbat eta hainbat baliabide ingeleserako, ezen esperimentuak egitea berehalakoa baita
- Beste hizkuntzetarako egindako ahaleginak ez dira alferrik (ikasteko ere balio dute!), baina “beste hizkuntza horrek” egiten du aurrera!
- Unibertsitate “ofizialaren” bidea ez da leunena euskararen gaineko ikerketarako:
 - meritu-sistema, argitalpenen “kalitatea” neurtzeko parametroak...
 - hizkuntza “exotikooi” buruzko ikerketa “saltzea” zailagoa da
 - militantzia behar?

Ikerkuntza: joera bat?

Ixa Taldearen argitalpenetan (1.071 artikulu, 1987-2017)
ikergai diren hizkuntzak: ehunekoak, urtez urte



Formazioa: graduak, masterrak, doktoregoa

- ⦿ Hizkuntza- eta informatika-gaiak lantzen diren fakultate, zentro, gradu eta abarretan: **diziplinartekotasuna beharrezkoa da**
 - Informatika Ingeniaritza
 - Euskal Hizkuntzalaritza eta Filologia
 - Itzulpengintza eta Interpretazioa
 - Hizkuntzalaritza
 - Telekomunikazioa
 - ...
- ⦿ **56 doktorego-tesi, 126 titulatu** UPV/EHUko graduondoko titulazio eta masterretan...
 - Ez da aski dauden beharrak betetzeko, **baina hala ere ikasle euskaldunen kopurua ez da handitzen ari!**

Ekimenak: zertan ari gara?

🕒 Euskaltzaindia

- Lexikoaren Behatokia
- OEH corpus kontsultagarri bihurtzeko proiektua
- Teknologia aurreratua hiztegigintzan
- e-AEB zerbitzua, hiztegiaren eta arau-erabakien banaketa elektronikoa
- ...

🕒 HiTZ: Hizkuntza Teknologiako Zentroa (UPV/EHU)

- Ixa Taldearen eta AhoLaben ekimenez sortua, unibertsitate barruan (2018-09-27)
- Arloko ikerkuntza eta formazioa sustatzeko eta bideratzeko tresna
- Teknologia-berrikuntza eta -transferentzia



Ekimenak: zertan ari gara?

- Elhuyar, UZEI, Ametzagaiña...: hizkuntzaren tratamendurako atalak, gurekin batera hainbat proiektutan
- Zentro teknologiko eta ikerkuntza-laborategi eta -taldeak: Vicomtech, Tecnalía...
- Eta industria, oro har?
 - Langune: Hizkuntzen Industriaren alorreko Euskal Herriko enpresen elkarte (2010, 35 enpresatik gora)
 - EIKENen (Euskal Ikus-Entzunezkoen Industrien Klusterra) barne, orain
 - Oro har, baina, bada hizkuntzaren (euskararen) aldeko kezkarik? Egungo merkataritzan bezeroaren hizkuntzak duen garrantziaz jabetzen ote dira?

Edonola ere, oso komunitatea aktiboa gara, eta erreferentziazkoa beste hizkuntza minorizatuetao ikertzaileentzat

Ekimenak: zertan ari dira erakundeak?

◎ European eta Spainian ari dira mugitzen!

- Europako Parlamentuak *Language equality in the digital age* txostena onartu berri du gehiengo handiz (2018-09-11):
The EP calls "on the Commission and the Member States to develop strategies and policy action to facilitate multilingualism in the digital market; requests, in this context, that the Commission and the Member States define the minimum language resources that all European languages should possess, such as data sets, lexicons, speech records, translation memories, annotated corpora and encyclopaedic content, in order to prevent digital extinction".
- Espainiako Gobernuak *Plan de Impulso de las Tecnologías del Lenguaje* delakoa du martxan (2015/20: 90 milioi euro)
"...tiene como objetivo fomentar el desarrollo del procesamiento del lenguaje natural, la traducción automática y los sistemas conversacionales en lengua española y lenguas cooficiales."

Ekimenak: zertan ari dira erakundeak?

◎ EAEko Legebiltzarra

- 2018-04-16an aurkezpena, Kultura, Euskara eta Kirol Batzordearen aurrean: ondorio handirik ez, oraingoz

◎ Eusko Jaurlaritza

- Baditu zenbait laguntza-programa (Etortek, Elkartek...), 2020ko Agenda Digitalak ere zabaltzen ditu aukera batzuk: **euskara kulturaren bidetik beti, ezen ez industriarenetik**
- EJren *Itzultzaile automatikoa*: Hori da bidea? Ez dago hemen pareko teknologiarik? Ez litzateke hobe software librearen bidetik jotzea?

◎ Euskararen Aholku Batzordea

- 2016ko txostenean, gomendio argia

Ekimenak: zertan ari dira erakundeak?

- 2016ko txostena: *Euskarazko IKTak. Gomendioak herri-aginteentzat.*

Euskararen Aholku Batzordea (IKT Batzorde-atala)

Hizkuntza Politikarako Sailburuordetza

- <https://labur.eus/OLLLG>

- Egoera aztertu eta gomendio batzuk eman ziren; lehentasun handienekoen artean, hauxe 12.a:

Hizkuntza-teknologiak garatzen lagundu eta berariazko plana garatu.

- Gomendio hori ez da oraingoz bete



Nondik jo: corpusgintza

- Funtsezkoa, ikerkuntzaren euskarri gisa
- Erreferentziazko **Corpusa sortu** (LB abiapuntu), Euskaltzaindiaren eskutik
 - handia, behar bezala orekatua eta etiketatua, publikoa
 - ahozko atala ere barne
- Badira **hainbat ekimen** corpusgintza-bide honetan (Euskaltzaindia, UPV/EHU, Elhuyar, Ixa Taldea...), nolabait **bateratu beharko liratekeenak**
- Eduki libreak, ikerketarako trabarik ez

Nondik jo: arauak, estandarra

- ◎ Arau, gomendio eta erreferentziazko lanen informazio-sistema integratua sortu behar litzateke
- ◎ Tresnak:
 - Idazketarako laguntza integratuak:
 - Laguntza ortografiko-gramatikala: testuingurua aintzat hartuko duten zuzentzaileak, hizkera espezializatueterako egokituak eta estilo kontuak ere tratatuko dituztenak
 - Hiztegi-laguntzak
 - ...
- ◎ Hedapenerako politika: eduki libreak

Nondik jo: hiztegitza, terminologia

- **Baliabide lexikal eta terminologiko eguneratuak:** EH, EuskalTerm, OEH, hiztegi elebidunak..., esportaziorako modu estandarrek, teknologia aurreratuz hornituak (hiztegiak editatu eta argitaratzeko ingurunea...)
- **Terminologia lantzeko tresnak:** UPV/EHUren TSE programan (Terminologia Sareak Ehunduz) garatutako GARATERM (lan-ingurunea eta corpora) eta TZOS (Terminologia Zerbitzurako Online Sistema), adibidez
- **Irakurketarako laguntzak:** e-liburuetan (mugikor, tableta eta ordenagailuetan) integratutako hiztegiak

Nondik jo: itzulpen automatikoa, itzulpenerako laguntzak

- ◎ Itzulpen automatikoaren (eta aldiberekoaren) erabilpena zabaldu eta orokortu gizarteko hainbat foro eta esparrutan
 - “en castellano, para que nos entendamos todos” gisako aitzakiak nolabait neutralizatzeko
- ◎ Ahoz aritzeko aukerak ere integraturik
- ◎ Euskal Herrian bada itzulpen automatikoaren ikerketan punta-puntan ari den jenderik, ez dago kanpora joan beharrik bila
<http://modela.eus/probatu.html>

Nondik jo: humanitate digitalak

- Hizkuntzaren prozesamendua giza zientzietara zabaltzen ari da:
 - artxibo digitalak, liburutegiak
 - testu-meatzaritza eta -analisiak
 - testu historikoen tratamendua
 - hizkuntzaren irakaskuntza
 - ...
- Mugimendu handia dago arlo honetan ere
 - CLARIN, Europa mailako azpiegitura: teknologiaren erabilera humanitate eta giza zientzietan, zerbitzuak eta tresnak



Ondorioak

- ◎ Hizkuntza-teknologiaren arloak garrantzi estrategikoa du hizkuntzaren etorkizunari begira
- ◎ Hizkuntza-teknologia ardatz bat izango da European –eta Espainian– hurrengo urteetan, eta inbertsioak egingo dira
 - Euskal Herrian ere pareko plan bat behar genuke lehenbailehen
- ◎ Tren hori ez galtzeko, ikerkuntza eta trebakuntza espezializatuak bultzatze behar dira, ez da aski egiten ari garena
- ◎ Egokiera onean gaude erronkari heltzeko, aurretik egindako lanean oinarrituta; baina –sinetsi eta– mugitu beharra dago!

Zenbait proposamen

- ⦿ Erreferentzia-zentro bat sortzea unibertsitatearekin eta Langune/Eiken-ekin lankidetzan:
 - HiTZ litzateke, UPV/EHU barruan, erreferentzia-zentro horri begirako lehen urratsa
- ⦿ Arloa Eusko Jaurlaritzaren hurrengo zientzia-planaren lerro nagusien artean sartzea, estrategikotzat hartzea
 - zenbait itsasargi-proiektu indartsu diseinatzea eta garatzea, eta erosketa publiko berritzailea sustatzea
- ⦿ Euskaltzaindia ere eragile izatea, duen erreferentzialtasuna administrazioen aurrean baliatuz