

# ...eta hizkuntza- teknologiak lantzen 20!

Elhuyar Fundazioak 50 urte bete ditu aurten, baina, horrez gain, 20 urte ere badira Elhuyarren hizkuntza-teknologiak ikertu, garatu eta merkaturatzeari ekin genionetik. 20 urte, eta asmo eta helburu bakarra: beste edozein hizkuntzak bezala euskarak ere hain beharrezko dituen hizkuntza- eta hizketa-teknologiak garatu eta gizartearen eskura jartzea. Jardun horren fruitu dira, besteren artean, oso ezagun eta askorentzat ezinbesteko bihurtu diren gure corpusak, hiztegiak, zuzentzaileak, itzultzaile automatikoak (Elia.eus), transkribatzaile automatikoak (Aditu.eus) edota hizketa-sintetizatzaileak.

2002. urtea mugarri garrantzitsua izan zen Elhuyarrentzat. 30 urte bete zituen urte horretan bertan eman zuen kultura-elkarte izatetik fundazio bihurtzerako pausoa. Baina, gainera, gerora oso garrantzitsuak izango zirela sumatuta eta etorkizunari aurrea hartuz, hizkuntza-teknologien arloa lantzeari ekin zion. Eta batekin nahikoa ez, antza, eta bi frontetan egin zuen hori.

## Eleka eta Elhuyar I+G

Alde batetik, Eleka enpresaren sorrera dago. [EHUko IXA Taldeak](#) bazeramatzan urte batzuk euskararentzako hizkuntza-teknologietan oinarrituko ikerketa egiten, eta jada eginak zeuzkan etiketatzailea ([orduan Euslem, gerora Eustagger](#)), zuzentzailea ([Xuxen](#)) eta itzultzaile automatikoa ([Matxin](#)). Baina horiek gizartearen eskura jartzeak edo merkaturatzeak eskatzen dituen lanak (plataforma ezberdinetarako egokitzapenak, bertsio berriak ateratzea...) unibertsitateko ikerketa-talde baten ohiko lanetatik kanpo zeudenez, ez zen nahi eta behar bezala egiten. Hutsune hori betetzen laguntzeko prestasuna agertu zuen Elhuyarrek, eta hala sortu zuten elkarrekin Eleka enpresa. Geroztik, ez soilik hasierako horiek, hizkuntza- eta hizketa-teknologietan oinarritutako beste tresna asko ere merkaturatzen jarraitu du, gaurdaino. Eta denborarekin, gero eta gehiago joan da bideratzen Elhuyarreko ikerketa-sail propioak egindakoak gizarteratzera, elkarlan estu eta oparoan.

Igor Leturia Azkarate  
Informatikaria eta ikertzailea



Izan ere, Elhuyar barruan sail berri bat sortu zen, Elhuyar I+G, helburutzat Elhuyarreko beste sailek behar zituzten hizkuntza-teknologiaren ikerketa eta garapena egitea zuena. Hasieran bereziki hiztegi-gintzarako tresnak egitera bideratu zuen jarduna, Elhuyarrentzat sail garrantzitsua baitzen hori: 1996an argia ikusi zuen euskara-gaztelania/gaztelania-euskara [Elhuyar Hiztegia](#) erreferente bihurtua zen, hiztegi-ekoizpena hizkuntza gehiagotara zabaltzeko asmoa zegoen, hiztegi terminologiko gehiago egitekoa... Hala, Elhuyar I+Gk testu-corpusak biltzeko lanei ekin zien ([ZT corpusa](#), adibidez), webetik automatikoki hainbat motatako corpusak (elebakar [orokor](#) zein [espezializatuak](#), elebidun [paraleloak](#) zein [konparagarriak](#)...) biltzeko teknikak eta tresnak landu zituen, [ontologiak automatikoki sortzeko](#)

[teknikak](#) garatu zituen, [corpusetatik hiztegiak automatikoki eraikitzeko tresnak](#) garatu... Horietaz gain, bilaketa-tresnetan ere ikertzen hasi zen berehala saila, eta [Zientzia.eus-en dagoen bilatzaile eleaniztuna](#) eta haren antzeko tresnak garatu zituen. Ikus daitekeenez, Elhuyarren beste sailen tresna ezagun horiek Elhuyar I+Gk egiten zuenaren erakusgarri eta erakusleho bikaina ziren. Baina jarduna ez zen horretara mugatu, eta asko dibertsifikatu da azken 20 urteotan.

Ikerketa-talde guztietan bezala, Elhuyar I+Gn ere nazioarteko joerei jarraituz egin dira ikerketan lanak, akademiaren munduko azken ikerketetan oinarrituz eta arloari ekarpenak eginez. Hala, [100 artikulua zientifiko baino gehiago argitaratu ditu kongresuetan edo aldizkari espezializatuetan, eta](#)

[7 doktore-tesi egin dira sailean](#). Elhuyar I+Gk elkarlan estua eta jarraitua izan du beti IXA Taldearekin. Eta, 15 urtean baino gehiagoan, IXA Taldearekin, [EHUko Aholab ikerketa-taldearekin](#) eta [Tecnalia](#) eta Vicomtech ikerketa-taldeekin jardun du ikerketa estrategikoko hainbat elkarlan-proiektutan, askotan lider gisa.

*“Hainbat teknologia garatu ditugu arrakastaz euskararentzat, eta gizartearen eskura jartzen dihardugu”*

### Sare neuronal sakonen egungo momentua

20 urteotan, hizkuntza-teknologietan erabili izan diren tekniketari eboluzio handia gertatu da. Honetan hasi ginenean, [erregeletan oinarrituak deritzen teknikak](#) ziren nagusi. Halakoetan, hizkuntzari buruzko ezagutza (hitzak, deklinazio-arauak...) ordenagailuek ulertzeko moduko lengoia eta egituretan jartzen zen. Metodo horiekin, gauza batzuk oso ongi egiten ziren (aipatutako etiketatzea, zuzenketa ortografikoa, bilaketa eta horrelakoak), baina itzulpen automatikoa edo hizketa-ezagutza, esate baterako, ez hain ondo. Geroago, [ikasketa automatikoko metodoak edo metodo estatistikokoak](#) agertu ziren, adibideetatik ikasten zutenak, baina horiekin ere ez ziren behar bezain emaitza onak lortzen hainbat atazatan, ez behintzat euskaraz.

Duela bospasei urte, [sare neuronal sakonak \(Deep Neural Networks\)](#) edo [ikasketa sakona \(Deep Learning\)](#) izenez ezagutzen diren metodoak azaldu ziren eszenan. Izatez, ikasketa automatikoaren kasu partikular bat dira, eta, gainera, [sare neuronalak](#) aspalditik existitzen ziren, baina, emaitza onik ematen ez zuten, baztertuta zeuden. Teknogiaren aurrerapenak sare neuronal handiago eta konplexuagoak (hortik “sakonak” izendapena) erabili ahal izatea ahalbidetu zuen, [GPU edo txartel grafikoek](#) asko azkartu zituzten beren ikasketa- edo entrenamendu-denborak, eta digitalizazioaren eta Interneten aurrerapen handiak askoz datu gehiago jarri zituen eskuragarri sareok entrenatzeko, eta, bat-batean, ataza konplexu askotan, hala nola itzulpen automatikoa edo transkripzio automatikoa, ordura arte baino emaitza askoz hobeen berri iristen hasi zen nazioartetik. Eta, probak eginda, ikusi genuen euskararekin ere gauza bera lortzen genuela. Ia egun batetik bestera, lan askotan emaitza onargarriak. Azkenean!

Ordutik, hainbat teknologia garatu ditugu arrakastaz euskararentzat, eta gizartearen eskura jartzen dihardugu: [Elia.eus itzultzaile automatikoa](#), [Aditu.eus transkribatzaile automatikoa](#), hizketa-sintesi pertsonalizatua, [BERT motako lengoia-eredu neuronalak](#), [txatbot-ak](#), [Mycroft.eus euskarazko bozgorailu adimenduna](#)... Horiek guztiek eragin nabaria izan dute euskal gizartean eta euskararen egoera digitalean, eta ziurrenik are handiagoa izango dute gerora.

*“Elhuyarrek beti jakin izan du, gizartearen eboluzioari adi egonez, euskararentzat eremu berriak irabazten eta euskararen behar berriei erantzuten”*

### **Etorkizuna Orai da**

Sare neuronal sakonek ekarritako leherketarekin, hazkunde handia izan du Elhuyarren hizkuntza-eta hizketa-teknologien jardunak eta lantaldeak: ikerketan eta tresna eta zerbitzuen garapenean aritzen direnak batuta, 20 pertsona inguru ari gara egun.

Eta aurten, Elhuyarrentzat hain esanguratsua den 2022. urte honetan, beste jauzi esanguratsu bat eman dugu: Elhuyarren I+G sailak izen eta marka berriak ditu orain, [Orai NLP teknologiak](#), hain zuzen. Baina izen-aldaketa soila baino gehiago da, izana ere aldatu da. Euskarak beharrezko dituen hizkuntza-baliabide eta -tresnak sortzeko lana utzi gabe, adimen artifizialaren eta hizkuntza-teknologien ahalmena euskal enpresetan aplikatzeko ahalegin handiagoa egin nahi da, neurrirako ikerketa eginez eta enpresa horiei lehiakorragoak izaten eta hizkuntza-oztopoak gainditzen lagunduz. Horrez gain, beste hizkuntza gutxituentzat bide erakusgarri eta bidelagun izateko jada ekindako bidean gehiago sakondu nahi da, eta, kide okzitaniar eta aragoiarrentzako itzultzaile automatikoak eta hizketa-sintetizagailuak garatu ditugun bezala, tresna gehiago egiten jarraitu nahi dugu horientzat edo beste hizkuntza minorizatu batzuentzat. Elhuyarren urteurreneko leloak dioen bezala, *50 urte aurrerago* begira dihardugu Orai-n ere, euskal gizarteak, euskarak eta beste hizkuntza txikiek ere etorkizun argitsua izan dezaten.

Elhuyarrek beti jakin izan du, gizartearen eboluzioari adi egonez, euskararentzat eremu berriak irabazten eta euskararen behar berriei erantzuten. Duela 20 urte etorkizun-sen handia erakutsi zuen hizkuntza-

teknologien aldeko pauso sendo eta ausarta ematean, eta, fedez edo egoskortasunez, eutsi egin dio apustuari urteotan zehar (urte batzuk benetako desertuko zeharkaldi izanik ere: krisia, itzulpen automatikoan eta beste zenbait atazatan betetzen ez ziren emaitza onen promesak...). Horren fruituak jasotzen ari gara gaur egun, eta zorionekoak gara gu geu, Elhuyarren gauza interesgarri eta harrigarri asko egiten ari garelako, baina baita euskara ere, egungo eta etorkizuneko mundu digitalean atzera ez gelditzeko tresnak eskuratzen ari delako; eta, atrebentzia onartzen badidazue, baita lerrook idazten dituen hau ere, ia hasieratik Elhuyarreko hizkuntza-teknologien I+Gko taldean lan egiteko pribilegioa eta ohorea izan baitut. Beraz, zorionak eta eskerrik asko, Elhuyar! Eta zorionak Orairi eta hizkuntza-teknologien taldeari ere, eta urte askotarako! ●



Mycroft euskarazko bozgorailu adimenduna garatzen ari da Orai. ARG.: Elhuyar.