



Adimen artifiziala euskarara irits dadin lanean ari da Hitz Zentroa

Jaurlaritzak 1,68 milioi euro eman dizkio EHUko zentroari teknologia berriak iker ditzan, euskara sustatzeko errealitate digitalean • Datu bilketa masiboa abiatzeko deia egin dute

Xabier Martin Donostia

Adimen artifizialari esker lor daiteke hiztun edo testu baten euskarazko gaitasun maila era automatikoa antzematea; edota euskarazko ahots pertsonalizatuaren teknologia desgaitasuna dutenengan iritsaraztea. Ez hori bakarrik: ahots espontaneoaren transkripzioa egiteko teknologia gara daiteke; baita euskara elkarriketa sistemarik ahaltsuenen pare jarri ere, esaterako. Ez dira erronka apalak, baina EHUko Informatika Fakultateko Hitz Zentroak aurre egin nahi die lan horiei. «euskara ingurune digitalean areagotzeko». Aurrerantzean, zentro horretan aritzen diren 70 ikerlarik –informatikariak, hizkuntzalariak eta ingeniariak– Eusko Jaurlaritzak zentroari eman dion 1,68 milioi euroko diru laguntzaren babesa izango dute beren zereginen jarraitzeko.

Diru ziztada horrek izango du bigarren helburu bat, gainera: Hitz Zentroak Clariah-eus sarea sortu du, Europako Clarin eta Dariah azpiegitura zientifikoetan erkatuta, eta hura bultzatu nahi dute. Azpiegitura horietako 137 ikertzailek eta euskara ikertzen ari diren 11 taldek eta erakundek babestu dute Clariah-eus sarea, besteak beste humanitate, arte eta zientzia sozialetan ikerketa sustatzeko eta elkarrekin partekatze asmoz datu zein tresna digitalen erabilera.

Jaurlaritzako Kultura eta Hizkuntza Politikako sailburu Bingen Zupiriak eta EHUko erreto-



Bingen Zupiria, Eva Ferreira eta Eneko Agirre, atzo, Donostian. BERRIA

re Eva Ferreirak bisita egin zieten Hitz Zentroko ikerlariari, atzo, Eneko Agirre zuzendaria buru izan zutela. Zupiriak esan zuen «eremu digitalean hizkuntza hegemonikoak» direla «jaun eta jabe», batez ere ingelesa, eta «beharrezkoa» dela euskarak ere «bere tokia» hartzea alor horretan: «Bestela, beste hizkuntza batzuek hartuko dute espazio hori». Sailburuak erantsi zuen «ezinbestekoa» dela hizkuntza teknologia lantzea, eta «trakzio lana» egin nahi duela Jaurlaritzak helburu horren alde. Adi-

men artifizialak sortutako bere ahots pertsonalizatuaren lagin bat entzuteko aukera ere izan zuen Zupiriak Hitz Zentroan.

Euskara eta batasuna

Eva Ferreira EHUko errektoreak, berriz, «benetako iraultza teknologikoan» bizi garela gogoratu zuen: «Errealitate materialari errealitate birtuala gehitu zaio, eta, gaur egun, ez dago bizirik iraungo duen giza errealitate mundu digitalean korrespondentzia egokirik ez badu». Eta horrek ondorio «argi bat» duela erantsi

zuen: «Euskarak hor egon behar du, eta indartsu, baliabide egokiak». Ferreirak esan zuen 1970eko hamarkadan euskara estandar bat egiten hasi zirela hizkuntzalariak, eta Koldo Mitxelena oso argi adierazi zuela batasunik gabe euskarak ez zuela gerorik izango; batasunarekin ere biziraupena ez zela bermatua ohartarazi zuen Mitxelena, baina batasunik gabe segurua zela euskararen galera. Bada, egun euskarak alor digitalean behar duen garapenarekin alderatu zuen ideia hori Ferreirak.

Izan ere, egungo eta etorkizuneko teknologiek testuzko eta ahozko datu masa erraldoiak eskatzen dituzte hizkuntza bakoitzeko, eta, aldiz, euskarak eskaini ahal dituen datuak ez dira nahikoak, teknologia horiei etekina ateratzeko, eta gailu horiek behar adina kalitate eskaintzeko. Hor dago Hitz Zentroaren benetako erronka, zentroko zuzendari Eneko Agirrek esplikatu zuenez: «Egungo teknologiak hiztun ugari hizkuntzetan oso ondo aritzen dira, baina hiztun gutxi-koetan kalitate eskasa ematen



Dugun testu masarekin ezin dira elkarriketa sistemen motore diren hizkuntza ereduak euskaraz entrenatu»

Eneko Agirre

EHUko Hitz Zentroko zuzendariak

dute. Gure ikerketa batek erakutsi du egun dagoen euskarazko testu masarekin ezin direla elkarriketa sistemen motore diren hizkuntza eredurik ahaltsuenak euskaraz entrenatu, eta ahotsarekin ere berdin gertatzen da». Horregatik, Hitz Zentroan eredu berriak eraikitze metodoak ikeritzen ari dira, eta, horren osagarri gisa, testuzko eta ahozko datu-masak biltzen ari dira.

Agirrek nabarmendu zuen eraikitzen dituzten eredu eta algoritmo guztiak inguruko zentro teknologiko eta enpresekin «era librean» partekatuko dituela EHUk, betiere euskararen presentzia handitzeko alor digitalean. Eta azken mezu bat ere helarazi zion euskalgintzari: «Euskarazko hizkuntza teknologia propioa eta kalitatezkoa nahi badugu, testuen eta ahotzen datu bilketa masiboa abiatu beharra dago, eta erakunde publikoen nahiz pribatuen eskuzabaltasuna gako izango da horretarako».