

Siri eta Cortana ere euskaltegira bidaltzeko garaia

Ahots bidezko ezagutza euskaraz posible egin ahal izateko 'gaitu.eus' ekimena abiatu dute. Ahalik eta euskaldun gehienen grabazioak bildu nahi dituzte, gailu teknologikoez elkarrekintzak ahalbidetzeko. Grabazioak egiteaz gain, grabazio horiek balidatuko dituen jendearen bila ari dira.

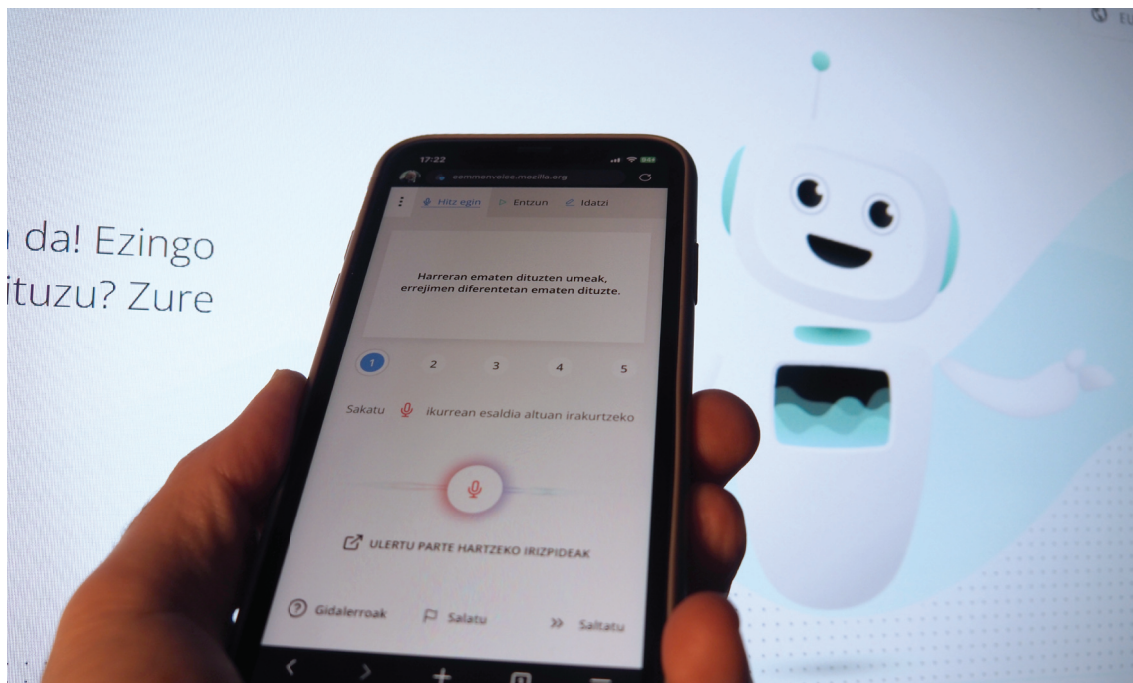
Testua: **Juanma Gallego.**

Oye, Siri. Zer da Arabako ALEA?". *No sé qué responder a eso.* Eskerrak oraingo honetan makina ez den haserretu. Eta arrazoirik ez luke faltan, testu bidez jasotakoaren arabera, ulertu duelako *Siri, cerda, arabaco.* Bigarren saiakeran ere ahalegindu da gizajoa, eta oraingoan bederen, ez du mezua iraintzat hartu. Fin eta zintzo erantzun du ispilu beltzaren barruan sartuta dagoen jeino magikoak. *No tienes recordatorios sobre barbacoa.*

Argi dago, Siri andrea, gure artekoak ez duela etorkizun askorik. Ez euskaraz, behintzat.

"Tren hau galtzen baldin badugu, euskara atzean geratuko da". Oraingoan hizketan dabilena ez da silizioan oinarritutako tramankulu bat, karbonoan oinarritutako adimen natural bat baizik. "Jende gazteak dagoeneko badauka teknologiarekin halako ideia bat buruan txertatuta", jarraitu du Librezaile elkarteko eta Iametzako kide Asier Iturraldek. "Teknologiari gatzeleraz, frantsesez edo ingelesez egin behar zaiolako ideia. Horrela jarraituz gero, teknologiarako balio ez duen hizkuntza zaharkituztat hartuko da euskara".

Agerikoa da zerumuga hori ez dela batere desiragarria. Hau gerta ez dadin, Librezaile elkarteak eta beste zenbait eragilek urte batzuk daramatzate lanean, ahots bidezko ezagutzaren mundura ere euskara eramaten. "Funtsean, helburua da teknologiarekin ahoz euskaraz komunikatu ahal izatea. Hots, euskaraz hitz egiten baldin badiogu, teknologiak uler gaitzala". Iturraldek gogora ekarri duenez, une honetan gero eta ohikoagoak dira ahotsa ezagutzen duten



teknologiak, baina, beste zenbait arazoren artean, dauden tresnek ez dute euskara ulertzen. Googleren Assistant, Amazonen Alexa, Appleren Siri, Microsoften Cortana... ahots laguntzaile asko daude, eta, Librezaile elkarteko kidearen ustez, horiek "arazo asko" dituzte. "Baina euskaldunoi gehien eragiten gaituen arazoa da euskara ez dutela ulertzen. Enpresa hauetako bakoitzak bere tresna entrenatzen du, testuak eta testu horiei lotutako audioak erabiliz. Horretarako hainbat hizkuntzatan irakurritako esaldi laburrez baliatzen dira. Baina enpresa hauek ez dituzte inorekin konpartitzen grabazio eta datu horiek. Beren tresna elikatzeke bai, baina beste inork ezin ditu erabili", argudiatu du.

Hau ez da, gainera, euskararen arazo soila, gainerako hizkuntza gutxitu gehienena baizik. "Dozena bat hizkuntzatan egon ohi dira erabilgarri, baina milaka hizkuntza daude munduan. Gehienetan mendebaldeko hizkuntza nagusiak dira aintzat hartzen dituztenak. Hitzun asko izan arren, munduko gainerako hizkuntza gehienak alde batera uzten dituzte".

Honi irtenbidea eman nahian, *Gaitu.eus* ekimena abiatu dute. Iturraldeak aipatu duenez, *Gaitu*-ren azpian Mozilla Fundazioaren *Common Voice* egitasmoa dago. Helburua da ahalik eta hizkuntza gehienetan datu sorta horiek sortzea, gero edonork erabili ahal izateko modukoak izan daitezen.

Prozesua ez da batere konplikatu. Behin webgune horretara joanda, proiektuaren orrialde nagusira bideratzen dute bisitaria. Bertan, bi atal nagusi ikusiko ditu: *hitz egin*, eta *entzun, lagundu ahotsak balioztatzen*. "Lehen kasuan, bost esaldi labur agertuko dira. Grabatzeko botoia sakatu, eta esaldia ozen irakurtzea baino ez

'Gaitu.eus' webgunera jo behar da 'Common Voice' proiektuaren baitan grabazioak egiteko. ALEA

**EKIMENAK BI ALDE
DITU: ESALDIK
GRABATZEA ETA
BESTEEN GRABAZIOAK
BALIDATZEA**



Parte hartzea bultzatzeko kanpaina bat egin dute; tartean, telebista eta sare sozialetarako spot batekin. IREKIA

dago, bai mugikorretik edo ordenagailuaren mikrofonotik".

Bigarren atala balidazioarena da. "Pertsona gehienek borondate onez egiten dute ekarpena, baina badira ere *troll* modura aritu daitezkeenak ere, proiektua kaltetu nahian. Horiek baztertzeko, egin behar dena da besteek egindako grabazioak entzutea, eta zuzenak diren edo ez esatea. Hau da, erabaki behar dugu grabazioa bat datorren testuarekin ala ez", azaldu du Iturraldek.

Balidazioa aintzat hartzeko, gutxienez bi pertsonak ontzat eman behar dute esaldia.

Horietan, aldaera fonetikoak onartzen dira. Iturraldek jarri duen adibideari jarraiki, hizkera jasoan euskaraz *geologia* modu leunean ahoskatzen da [/geologia/, Nazioarteko Alfabeto Fonetikoan], baina onartzen da ere modu gogorrean ahoskatuta [/xeologia/, fonetikoan]. Ipar Euskal Herriko hiztun batek *h*-a hasperendu egiten baldin badu, hori ere ondo dagoela gaineratu du. "Hizlariak hitzak aldatu baldin baditu, orduan ez dira onartuko. Kasurako, gipuzkoar batek *det* ahoskatzen badin badu, berriz, proiektu honetarako ez da ontzat hartu behar".

Lehenengoz, batua

Azken adibideak euskalkien gaira eramaten gaitu. Proiektuaren atzean daudenek oso argi dute momentuz bederen ahalegina batuan egin behar dela. "Berez Mozillaren proiektuak hizkuntzaren aldaerak —gure kasuan, euskalkiak— onartzen ditu, baina prozesu osoa berriz errepikatu beharko litzateke hasieratik. Lan nekeza da, eta euskara baturako ere kostatu zaigu esaldiak biltzea. Milaka esaldi bildu behar dira", azaldu du Iturraldek.

Erronkaren tamainaz jabetzeko, komeni da atzera egitea. "Hasiera batean Librezele elkartetik esaldiak eskuz biltzen hasi ginen, baina konturatu ginen lan handiegia zela. *Argia*-k esaldi mordo bat eman zizkigun, eta gutxienera langara iritsi ginen. *Gaitu*-ren barruan EITB taldeak ere dohaintzan pila bat esaldi eman dizkigu. Baina bizkaieraz edo zubereraz egin beharko bagenu, aipatu bezala, prozesu osoa errepikatu beharko litzateke. Horregatik, hasiera batean bederen, indarra euskara batura bideratu behar dugu. Gutxienez teknologiarekin batuan hitz egin ahal izateko".

Kasu honetan, zenbaki mardulak garrantzitsuak dira, baina baita aniztasuna bermatzea ere. "Komeni da ahalik eta jende gehienak ahalik eta testu gehien irakurtzea. Pertsona bakar batek mila ordu grabatzen baldin baditu, pertsona horren neurriera erabat egindako hizketa ezagutza teknologia bat garatuko genuke, eta hori ez da nahi duguna. Hizkuntza bakoitzaren barruan ere ahots aniztasuna

bermatu nahi dugu. Silicon Valleyn normalean gizonezko zuri, ingeles hiztunak dira... aniztasun gutxi dago. Emakumeen ahotsak behar dira, kasurako". Iturraldearen arabera, "Siri eta halakoek emakumezkoen ahotsak okerrago ezagutzen dituzte, gizonezkoena baino, entrenatzeko garaian historikoki batez ere gizonezkoen ahotsak erabili izan direlako".

Dena dela, eta kontutan izanda halako gailuen garapena momentuz ekoizle handi batzuen esku dagoela, zertarako osatu halako prozesu bat? Are, zergatik eskaini behar da borondatez ezaugarri biometriko bat? Zalantza hauek argitu ditu Iturraldeak. Dioenez, Elhuyar bezalako enpresek garatu ahalko dituzten produktuetan erabili ahalko litzateke grabazioen *corpus* hau. "Kontrara, Microsoft, Google edo Amazonen daturik ezingo lituzkete erabili, pribatuak direlako. Aintzat hartu behar da *CommonVoice*-ren beste ideietako bat hizketa teknologiak demokratizatzea dela, edonork tresna propioak garatu ahal izan ditzan. *Creative Commons 0* lizentziarekin doaz, jabari publikokoak izatearen parekoan".

Pribatutasunari buruzko kezkei dagokienez, Iturraldeak uste du alde onek alderdi alde txarrak gaintzen dituztela. "Kezka hori zilegia da, eta partekatzen dugu. Baina Mozilla Fundazioak pribatutasuna bermatu du. Ez dago, ahotsaz gain, pertsonalki identifikatuko zaituen ezer. Parte hartu dezakezu modu anonimoan, inolako daturik eman gabe. Edo, nahi izanez gero, eman ahal dituzun datuak oso orokorrak dira: adin tartea, zer motatako azentua daukazu eta gizonezko, emakumezko edo bestelakoa bezala identifikatzen ote zaren. Hau da, ezin da jakin ahotsa norena den. Nik uste dut Mozilla nahiko modu txukunean ari dela gauzak egiten".

Jaurlaritzak eta beste zenbait erakundeek proiektuari eman dioten bultzada zenbakietan nabaritu dela dio Iturraldeak. "Denbora gutxian, 150 bat grabazio ordutatik 250 ingurura pasa gara, baina oraindik ere astindua eman behar diogu honi", nabarmendu du.

Zeregin horretan, herrietako ekimenek garrantzia handia izan ohi dute. Gasteizen kasuan, GEU elkarte ari da proiektua sustatzen. "Estrategikoa iruditzen zaigu eremu digitalean ere euskararen presentzia sendotzea", azaldu du GEU elkarte kide Azaitz Unanuek. "Hainbat eremutan hankamotz gaudela uste dugu, eta hau da horietako bat. Gainera, guk ere bat egiten dugu software librearen inguruko filosofiarekin".

Unanuek azpimarratu du ekimena horren sinplea izateak aukerak errazten dituela. "Militatzeko ahalmen mugatua duten hainbat lagunek euskararen aldeko ekintza erraza eta xumea egin ahal dute honen bitartez. Haur txikiak dituztenak, edo etxetik lasai zerbait egin nahi dutenek ez dute zertan inora joan parte hartzeko. Gainera, bost minutu baino ez dira behar esaldi asko grabatzeko". Auzoetako euskara elkarteekin ari dira ekimena zabaltzen. Kasurako, Sarburu elkartearekin batera egin dute aurrera joan den ostiralean, eta Zabalganean, Judimendin edo Abetxukun ere gauzak prestatzen ari dira. Hilaren 17an, gainera, Iturralde bera gonbidatu dute Landatxon gaiari buruzko hitzaldia eman dezan, eta aurrera begira grabazio maratoni bat egitea dute buruan. ■■■

**GASTEIZKO AUZOETAN
EKIMENA ZABALTZEAZ
GAIN, GRABAKETA
MARATONI BAT EGITEA
DUTE BURUAN**