



Olatz Arregi Uriarte

Donostiako Informatika Fakultateko irakasle eta ikertzailea

Euskararen iraungipen digitala ekidin dezagun

Hizkuntza txikien iraungipen digitalaren atarian zer egin. Hizkuntza-teknologiak gero eta txertatuago daude gure eguneroko bizitzan, testuak automatikoki zuzentzean, webguneak automatikoki itzultzean, telefono mugikorrean ahots-aginduak ematean, GPS nabigatzaileak erabiltzean... Eta teknologia horiei esker, gero eta modu naturalagoan komunika gaitezke ordenagailuekin, etxetresnekin edota ibilgailuekin. Jokaleku horietan, Interneten bezalaxe, ingelesa da nagusi, eta, arrisku handia dago hizkuntza hegemonikoez guztiz berengana dezaten atentzioa eta sostengua, eta hizkuntza minorizatuak, euskara barne, zokoratu gera daitezen.

Indartsu datoz adimen artifiziala eta hizkuntza-teknologiak, eta abagune honek politika ondo diseinatuak eta eraginkorrek behar ditu. Estatu ahaltsuek eta, batez ere, erraldoi teknologikoez –Google, Microsoft, Facebook, Amazon eta IBM bezalakoek– hartu dute beren lekua, uhin honek ihes egin ez diezaien. Agertoki globalean gertatzen ari da eraldaketa teknologikoa, baina ez horratik modu homogeneoan.

Hain zuzen ere, kalitate handiko hizkuntza-teknologia garatzea ezinbestekoa eta presazkoa da euskararen kasuan. Ezinbestekoa, hizkuntzaren beraren irautea bermatu nahi bada; eta presazkoa, inondik ere, teknologia hauen bilakabidea eta berritze-abiadura ikusita atzean geratzeko arriskua benetakoa baita. Aurrerakuntza teknologikoa behar du euskarak, era digitalean atzean geratuko ez bada.

Hizkuntza-teknologiak, erronka eta aukera hizkuntza txikientzat. Hizkuntza txikiek jauzi kualitatiboa eta kuantitatiboa eman behar dute ekosistema digitalean. Izan ere, mundu digitala aukera eta erronka da gaur egun euskara bezalako hizkuntzentzat. Per-

tsonen eta teknologiaren arteko elkarrekintza fase berri batean sartu da, gero eta gutxiago erabiltzen dira teklatuak, saguak edo ukipen-pantailak, eta gehiago ahotsa interakzio naturalagoa lortzeko. Eta horrek garrantzi berezia ematen dio hizkuntzari, elkarrekintzarako bide nagusietakoa bihurtzen ari baita. Eta hori, euskaraz egin nahi dugu.

Argi dago plan bat garatu behar dela euskara esparru digitalera eramateko, eta une honetan, badugu nora begiratu. Katalanek proiektu erraldoi bat aurkeztu berri dute asmo horrekin, AINA proiektua.

Adimen artifizialean eta hizkuntza-teknologietan oinarritutako proiektua da AINA, eta helburua da herritarrek, beste hizkuntza globaletan gertatzen den bezala, katalanez parte hartzeko aukera izatea mundu digitalean, eta hartara, hizkuntzaren iraungipen digitala saihestea.

Proiektua Kataluniako Generalitateko Politika Digitalen Sailak bultzatu du. Guztira, 2020tik 2024ra, 13,5 milioi euroko aurrekontua du, gehiena Europar Batasuneko Next Generation EU funtsetik datorrena. Ez da makala, beraz, katalanak esparru digitalean behar duen presentzia izateko egingo den inbertsioa.

AINA garatu ahala, adimen artifizialean eta hizkuntza-teknologietan oinarritutako hizkuntza-baliabideak eta baliabide digitalak sortuko dira katalanerako, hala nola ahots-laguntzaileak, itzultzaile automatikoak edota elkarrizketa-agenteak. Kontuan izan behar da hizkuntza-teknologia aurreratuek kalitatezko datu multzo handiak behar dituztela aplikazio eraginkorrek sortu nahi badira.



Katalanak hasiak dira lanean, eta sortu duten lehenengo baliabidea corpus (nolabait egituratutako testu sorta handia) erraldoi bat da, 1.770 milioi hitzekoa, orain artean katalanerako sortu den handiena. Corpus hori, adimen artifizialean erabiltzen diren algoritmoak elikatzeko erabiltzen hasia da, eta horri esker, hemendik aurrera sortuko diren tresnek orain artekoek baino kalitate hobea izango dute. Corpus berriak ere garatuko dira, non katalanaren aldaera dialektalak jasoko diren, baita erregistro linguistiko ezberdinak ere (lagunartekoa, administratiokoa, jaso eta abar). Gainera, irudi- eta ahots-fitxategiak ere bilduko dira. Hurrengo urratsetan, oinarritzko beste hizkuntza-baliabideak sortuko dira; horrela, enpresek adimen artifizialean oinarritutako aplikazioak katalanez sor ditzakete, hala nola ahots-laguntzaileak, itzultzaile automatikoak, elkarrizketa-eragileak eta abarrak.

Tresna horiek guztiak sortzeko behar diren adimen artifizialeko algoritmoak entrenatzeko Kataluniak badu azpiegitura egokia, Bartzelonako superkonputazio zentroa (BSC)². BSCk parte-hartze aktiboa du proiektuan eta dagoeneko, 2.000 prozesu-ordu erabili dira goian aipatutako corpus erraldoia garbitzeko: bikoizketak ezabatzeko eta katalanez ez dauden esaldiak kentzeko.

Euskarari ere ondo etorriko litzaioke horrelako plan zehatza eta diruz ondo hornitua, iraupen digitala ekidin nahi badu. •