



Maite Goñi:

«Ingurune digitala
estrategikoa da:
edo hor zaude edo
hizkuntzak ez du
biziraungo»





Koordinatzailea: Ander Gartzia Badiola.
Azaleko argazkia: Oier Aranzabal.
Diseinua eta maketazioa: Concetta Probanza.

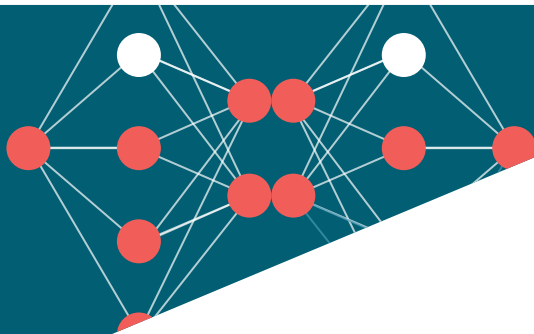




Aur kibi dea

- 4** Editoriala
- 6** Elkarrizketa: Maite Goñi
EUSKARAREN INGURUNE DIGITALA:
GOMENDIOAK 2021-2024
- 13** Euskarazko Wikipedia, euskararen sareko
egoeraren adierazle gisa
GALDER GONZALEZ LARRAÑAGA
- 18** Hizkuntza-teknologiaren egoeraren
adierazleak
XABIER ARREGI IPARRAGIRRE
OLATZ ARREGI URIARTE
- 23** Euskara Ingurune Digitalean

Edi tori ala



Euskararen Adierazle Sistema (EAS)



Hizkuntzak ez dira kalean, tabernan, etxeke egongelan edo ikastetxeke patioan soilik entzuten eta mintzatzzen. Hizkuntzak internet zabalean ere bizi dira. Bizi dira Twitterren, bizi dira whatsapp-ean, eta oharkabea bada ere, bizi dira gure telefono mugikorretako sistemetan. Sakeleko telefono baten hizkuntza konfiguratzen dugun aldiro ari gatzazkio esaten nola hitz egin behar duen gurekin, esaten ari gatzazkio aplikazioak euskaldunak ala erdaldunak nahi ditugun gure telefonoan. Mintzaira digitalaren aroan gaude murgilduta, eta itsaso zabal horretan euskara bere lekua hartu nahian dabil. Horregatik da ezinbestekoa ezagutzea euskararen egoera nolakoa den Ingurune Digitalean, nolakoa ari den izaten haren bilakaera.

Beharizan horretatik abiatuta, **Euskararen Adierazle Sistemak (EAS)** Ingurune Digitalari lotutako zortzi adierazle berri osatu ditu, eta haien jarraipena egingo du aurrerantzean ere, arlo bakoitzean euskara ematen ari den pausoak gertutik ezagutzeko. Izan ere, datuen ezaguera horri esker jakingo dugu hobeto zein den gaur gaurkoz euskararen posizioa Interneten: ondo kokatuta al dago euskara mundu digitalean? Zein da gure inguruko gaintzeko hizkuntzen egoera? Euskaldunek euskaraz egiten al dute sare sozialetan? Zer erronka ditu aurrez aurre euskarak? Galdera horiei erantzuna aurkitzeko eta gaian sakontzeko asmoarekin Ingurune Digitalean arituak diren adituengana jo dugu, eta ale monografiko hauxe da horren emaitza. Espero dugu irakurleak zenbait gako nagusi aurkitzea orrialde hauetan. 🌟

Elk arri zke ta



Maite Goñi

«Guk konfiguratuta badaukagu ordenagailua euskaraz, munduari ematen ari gatzaizkion mezua da euskaraz hitz egiten ari garela, batez ere konpainia handiei»

Eusko Jaurlaritzak «Euskararen Ingurune Digitala: gomendioak 2021-2024» dokumentua aurkeztu zuen joan den maiatzaren 21ean. Euskararen Aholku Batzordeko (euskararen erabileraren normalizazio-lanean laguntzen duen harreman organoa) Ingurune Digitalean Euskara Sustatzeko Batzorde-atal Bereziak egindako txostena da eta datozen lau urteetarako, 2021-2024 aldirako, euskararen erabilera ingurune digitalean sustatzeko bide orria izango da. Batzorde-ataleko zuzendaria da Maite Goñi.

«Gazteek gaur egun ez diote bereziki hizkuntzari erreparatzen, gai direlako hizkuntza ezberdinetan edukiak kontsumitzeko. Bai erreparatzen diote, aldiz, edukia erakargarria izateari»

Zertan datza aurkeztu duzuen bide-orria?

Aurkeztu dugun dokumentua Ingurune Digitalean Euskara Sustatzeko Batzorde-atal Bereziak egin-dako txostena da eta datozen lau urteetarako, 2021-2024 aldirako, euskararen erabilera ingurune digitalean sustatzeko 4 lan-lerro estrategikotan banatzen diren 10 gomendio eta 24 ekintza biltzen ditu.

Zertan lagundu dezake?

Bide-orriaren garrantzia da administrazioari ipar bat markatzea, guztiok egiteko tira hizkuntzaren mundu korapilatsu honetan. Administrazioari sail bakoitzak erabakiak hartzen baditu bere aldetik ziurrenik kontraesanak egongo dira. Begirada bateratu bat edukitzea garrantzitsua da, bai udaletan, bai Aldundietan, bai administrazioan, eta baita gizartean ere diru bat eskaini behar bazaio euskarari jakiteko zer gauza falta diren eta zein gauzek emango luketen bidea aurrerapauso handi bat emateko euskararen arloan, eta kasu honetan ingurune digitalaren esparruan. Zer erabaki hartu behar dira? Hemen badaude hutsune batzuk, hutsune asko daude, kontua da behar ugari daudela euskararen esparruan ingurune digitalean, baina baliabideak mugatuak dira, orduan jakin egin behar da baliabide horiek aukeratzeko. Horretan lagundu dezake, begirada bateratu bat emateko une horretan.

Zergatik dator hau orain?

Euskararen Aholku batzorde orokorraren azpian batzorde atalak daude, arlo desberdinei dagozkienak; lehenago ere erabaki zen Ingurune Digitalarentzat IKT batzordea eratzea. Hor profesional ezberdinak

gaude, begirada ezberdinekin, baina denok daukagu komunean esparru honetan arituak eta adituak garela: enpresa munduko jendea dago IKTeetan esperientzia eta eskarmentu handikoa; badago ere unibertsitateko jendea, horretan egunero aritzen direnak. Talde horri eskatu zitzaion orain bi legegaldi gomendio batzuk ematea Eusko Jaurlaritzari, norabide batean jotzeko horren aldakorra den esparruan: ingurune digitalean, hain zuzen ere. Azkenean gomendio horiek bide-orri bihurtu ziren gobernuarentzat, aintzat hartu zituzten. Orain berriro egin ditugu legegaldi honetarako, Ingurune Digitalean oso azkar aldatzen direlako beharrak. Orain erabaki dena da gomendioen sailkapenean haratago joatea eta baita sailkapen horretatik eratorritako ekintzetan ere.

Garbi daukaguna da Ingurune Digitalean izugarrizko mehatxua dagoela hizkuntza hegemonikoekin, ingelesarekin batez ere, baina baita gaztelaniarekin ere. Gazteek gaur egun ez diote bereziki hizkuntzari erreparatzen, gai direlako hizkuntza ezberdinetan edukiak kontsumitzeko. Bai erreparatzen diote, aldiz, edukia erakargarria izateari. Gazteek hartzen dituzten erabakiak garrantzitsuak dira euskararen norabiderako, ze ikusten badituzte film guztiak ingelesez, gaztelaniaz edo japonieraz —badakigu ikusten dituztela—, hor oso toki txarrean geratu daiteke euskara, desabantailarekin, ez daudelako hainbeste eduki. Horregatik garrantzitsua da norabide finko bat edukitzea, jakiteko zer erabaki hartu.



«Euskara Ingurune
Digitala: gomendioak
2021-2024»
bide-orriaren
aurkezpena,
Donostian,
San Telmo
museoan.

Lau lan lerro aipatu dituzu. Zeintzuk dira zehazki?

Alde batetik, *euskarazko eduki digitalen sustapena*: herritarrek beste hizkuntza batzuetara joko dute, eta hau geratzen ari da. Bigarrenik, *euskarazko tresna, aplikazio eta baliabide digitalen sustapena*: teknologiak euskaraz hitz egitea eta guk teknologia-rekin euskaraz hitz egin ahal izatea garrantzitsua da. Hirugarrenik *euskararen eta euskarazko baliabideen ikusgarritasuna eta zabalkundea*. Eta azkenik, *zaintza teknologikoa ziurtatzea*: ezin gara horretan geratu, hurrengo bide orria pentsatzen hasi behar gara. Ez dugu ikusi behar bakarrik beste herrialde batzuetan, hizkuntza gutxituak erreferentziatzen hartuta, zer egin ari diren; interesatzen zaigu ere hizkuntza hegemonikoetan teknologiaren inguruan zertan ari diren. Baina ez bakarrik hizkuntzan, baizik eta teknologian ere bai. Horregatik da euskararentzat garrantzitsua zaintza hori.

Lan-lerro horietan zentratu zarete gomendioak emateko.

Lau lerro horiek oso garrantzitsuak dira, eta gu saiatu gara horren barruan lan-lerro bakoitzerako gomendio batzuk ematen. Horrela hamar gomen-

dio atera zaizkigu, baina ez gara horretan geratu. Gomendioak ematea erraza da; guk ekintzetara iritsi nahi izan dugu. Nahi izan dugu zehaztu ekintzak zergatik egin behar diren eta nola egin behar diren. Kontuan hartu behar da ekintza horiek malgutasunez hartu behar direla. Hausnarketa egin genuen momentuan ekintza horiek interesgarriak izan zitezkeen, baina gerta daiteke hilabete batzuk pasatu eta helburu horietara iristeko beste formula batzuk edo aldaera batzuk aurkitzea, interesgarriagoak izan daitezkeenak.

Premia handia al dago Ingurune Digitalean euskararen presentzia sustatzeko? Une erabakiorrean al gaude?

Estrategikoa da. Estrategia hitza erabiliko nuke. Ingurune digitala estrategikoa da une honetan: lehen hautua zen ea ingurune digitalean egotea beharrezkoa zen, eta teknologiarekin lan egitea; gaur egun ez da hautu bat: edo hor zaude edo hizkuntzak ez du biziraungo. Baina ez du biziraungo ez ingurune digitalean, ez beste inongo esparrutan ere, dena ari baita bihurtzen ingurune digitala. Beraz pentsa zenbateraino den estrategikoa ingurune hori.



Adibidez, gaur egun ia denok egiten dugu lan gehiago edo gutxiago mahai gaineko ordenagailuarekin edo gailu mugikorrekin, eta hor barruan erabiltzen dugunak asko markatuko du. Adibide bat: guk konfiguratuta badaukagu ordenagailua euskaraz munduari ematen ari gatzazkion mezua da euskaraz hitz egiten ari garela, batez ere konpainia handiei; konpainia handiek, hortaz, badakite gure aparatua euskaraz ari den, gaztelaniaz ari den, frantsesez ari den... eta kopuru horiek aintzat hartzen dituzte beraien produktuak garatzeko; beraz indarra gal dezakegu plataforma ezberdinetan euskararen presentzia aldarrikatzen dugunean gero guk ordenagailuak gaztelaniaz konfiguratuta baditugu, mezu kontraesankorra izan daitekeelako. Euskaraldian helburua da lehen hitza euskaraz egitea, bada ordenagailuetan edo ingurune digitalean horren parekoa izango litzateke konfigurazio hori euskaraz edukitzea. Konfigurazio horrek, neurri handi batean, markatzen die konpainia handiei zein nahi eta interes dagoen hizkuntza horrekiko. Guk ez badugu esaten interesatzen zaigula, konpainia handiek zer eskainiko digute bada? Gauza bat da administrazioak egin dezakeena, baina gero beste gauza bat da norbanakook ere zer egin dezakegun esateko ingurune digitalean ere euskaraz egin nahi dugula.

Hau da, zuek gomendatzen duzuen da IKTek euskaraz egitea.

Gomendio horrekin bilatzen dena da administrazioak gutxienez bere baitan dagoen hautua edo erabakia hartzea eta ordenagailuak euskaraz kon-

«Inori ez dio ihes egiten, euskararen itsasoan, olatuak zakar samar datozela azken boladan. Hori dela eta, guztiok norabide berean arraun egitea komeni da. Beharrak ugariak dira; baliabideak, berriz, urriak. Guztiok gara aliatuak eginkizun honetan»

figuratzea. Horrela "ni euskaraz ari naiz" mezua zabaltzen arituko ginateke. Alegia, esaten arituko ginateke badagoela komunitate bat edo gogo bat.

Bai gomendioak, bai ekintza horiek zehazteko orduan, adostasuna errez lortu al da taldean? Guztiok ikusten al dute egoera modu bere-tsuari?

Kontsentsuz hartu dira erabaki guztiak, baina informatu egin gara lehendabizi: txostenak enkargatu ditugu eta egoera ezberdinak aztertu ditugu. Aditu eta aritu ugari ekarri ditugu batzordera, eta esparru ezberdinetako jendea pasatu da. Jardunaldi bat ere antolatu genuen Hizkuntza Teknologien inguruan, ikusteko beste herrialdeetan zer egiten ari ziren. Horiek denak mahai gainean edukita joan gara erakitzen bide-orri hau, eta erabat kontsentsuz hartu dira erabakiak.

Gomendio hauek administrazio publikoari zuzenduta daude ala eremu publikotik haratago doaz?

Ekintza-proposamen horiek guztiak Eusko Jaurlaritzari emandakoak edo zuzendutakoak dira nagusiki, hori baita EABren Batzorde-atalak duen eginkizuna. Halere, ezin dugu ahaztu ingurune digitalak ez daukela muga fisikorik. Horregatik, Euskal Herriko erakunde publiko zein pribatu guztien esku utzi nahi izan ditugu gomendio hauek, bakoitzak bere eskumen eta ahalmenen barruan jardun dezan. Inori ez dio ihes egiten, euskararen itsasoan, olatuak zakar samar datozela azken boladan. Hori dela eta, guztiok norabide berean arraun egitea komeni da. Beharrak ugariak dira; baliabideak, berriz, urriak. Guztiok gara aliatuak eginkizun honetan.

Gomendioez gain bestelakorik aipatu da?

Badira ingurune digitaleko euskararen sustapenean estrategikoak eta zeharkakoak iruditzen zaizkigun bi gai eta horiek ere jaso ditugu dokumentuan. Batetik, burujabetza teknologikoa eta, bestetik, software eta lizentzia libreak.

Burujabetza teknologikoaren gaia aurretik ere aipatu izan da. Jakina da informazioaren eta komunikazioaren teknologiak eta azpiegiturak, datuak biltzeko eta gordetzeko erabiltzen diren teknologiak, erabiltzaileak dauden herrialdeko legeekin, beharrek eta interesekin bat etortzea garrantzitsua dela; izan ere, erabiltzaileei buruzko informazioa, besteak beste, ez litzateke geratu behar erabiltzaileek kontrola ezin ditzaketen eragileen esku. Edono-

la ere, esan beharra dago gai honek gainditu egiten duela euskararen sustapena ingurune digitalean, haratago doala; baina zalantzarik ez dago esparru teknologikoan euskararen erabilerari ere eragiten diola, teknologia burujabe hauek, enpresa teknologiko handienak ez bezala, euskaraz egoteko aukera bermatzen dutelako. Hortaz, interesgarria litzateke estrategia honen inguruko eztabaida zabaltzea.

Software librea eta lizentzia librean sustapenari dagokienez, oraingoan ez dugu gomendio zehatz bihurtu, iruditzen baitzaigu sustatzen diren egitasmo guztietan kontuan izatea komeni dela. Garrantzitsua da azpimarratzea software librea eta lizentzia libreak sustatuta bidea zabaltzen zaiela euskarazko eduki, baliabide, aplikazio eta tresnei. Software libreak eta lizentzia libreek euskararen normalizazioan izan dezaketen eragina handia da, horiek euskaraz ipin daitezkeelako. Bestelako software eta lizentzien kasuan, aldiz, euskararen presentzia ez dago bermatuta.

Ekintzen artean badago bat EAS Euskararen Adierazle Sistemarekin lotura izan dezakeena, 10.1 puntua: *Ingurune Digitalean euskararen inguruko datuen neurketa sistema antolatu: adierazleak identifikatu eta sistematizatzeke lana egin.*

Halaxe da, zalantzarik gabe. Edozein esparrutan, ez bakarrik euskarari dagokionean, ezinbestekoa da datuetan oinarritzea politika eraginkorrek egin ahal izateko. Horregatik, bide-orrian toki garrantzitsua du Ingurune Digitalean euskararen inguruko datuen

neurketak eta haren jarraipenak. Eusko Jaurlaritzak hainbat lan gauzatu ditu azken urteotan arlo honi dagokionez. Edonola ere, gomendio honekin lan-leerro honen garrantzia azpimarratu nahi izan dugu, jarraipena izan dezan. Horrenbestez, Ingurune Digitala ikertzeko adierazleak definitzea eta horiek Euskararen Adierazle Sisteman (EAS) integratzea, neurketa sistematizatu bat egin ahal izateko, ezinbestekoa da datu kontrastatuetan oinarritutako erabakiak hartu ahal izateko.

Zer garrantzia dauka datu horiek lortzeak eta sistematizatzeak? Zer zailtasun daude gisa horretako datuak eskuratzeko?

Ingurune digitalean zer, nola eta zergatik kontsumitzen den, alegia, erabilera ohiturak, kopuruak, joerak eta arrazoiak zeintzuk diren jakiteko gai bagara, sustapen bideak datu horietan oinarritzeko eta begirada behar handien dagoen arloetan beren-beregi ipintzeko aukera egongo da. Aurrerapauso nabarmena izango dela iruditzen zaigu. Edonola ere, ez da erraza adierazle horiek zehaztea. Alde batetik, identifikatutako adierazle batzuk ezin direlako neurtu eta, bestetik, etengabe eta abiada bizian aldatzen doan ingurune honetan, ez delako erraza jarraipena izango duten adierazleak topatzea. 🌟

«Ingurune Digitala ikertzeko adierazleak definitzea eta horiek Euskararen Adierazle Sisteman (EAS) integratzea, neurketa sistematizatu bat egin ahal izateko, ezinbestekoa da datu kontrastatuetan oinarritutako erabakiak hartu ahal izateko»







Euskarazko Wikipedia, euskararen sareko egoeraren adierazle gisa



Galder Gonzalez Larrañaga

Euskal Wikilarien Kultura Elkarteko programa arduraduna

2001ean Wikipedia sortu zenetik, nabarmen aldatu da Interneten erabilera, presentzia eta izaera. Hogei urte igaro diren honetan, Internet leku guztietan dago, baita ia herritar guztion poltsikoetan dauden telefono adimentsuetan ere. Hasierako norabide bakarreko sareari (1.0 weba zena) aldebiko komunikazioa gehitu zitzaion 2.0 sarearekin. Sare sozialak, bideo sorrera plataformak, blogosferak eta wikiak gehitu zitzaizkion sareko ekosistemari. Wikipedia ere testuinguru horretan jaio zen, eta hasieran edukiaren kalitatearen inguruko kezka handia bazen ere, denborarekin kontrako bidea hartu du, informazio egonkorra, fidagarria eta gaurkotuaren sinonimo gisa agertuz.

Wikipediari erreparatzea garrantzitsua da hizkuntza baten sareko egoera zein den hobeto jakin nahi badugu. Alde batetik, hizkuntza horren presentzia zein den ezagutu dezakegulako, neurketa mota askorekin; bestetik, sareko ekosisteman dauden beste eragile batzuek, tartean Google bezalako enpresa handiek, Wikipedia erabiltzen dutelako euren zerbitzuen hizkuntza zehazteko; azkenik, komunitate batek egiten duelako Wikipedia eta, beraz, komunitate horrek ere esaten digulako zein den gure hizkuntzaren egoera.

Wikipedia eta hizkuntzen osasuna

Wikipedia 300 hizkuntza baino gehiagotan existitzen da, eta handiagoa den Wikimedia familiaren parte da. Familia osoak ehunka hizkuntza hitz egiten ditu, hamarnaka alfabetotan. Helburua munduko ezagutza guztia biltzea da, gure kasuan, euskaraz. András Kornai hungariar hizkuntzalari matematikoak ikertu du hizkuntza batek XXI. Mendea bukatu aurretik bizirik irauteko behar dituenak, eta Wikipedia bezalako plataforman egotea funtsezkoa dela ondorioztatu du. Hizkuntza batentzat, modu librean eta atzigarrian sortzen diren edukiak egotea nahitaezkoa da, beraz.

Euskarazko Wikipedia 2001ean sortu zen, eta hazkuntza geldoa izan zuen 2006ra arte. Hortik aurrera etengabe hazi da, bai artikuluko kopuruan, editore kopuruan zein kalitatean. Hezkuntzan erabiler zabalekoa da, eta ikasleek euren egunerokoan ikasitakoa osatzeko erabiltzen dute. Boluntarioek

egindako lana denez, zaila da ziurtatzen gaztetxoek behar duten edukia txukuna dela; horregatik, Eusko Jaurlaritzaren laguntzarekin, «2.000 artikuluko 12-16 urteko ikasleentzat» proiektua jarri genuen martxoan 2017an, oinarrizko edukien kalitatea igotzeko.

Hezkuntza Programa horren baitan 5.000 ikasle baino gehiagok hartu dute parte, 6 milioi hitz baino gehiago idatzi dituzte (euskarazko Wikipediaren %10) eta 4.820 artikuluko hobetu dira, Txikipedian egindakoa kontuan hartu gabe. DBHko ikasleentzat interesgarri diren 2.000 artikuluko horiek %155,4 handitu dira byte kopuruan eta dagoeneko zerrenda horretako 1.308 artikuluko ditugu 5.000 byte baino gehiagorekin (%65,4).

Artikulu kopurua neurgailu gisa

Wikipediako unitate minimoa artikulua da. Artikulu kopuruak, bere horretan, ez du gehiegi esan nahi, artikuluko asko modu masiboan sor daitezkeelako robotak erabilia. Horregatik, beste neurketa batzuekin alderatu behar da, Wikipedia osasuntsu dagoen edo ez jakiteko. Euskara 321 hizkuntzen artean 32.a da artikuluko kopuruan¹ (2021eko ekainaren 7an), hau da, artikuluko gehien dutenen %10ean sartzen da. Une honetan 375.619 artikuluko daude, eta turkieraren (406.413) eta errumanieraren (420.612) atzetik dago. Hizkuntza eraikiak eta hilak kontuan hartu gabe, 8. hizkuntza gara hiztun kopuruarekiko artikuluetan,

1. https://meta.wikimedia.org/wiki/List_of_Wikipedias

1. irudia. Hilabetero sortutako artikuluko berriak euskarazko Wikipedian



Iturria: <https://stats.wikimedia.org/#/eu.wikipedia.org>

eta bigarrena 100.000 artikulutik gora duten hizkuntza natural bizien artean². 521,8 artikuluko ditugu 1.000 hiztuneko (720.000 euskaldun alfabetatu dau dela kalkulatu).

Hala ere, artikuluko modu automatikoan sortzea posible denez, ez da egin behar den neurketa bakarra. Azken 12 hilabeteetan, hala ere, ez da artikuluen sorrera automatikorik egin, eta ikus daiteke hazkuntza orekatuagoa dela, 1.000 artikuluko berri inguru hilabeteetan³.

2. https://meta.wikimedia.org/wiki/List_of_Wikipedias_by_speakers_per_article

3. Grafika zein datuak <https://stats.wikimedia.org/#/eu.wikipedia.org> webgunean kontsulta daitezke.

Sakonera adierazle gisa

Artikulu kopurua erraz puztu daitekeen elementua denez, artikuluko horien sakonera neurtzea izan daiteke adierazle fidagarriago bat. Hainbat modu daude hau neurtzeko, horietako bat funtsezko artikuluen sakonera da. Funtsezko artikuluko Wikipedia guztiek izan beharreko 1.000 artikuluko dira. Zerrenda hau elkarlanean eraikitzen da, nazioartean garrantzia duten gaiak erabaki ahal izateko. Ondoren, artikuluen sakonera neurtzen da, dituzten byte kopuruen arabera, hizkuntza bakoitzak gauza berbera esateko behar dituen letra kopurua kontuan hartuta. Eraitza ehunekoan ematen da. Katalana lehenengoa den zerrenda horretan, euskara 39.a da, %43,64rekin. Azken hiru urteotan euskara da

gehien hazi den hizkuntza. Batezbestekoa 23.815 byte da (35.a) eta mediana 13.414 byte (35.a)⁴. Ariketa berbera 10.000 artikulurik garrantzitsuenen zerrendarekin eginda, 32. postuan gaude, %50,73 beteta.⁵

Artikulu guztietan dauden hitz kopurua ere neurtu daiteke, eboluzioa ikusteko. Une honetan Euskarazko Wikipediak 62.227.120 hitz ditu, artikulu bakoitzeko 165,6 hitz. Wikipedia guztien batezbestekoa 3576 da, beraz horren azpitik gaude.

Bisita kopurua neurgailu gisa

Bisita kopurua neurtzea zaila da, Wikipediak ez duelako bere erabiltzaileen estatistikarik egiten. Kargatutako orrialdeak ikus daitezke, azken urtean 144 milioi (12 milioi hilabeteetan, 394.000 egunean)⁶. Estatistika honetan artikulua zein artikulu ez direnak (bilaketak, kategoriak, azal nagusia...) kontuan hartzen dira.

2. irudia. Bisitatutako orrialde guztiak, artikulua eta orrialde bereziak barne, hilabetero, euskaraz



Iturria: <https://stats.wikimedia.org/#/eu.wikipedia.org/reading/total-page-views/normal|bar|2020-05-31~2021-09-30|~total|monthly>

4. https://meta.wikimedia.org/wiki/List_of_Wikipedias_by_sample_of_articles

5. https://meta.wikimedia.org/wiki/List_of_Wikipedias_by_expanded_sample_of_articles

6. <https://stats.wikimedia.org/#/eu.wikipedia.org/reading/total-page-views/normal|bar|2020-05-01~2021-07-01|~total|monthly>

Mugikor bidez egindako sarrerak fidagarriagoak dira neurgailu gisa (bisita automatizatuak ez direlako ohikoak). Kasu horretan 23 milioi orrialde zerbitzatu dira urte batean, ia 2 milioi hilabeteetan, 63.000 baino gehiago egunean).

Hala ere, bisita kopuru hau ere mugikorren eta bilaketen hizkuntzaren arabera egin ohi dira. *Herri txiki, infernu handi* saioan agertzen diren herrien bisitak alderatzen badira, oro har, gaztelaniaz 6 aldiz bisita gehiago dituzte euskaraz baino herri horiek emisio egunean. Gailuen hizkuntza hautuak eragina izango du bisita kopuruan, beraz.

Editore kopurua neurgailu gisa

Euskarazko Wikipediak 211 editore aktibo izan ditu (5 edizio gutxienez) bataz beste azken 12 hilabeteotan. 2021eko apirilean inoizko kopururik handiena lortu zen, 418.

Editore kopurua, hiztun kopurua eta Barne Produktu Gordina alderatu daitezke, The Economistek egin zuen bezala. BPGa Gaindegiaren datuetan oinarrituta egin da (hiztunen BGPa kalkulatzea ia ezinezkoa den ariketa da).

Euskarak 293 editore aktibo ditu milioi hiztuneko (72.000 hiztunekin kalkulatuta). Euskal Herriaren BPGa Kataluniarena baino zertxobait handiagoa da, eta hiztun kopurua ordezkatzeko duen zirkulua islandierarena halako bi.

Ondorioak

Wikipediak aukera asko ematen ditu hizkuntza baten sareko egoera neurtzeko, artikuluko kopuruaz haratago. Wikipediaren kalitatea neur daiteke artikuluko sorta bat hartuta, edo artikuluen luzera kalkulatuta; posible da komunitatea bera neurtzea, parte-hartzaile kopurua ikusita; eta bisita kopuruak ere aukera ematen digu komunitate horrek egiten duen lanaren ikusgarritasuna neurtzeko, zein euskarazko bilaketa ohituretan dauden aldaketak ikusteko.

Euskarazko Wikipedia proiektu bizia da, bere antzeko egoera soziolinguistikoan dauden hizkuntza asko baino osasuntsuago, baina erronka handiak ditu aurretik, bereziki kalitatea handitzeko bidean eta irakurleak lortzeari begira. 🌟



Hizkuntza- teknologiaren egoeraren adierazleak

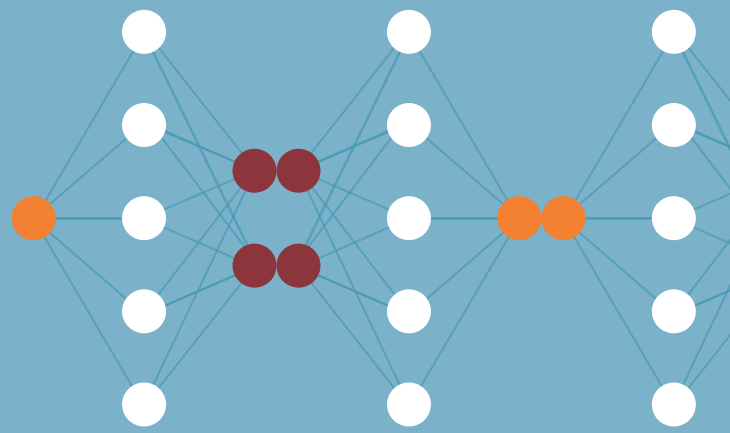


Xabier Arregi Iparragirre
Olatz Arregi Uriarte

Ixa taldea, Hitz zentroa, UPV/EHU



Hizketan aritu gaitzke gaur egun era askotako gailuekin. «Zein da tenperatura?» galdetuko diogu autoari, «jarri albistek» aginduko telebistari, «deitu urliari» eskatuko smartphoneari, «nor da Katalin Karikó?» itaunduko diogu etxeko bozgorailuari. Trepeta horiek denak solaskide ditugu.



Alexa, Siri edo horrelako beste laguntzaile birtualek gaitasun handia dute. Solasaldiak jarraitzeaz gain, gai dira audioa erreproduzitzeko, etxeko gailuak kontrolatzeko, erosketak egiteko edota posta elektronikoa kudeatzeko. Bestalde, idatzizko testueta-rako ere, badira hainbat sistema laguntza eskaintzen digutenak, zuzentzaile automatikoak, itzultzaileak... Baina, tresna eta aplikazio horiek guztiek nola funtzionatzen dute? Nola "ulertzen" dute hizkuntza? Eta zergatik dira gai gure galderak erantzuteko?

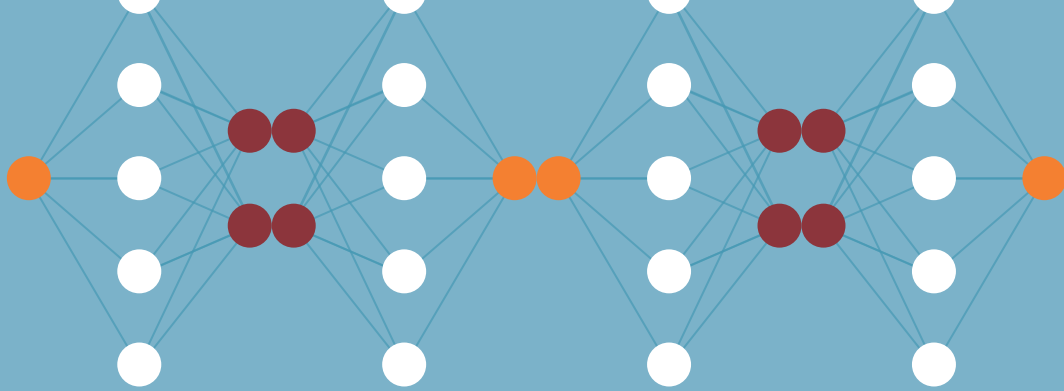
Horren guztiaren atzean hizkuntza-teknologiak daude, hau da, testuak eta audioak automatikoki lantzeko dauden programak, tresnak eta aplikazioak.

Digitalizazioaren garapenaren ondorioz eskura dugun testu- eta ahozko informazio izugarria automatikoki ustiatzeko aukera ematen dute hizkuntza-teknologiek, eta horiek gabe, ezinezkoa litzateke datu horiek guztiak tratatzea. Teknologia horien parte dira, hizkuntza naturalaren prozesamendua, itzulpen automatikoa edota elkarrizketa sistemak. Ez dira nahastu behar hizkuntza-teknologiak eta IKTak, Informazioaren eta komunikazioaren teknologiak. Azken horiek, informazioaren tratamenduan eta transmisioan erabilitako gailuak eta teknikak biltzen dituzte. Funtsean, IKTak makinaren eta pertsonen arteko komunikazioa ahalbidetzen duten baliabide informatikoak dira, adibidez, arbel digital interaktiboa, multimedia baliabideak, merkataritza elektronikoko plataformak, eta abar. Horietan guztietan ez du zertan hizkuntzaren tratamendua egon behar.

Oinarrian hizkuntza-teknologiak dituzten hainbat tresna erabiltzen ditugu gure egunerokoan, eta beste hainbeste erabiliko ditugu datozen urteetan arloak hartu duen garrantzia kontuan izanik. Dagoeneko aipatu diren laguntzaile birtualez edota itzulpen automatikoaz gain, badira hizkuntza-teknologiek lagundu dezaketen hainbat arlo, izan administrazioan, osasungintzan, hezkuntzan edota aisialdian.

Osasungintza arloari erreparatuta adibidez, badira egingarriak diren hainbat ataza onura handiak ekarriko lizkiokeenak gizarte osoari. Mediku-konsultetan sortzen diren osasun-txostenak modu automatikoan prozesatzeko gai gara egun dagoen teknologiarekin. Txosten horietan dagoen informazioa ondo erauzten bada, hau da, zein diren sintomak, zein gaixotasuna, zer-nolako tratamendua eman zaion eta zein izan den pazientearen eboluzioa, datu horiekin guztiekin medikuek tratamendu egokiagoak aplikatu ditzakete besteen esperientzian oinarrituta, modu eroso eta "erraz" batean. Arlo honekin jarraituta, mediku-txostenaren laburpen sinplifikatua automatikoki egitea ere gaur egun egingarria da. Horrek, pazienteari informazio ulergarria emango lioke zehaztasunik galdu gabe. Gainera, laburpen hauek ahots automatikoaren bidez ere jasotzeko gaiak, modu honetara, irakurtzeko zailtasunak dituenak ere ez luke arazorik izango informazio hori eskuratzeko.

Antzeko adibidea justizian topa dezakegu. Osasun-txostenen ordez, epai-txostenak ditugu he-



men gai nagusi, eta gaixotasunen ordez, delituak; tratamenduen tokian, epaiak eta abar. Delitu berdinen gainean, ahal dela, epai bera ematea errazagoa litzateke horrela.

Turismo alorrera jauzi eginez, asko dira hizkuntza-teknologiaren laguntzarekin burutu daitezkeen atazak. Txatbot eleaniztunak, turisten iritziak prozesatzeko eta informazio baliagarria ateratzeko tresnak,

garraioarekin lotutako lokalizazio tresna adimen-duak...

Adibideak asko eta asko dira, baina, esan beharra dago, hemen planteatutakoak egiteko teknologia egon arren, hau guztia ingelesa ez den hizkuntzetarako landu beharreko gaia dela. Corpus espezializatuak eta hizkuntza zehatzerako entrenatutako tresnak edo aplikazioak ezinbestekoak dira.

Adierazleak

Adimen artifizialaren ezta edo bete-betean eragin die hizkuntza-teknologiei. Gaur egun, hizkuntza prozesatzen duten teknologia neurona-sareetan oinarritzen dira. Hurbilpen horrek ezinbestean eskatzen ditu datu-multzo handiak, horietaz baliatzen baitira sistema neuronalak entrenatzeko eta ikasteko.

Hori horrela, euskararen euskarri teknologikoa zertan den jakiteko, lehenik oinarritzko azpiegiturari, baliabideei eta tresnei erreparatu behar zaie.

Horiekin batera, hizkuntza-teknologietan oinarritutako aplikazioen egoera ere aztertu beharrekoa da. Jakin nahi dugu, batetik, munduan erabiltzen diren aplikazioetatik zein ditugun erabilgarri euskaraz, eta, bestetik, egiatan zenbateko erabilera duten gurean.

Arlo honetako euskararen egoera ezin daiteke ingelesarekin edo munduko hizkuntza boteretsuenekin konparatu, ez behintzat berdinen arteko konparazioa balitz bezala. Hala ere, komeni da baliabide urriko beste hizkuntza batzuekin alderatzea, erkaketa horrek adieraziko baitu zein den euskararen egoera erlatiboa, eta zein diren gure ahalak eta ahuleziak besteen aldean.

Bada kontuan hartzeko beste faktore bat adierazleen jarraipena egitean: arlo honen dinamikotasuna. Bilakaera azkarreko teknologiaz ari gara, eta horrek eskatuko du adierazleak maiz errebisatzea.

Horrela bada, euskararen teknologien adierazle-sistema diseinatzean, honako bost atalei erreparatzea proposatzen dugu:

1.

Oinarrizko baliabideen eskaintza

Atal honetan ikusi behar da zenbat baliabide eta tresna eskaintzen diren horretara bideratutako biltegi edo gordailuetan.

Askotariko baliabide horien artean, esan dugunez, interes berezia dute corpusek eta datu-multzoek, horietan oinarritzen baita sistema neuronalen ikasketa-prozesua. Horiekin batera, oinarrizko hizkuntza analisia eta prozesamendua egiteko bestelako baliabideak ere sail honetan sartzen dira: hizkuntza-ereduak, ahotsaren analizatzaileak eta sintetizatzaileak, galderak erantzuteko sistemak...

Hizkuntza diogunean, ahotsezko nahiz idatzizko hizkuntzaz ari gara. Hau da, hizketaren eta testuaren tratamendurako azpiegituren eskaintza neurtu nahi da.

Badaude hizkuntza baliabideen eta tresnen nazioarteko gordailu homologatuak, adibidez, *huggingface*¹, *CLARIN*², *ELG*³ edo *ELRA*⁴, eta Euskal Herrian ere badugu Eusko Jaurlaritzaren Euskararen tresnak⁵ biltegia. Horietan oinarrituta euskararen (beste hizkuntzekiko) presentzia erlatiboa eskura daiteke. Planteamendua ez da, beraz, baliabideen katalogo edo zerrenda berri bat osatzea, baizik eta daudenak baliatzea euskararen presentzia erlatiboki kuantifikatzeko.

1. Hugging Face: <https://huggingface.co/>
2. Common Language Resources and Technology Infrastructure: <https://www.clarin.eu/>
3. European Language Grid: <https://www.european-language-grid.eu/>
4. European Language Resources Association: <http://www.elra.info/>
5. <https://www.euskadi.eus/eusko-jaurlaritza/euskara-tresnak/>

2.

Oinarrizko baliabideen erabilera

Eskainitako baliabideek eta tresnek zenbateko erabilera duten jakitea ere beharrezkoa da. Aipatu ditugun gordailuen bidez izango genuke aukera informazio hori eskuratzeko, adibidez, zenbat deskarga izan dituzten jakinez gero. Gainera, kalitatea neurtzeko gabezia nolabait arintzeko ere baliagarria izan daiteke erabileraren neurria: asko erabiltzen dena, on den seinale.

3.

Hizkuntza-teknologietan oinarritutako aplikazioen eskaintza

Nazioartean arrakasta duten aplikazioen artean, euskaraz erabilgarri daudenak zein diren jakitea adierazle esanguratsua da. Lehenbizi aplikazio erreferentziakoen zerrenda hori zehaztu egin beharko litzateke, eta ondoren ikusi horietatik zein dauden euskaraz. Esate baterako, itzultzaile automatikoak, txatbotak edo elkarrizketa-sistemak, eta laguntzaile birtualak edo bozgorailu adimendunak (Amazon Echo Alexa edo Google Home bezalakoak) erreferentziakoak dira gaur egun.

Horietako bakoitzaren hainbat aldaera edo azpi-produktu bereiz daitezke. Esate baterako, itzultzaile automatikoen artean: ahotsetik ahotserakoak, aldiberekoak, euskalkietarako eta hizkuntza-erregistroetarako doituak, domeinuetara egokituak, eta abar.

4.

Hizkuntza-teknologietan oinarritutako aplikazioen erabilera

Dagoeneko euskaraz erabiltzen diren aplikazioetan, komeniko litzateke erabileraren neurria ere jasotzea. Informazio hori eskuratzeko, bertako enpresen produktuei begiratzea izango da modurik errazena.

Euskal Herrian bertan sortutako itzultzaile automatikoak baditugunez, esate baterako, elia¹, itzuli², eta batua³, aukera egokia da itzultzaile automatikoen erabilera neurtzea. Galdera da: Zenbat testu-hitz itzultzen dituzte urtero euskaratik edo euskarara itzultzaile automatikoek? Oso interesgarria litzateke hori jakitea, eta posible da.

Gainera, puntu honetan euskarak zein rol jokatzen duen jakitea ere oso interesgarria da; alegia, zenbatetan den jatorri-hizkuntza eta zenbatetan xede-hizkuntza. Hein batean, tasun horrek adieraz dezake zein den sortze-hizkuntza, eta euskararen erabileran teknologiak duen eragina argitzeko ere lagungarria izan daiteke.

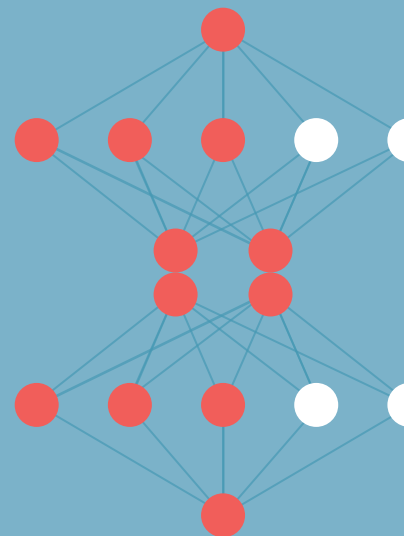
1. elia (<https://elia.eus/itzultzailea>)
2. itzuli (<https://www.euskadi.eus/itzuli/>)
3. batua (<https://www.batua.eus/>)

5.

Alorreko enpresa ahaltzuen eskaintzan, euskarak duen estatusa

Enpresa ahaltzu horiek erraldoi teknologikoak dira. Munduko ehunka hizkuntza tratatzeko gaitasuna dute, eta oso adierazgarria da beren lehentasunetan agertzea, horrek ere hizkuntza baten presentzia teknologikoaren neurria ematen baitu. Google enpresa har genezake adibide gisa, eta aztertu zein den euskararen estatusa Googlek duen eskaintza zabalean (bilatzailea, drive-ko hizkuntza-zerbitzuak, ahotsa tratatzeko aukerak...). Horietako zenbat eskaintzen dira euskaraz? Horietako zenbatetan dago euskara tratatzeko tresnaren bat...

Hizkuntza-teknologien alorrean euskararen egoera zein den jakiteko adierazleak proposatu dira. Ez da baina, adierazle horiek nola neurtu beharko liratekeen zehaztu. Horretarako, aipatu diren baliabide zein tresna bakoitzeko, neurri zehatzak erabaki beharko lirateke eta horiek lortzeko zein izan daitekeen modu egokiena zehaztu. Neurri horiek, kantitateari ez ezik, kalitateari ere erreparatu beharko liokete euskararen egoera teknologikoaren argazki fidagarria lortu nahi bada. ★



Euskara Ingurune Digitalean



Interneten euskarak duen presentzia neurtzeko eta euskal hiztunen komunitatea ezagutzeko asmoz, EAS Euskararen Adierazle Sistemak hainbat adierazle berri sortu ditu 2020. urtean. Hurrengo orrialdeetan dituzuen lau grafikoetako datuak dira horren lekuko:

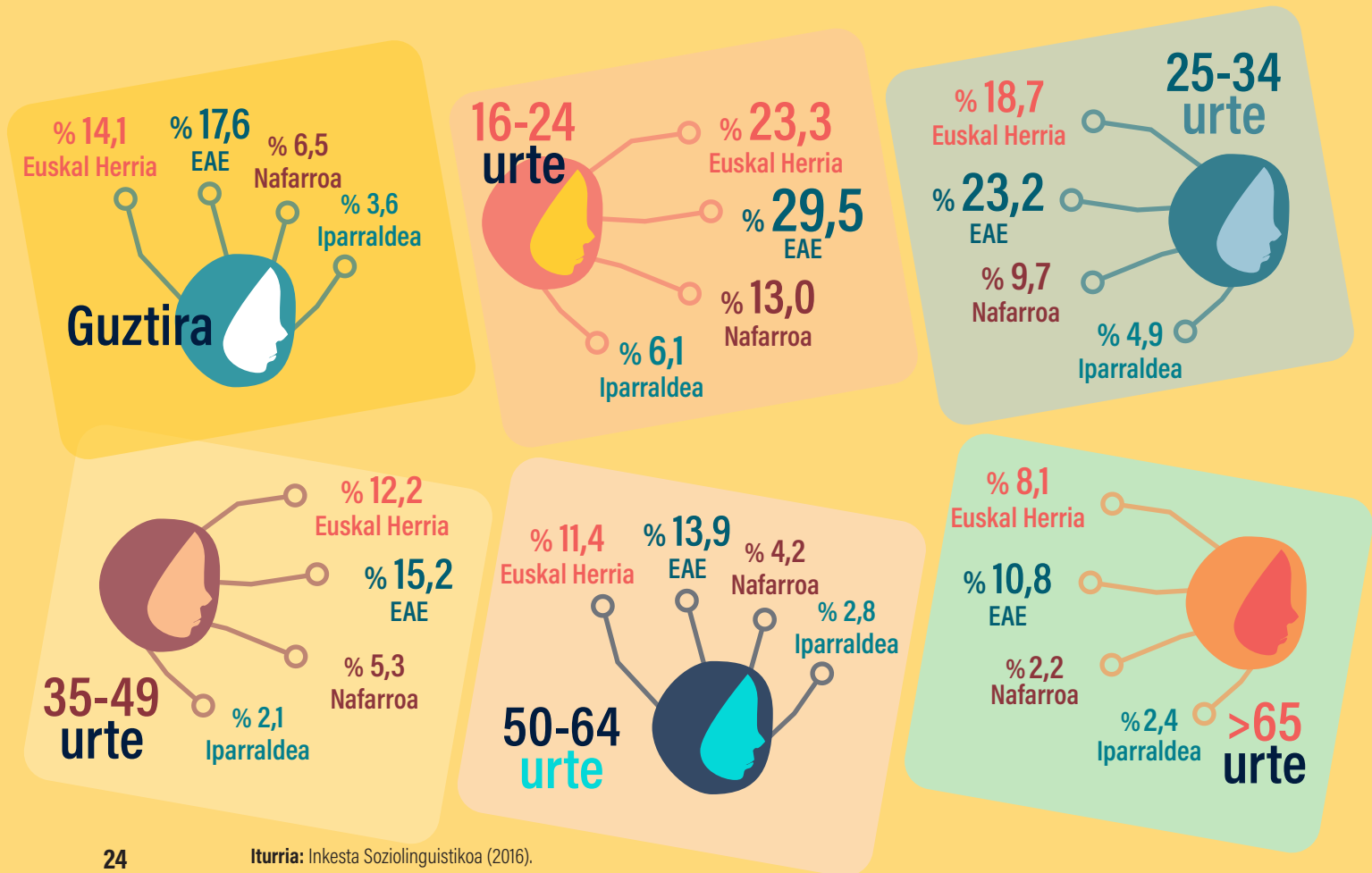


1.

Euskararen erabilera sare sozialetan

Sare sozialetako erabiltzaileen artean euskara nagusiki zenbatek darabilten neurtu da, adin-taldeka. Horretarako, erabiltzaileen artean euskararen erabilera %50ekoa edo handiagoa diren kasuak zenbatu dira. Euskal Herrian, batz bestea, erabiltzaileen %14ak darabil euskara nagusiki; gazteenen artean (16-24 urte bitartekoak), 9 puntu gehiago (%23).

1. grafikoa. Sare sozialetan euskara nagusiki darabilten erabiltzaileak, adinaren arabera. 2016

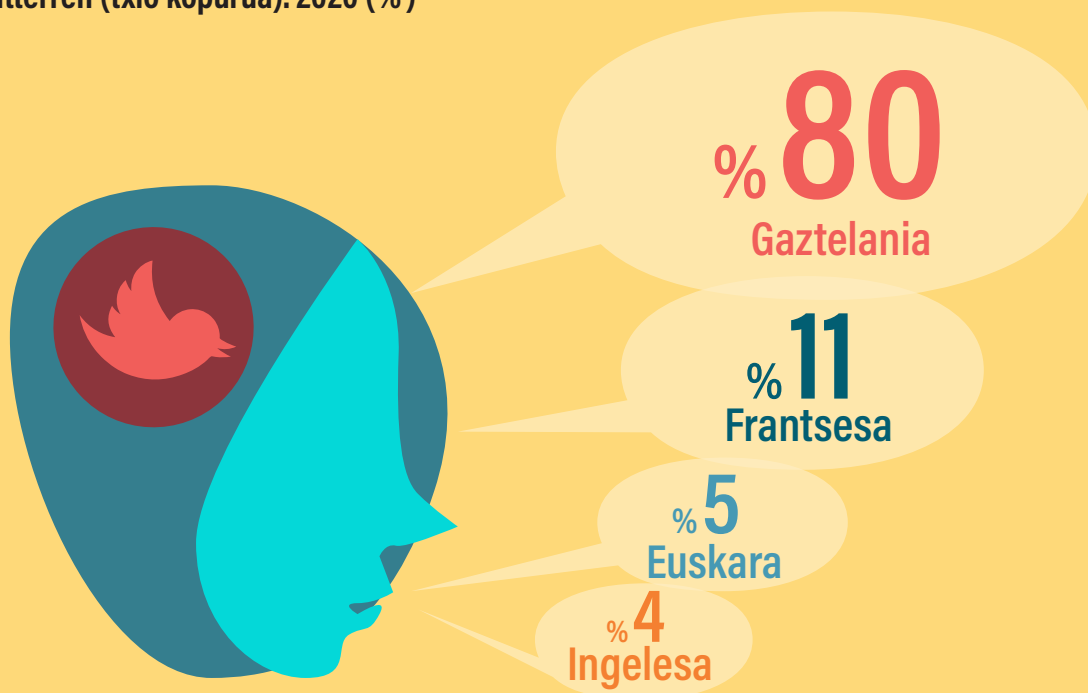


2.

Hizkuntzen erabilera Twitterren

Euskal Herrian hizkuntza bakoitzean txiokatutako mezu kopurua neurtu da. Euskal Herrian 2020. urtean 3.800 twitter erabiltzailek argitaratutako txio guztiak jaso dira neurketa hori egiteko. Euskarazko, gaztelaniazko, frantsesezko eta ingelesezko jarduna hartu da kontuan.

2. grafikoa. Hizkuntzen erabilera Twitterren (txio kopurua). 2020 (%)

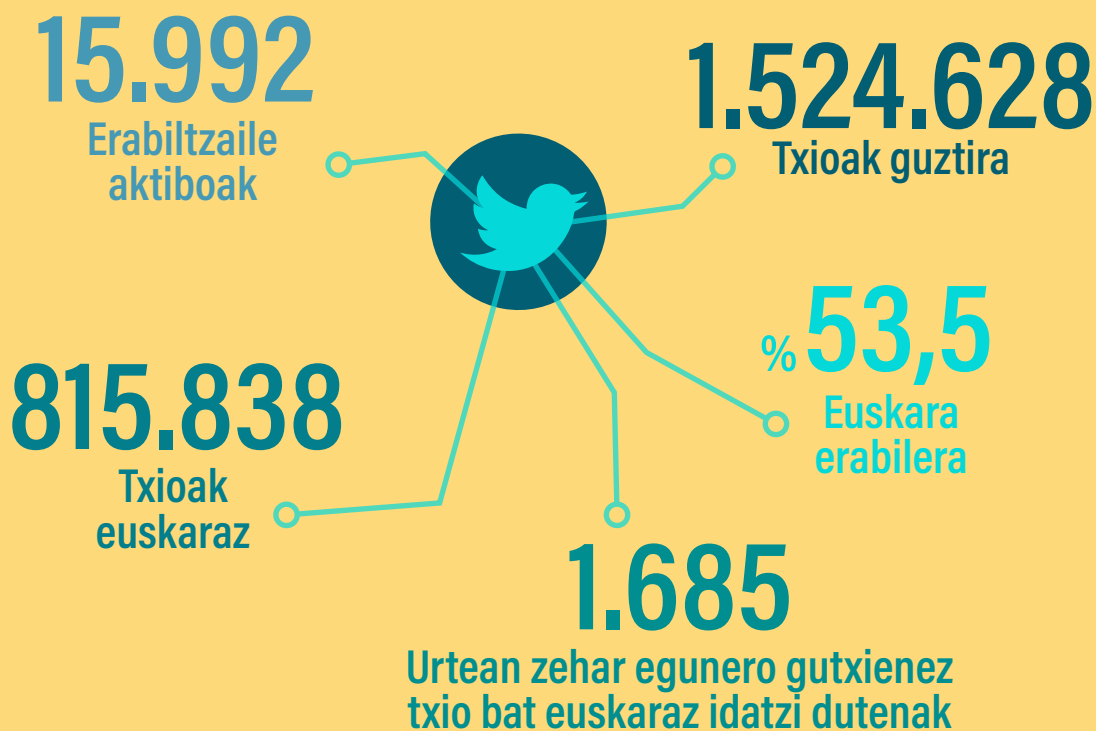


3.

Euskal txiolarien euskarazko jarduna

Umapeko datu-basean 60.000 euskaldun potentzial daude identifikatuta, eta txiolari aktiboak ia 16.000 dira. 2020ko datuen arabera, 6.563 txiolari euskaldun izan ziren beren txioak % 50 baino gehiagotan euskaraz idatzi zituztenak, eta 1.685 txiolari aritu ziren egunero txioren bat gutxienez euskaraz idazten. Euskararen erabilera % 53,5ekoa izan zen euskal txiolarien artean.

3. grafikoa. Euskal txiolarien euskara erabilera. 2020

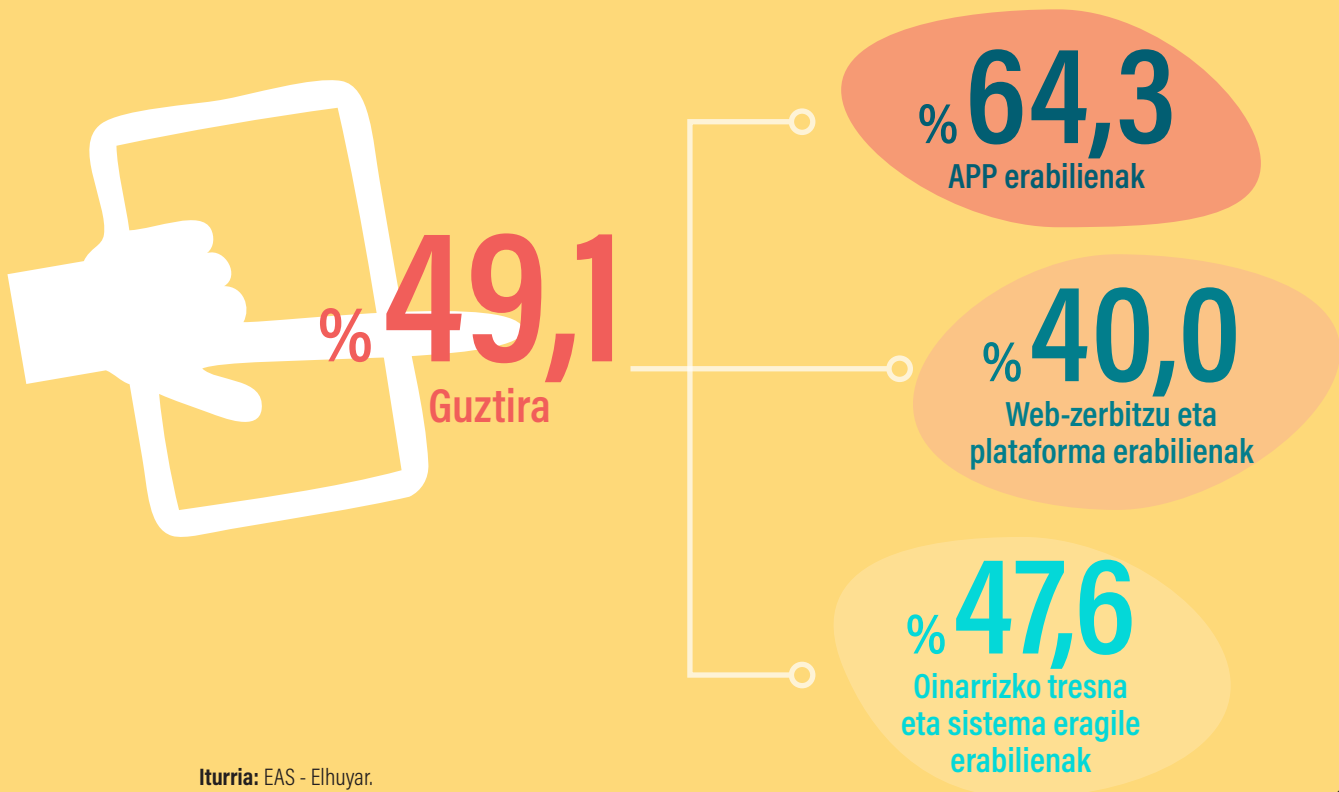


4.

Euskararen eskaintza aplikazio mugikorretan, web zerbitzuetan eta oinarrizko tresna eta sistema eragileetan

Espanian eta Frantzian gehien erabiltzen diren APP-en, Web-zerbitzuen eta plataformen, eta sistema eragileen eta oinarrizko tresnen artean euskaraz zenbat diren adierazten du adierazle honek. Orotara 28 APP, 40 web-zerbitzu edo plataforma eta 42 sistema eragile edo tresna aztertu dira. Euskal Herriko datuak aztertuta, eskaintzaren ia erdia da euskaraz.

4. grafikoa. Euskarazko eskaintza aplikazio mugikorretan, web zerbitzuetan eta oinarrizko tresna eta sistema eragileetan. 2020





HIRUKO
ITUNA
Euskara sustatzeko ituna
3kn

