

Makina bat neurona euskaraz

Urte bizia izan da itzultzaile automatikoen agendan. Iazko azaroan, Modela aurkeztu zuten; aurten, Batua sareratu dute; eta, iragan astean, Eusko Jaurlaritzak berearen berri eman zuen. Hiruretako edozein probatu duenak suma dezake aldea dagoela orain arteko euskararen eta gaztelaniaren arteko itzulpenekin. Bingen Zupiria Eusko Jaurlaritzako Hizkuntza Politikako sailburuak aurkezpenean nabarmendu zuen «jauzi kualitatibo handia» izan dela azken urtean aurkeztu duten teknologia berria, eta ideiarekin bat datoz garatzaile guztiak. Amaitu dira arau gramatikalei heldu eta estatistikek diotenari besterik jarraitzen ez dioten programak; sare neuronal bidezko itzultzaile automatiko berriek pentsatu egiten dute.

Izenetik beretik abiatuta, «gure garunean ditugun neuronon funtzionamendua emulatzen dute» itzultzaile berri horiek, Araceli Diaz de Lezanaren hitzetan; Hizkuntza Plangintzako arduraduna da Eusko Jaurlaritzako Euskara Sustatzeko Zuzendaritza. Hark azaldu du sistema «entrenatu» egiten dela, eta, entrenamendu horien ondorioz, «ikasi» egiten duela.

Ikasteko, irakatsi egin behar. «Lehen, itzultzaile automatikoak hizkuntzalaritza ziren ia-ia: gramatikaren arauari jarraituz, testu bat beste bat bihurtzen zuten. Egun, ez da hainbeste hori: sare neuronalak matrizeak eta matrizeak dira, bata bestearen atzetik». Xabier Garcia de Kortazar Transferentzia teknologikoan aritzen da lanean Vicomtech enpresan. Haiek parte hartu zuten Modela proiektuan, beste lau eragilerekin —Ametzagaiña, Elhuyar, ISEA Enpresako Zerbitzu Aurreratuetan Berrikuntza eta Ixa taldea—.

Sare neuronalen bidezko itzultzaile automatikoek corpus elebidunak behar dituzte, kalitate handikoak. Hizkuntza bakoitzeko itzulpen unitateek parekoak izan behar dute; kalitatezko itzulpenak lortzeko, gutxienez hiru milioi itzulpen unitate behar dituzte, Diaz de Lezanaren arabera. Haren esanetan, orain arteko sistemen arazoa da hizkuntza unitateak «modu soltean» tratatzen dituztela. Sare neuronalen bidezko itzulpen sistemen ereduak, berriz, «*deep learning* eta *representation learning*» teknikak erabiltzen dituzte: «Datuak modu hierarkikoan antolatzen eta erabiltzen ikasten dute. Horien arabera eraikitzen dituzte ereduak, eta horien arteko funtzionamendua prediktiboa da». Horra arrakastaren gakoa: «Itzulpen horiek askoz ere naturalagoak dira, koherenteagoak; erraz irakurtzen dira, esaldi barruko elementuen artean kohesioa eta lotura dagoelako».

Eusko Jaurlaritzarena da berriena: iragan astean aurkeztu zuten. Jaurlaritzak 2012. urtetik du itzultzaile gramatikala webgunean, gramatika eta lexiko konputazionala oinarri dituena. Oraingo itzultzaile neuronalak Vicomtech enpresak garatu du. Hortaz, funtsean teknologia bera da Modelatik Batuara eta hortik Eusko Jaurlaritzaren itzultzailera garatu dutena.

Jaurlaritzaren itzultzaileak 4.000 karaktereko muga du, eta www.euskadi.eus/itzultzailea helbidean erabili daiteke, doan.

Modelak ekarri zuen aldaketa, iaz. «Ikusita ikerketak zer bide hartu duen nazioartean, nolabait erreplikatu egin genituen ikerketa ildo horiek euskararen eta gaztelaniaren artean», esplikatzen du Itziar Cortes Elhuyarrek Itzulpen gintzarako Teknologien arloko arduradunak. Eusko Jaurlaritzako Industria sailak finantzaturako deialdi batera aurkeztu ziren bost kideak, eta elkarlanean garatu zuten proiektua. Ikerketa bukatu zenean, erabaki zuten *Modela.eus* sareratzeari; «emaitzak benetan onak ziren, eta gizartearen esku jarri genituen».

Elkarlanak eman zuen aurrera egiteko modua, baita teknologia aurrerapenek ere. «Duela urte batzuk, oso zaila zen hardware aldetik erantzun ahal izatea horrelako ikasketa prozesu zorrotzei, eta baliabide asko behar zituzten», Cortesen hitzetan. Gaur egun, horrelako makinak entrenatzeko aukera dute garatzaileek.

Orain, sarean dago Modela: Elhuyarrek eta Mondragon Linguak doako zerbitzu mugatua jarri dute, hilean 2.000 hitzekoa. Ordainpeko eskaintzak ere jarri dituzte. Izan ere, Cortesek adierazi du aurrez aipaturako hardware hori eskuragarri izateak ez duela esan nahi merkea denik. Zerbitzuak atzetik duen lana mantentzeko mugatu dute doako eskaintza.

Modelaren atzetik etorri dira beste batzuk, ordea. Ikerketa proiektua amaitu ondoren, Vicomtech enpresakoek hobetzeko bidea ikusi zioten hari, eta aurrera segitzea erabaki zuten. «Apustu bat egin genuen etxe barruan», azaldu du Garcia de Kortazarrek. «Jende gehiago hartu genuen, beste profiletakoa, eta jarraitu egin genuen. Horren ondoren, Batua atera dugu». Kalitatean hobetu dutela esan du, baita zerbitzuaren abiaduran ere. Halere, haien helburua ez da produktua ustiatzea: «Gu ikerketa zentro bat gara; horregatik utzi dugu libre eta irekia, jendeak bere ekarpenak egin ditzan eta ikas dezagun». *Batua.eus* helbidean dago beta bertsioa: 1.000 hitz itzuli daitezke aldiro; hori da erabilera muga bakarra. Hala ere, Garcia de Kortazarrek nabarmendu du haien lana teknologia garatzea dela, eta ez merkatura eramatea.

Corpusa gizentzen

Itzultzaile neuronalean, corpusa da garrantzitsua. Eusko Jaurlaritzak iturri garrantzitsua izan du, urteetan egindako itzulpenak erabili baitituzte proiekturako: HAEE Herri Ardularitzaren Euskal Erakundeak hogeitun urtean sortutako itzulpen memoriak. Denera hamar milioi segmentu dituen datu basea dago oinarrian.

Euskaratik gaztelaniara eta gaztelanitik euskarara aritzen dira; oraingoz ez dago beste hizkuntzarik itzultzaile horietan. Hala ere, Garcia de Kortazarrek adierazi du ez litzatekeela horretarako arazo handirik egongo: «Berez, ia edozein hizkuntzatan lan egin dezakegu. Dena den, beharrezkoa da hizkuntzalariek hasierako corpus horiek aztertzea».

Elhuyar fundazioa aitzindaria izan da itzultzaile automatikoetan. Modela izan da azken boladak aurrena, baina, horretara iristeko, Matxin izan zen lehen urratsa. 2007an sortu zuten lehen bertsioa, EHUko IXA ikerketa taldeak eta Elhuyarrek: Matxin 1.0.

«Arauetan oinarritutako sistema bat» zen hura, Cortesen hitzetan. Hizkuntzalariek sortzen zituzten itzultzaileentzako arauak. «Informatikarien edo ingeniarien lana zen hizkuntzarekin lotutako baliabide horiek guztiak martxan jartzea, eta hortik itzultzaile bat sortzea».

Duela hamabi urteko kontuak dira horiek; «Googlek ekarri zuen aldaketa». Cortesek berak azaldu du: «Googlek edozein hizkuntzatako itzultzaileak sortzen zituen hizkuntzalarien beharrik gabe, hizkuntza jakin gabe». Teknologian jauzi bat egin zuten: «Aurrez itzultitako testu bilduma erraldoiak hartzen zituzten, eta, metodo estatistiko bidez, nolabait ikasi egiten zuen itzultzaileak». Metodo horrek ez zuen emaitza oso onik eman euskaraz, haren esanetan.

Alde guztiek adierazi dute itzulpen neuronalak aurrerapauso nabarmena eragin duela. Orain, hainbat jakintza eremutako corpus zehatzekin entrenatu behar dituzte makinak. Garcia de Kortazarrek, ordea, ikusi dizkio «hutsuneak, edo akatsak»: irakurterraztasunean eta naturaltasunean irabaztearen truke, zehaztasunean pixka bat galdu dela uste du. «Testuinguruaren arabera egiten du lan beti, orokorrean ikusi duenaren arabera». Testu arruntetan ez dago arazorik, baina kasu batzuetan ez ditu itzulpen zehatzak egingo; legeen izenak edo departamentuak itzultzean, esaterako.

Elkarlana eta lehia

Topikoa den arren, Cortesek eta Garcia de Kortazarrek nabarmendu dute itzulpen automatikoa garatzeak ez duela inolako asmorik itzultzaile profesionalen lana ordezkatzeko. «Akatsak hor daude, eta akats horiek zuzentzeko gai izan beharko genuke», esan du Cortesek. «Norberaren erantzukizuna da tresna hauek ondo erabiltzea».

Teknologia garatzeak bat-batean hiru produktu antzeko kaleratzea eragin du. Cortesentzat, «gauza normala» da hori, eta bateragarria iruditzen zaio elkarlanean aritzea eta lehiakorrek izatea: «Batzuetan elkarlana egingo da, eta beste batzuetan lehiatu egingo dugu».

Sareko BERRIAzalea:

Irakurri berri duzun edukia eta antzekoak zure interesekoak badira, eskari bat egin nahi dizugu: *Berria* diruz babestea.

Zuk eta zure gisako sareko milaka irakurlek egindako ekarpenarekin, eduki gehiago eta hobeak sortuko ditugu. Eta, zuekin osatutako komunitateari esker, publizitateak eta erakundeen laguntzek bermatzen ez diguten bideragarritasuna lortuko dugu.

Euskarazko kazetaritza libre, ireki eta konprometitua eskaini nahi dizugu egunero; bizi zaren munduaren eta garaiaren berri ematen segitu.

[Izan BERRIAlaguna](#)

Gaiak

Albiste gehiago



[Munduko aberatsak txiro irudikatu ditu artista batek, adimen artifiziala erabilia](#)

Edu Lartzanguren

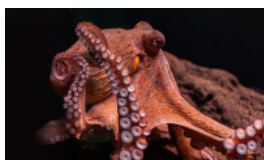
Besteak beste, Donald Trump, Bill Gates, Mark Zuckerberg, Jeff Bezos eta Elon Musk Indiako txabola auzoko biztanle gisa irudikatu ditu Gokul Pillai sortzaileak.



«Jazz musika ez dago zertan ulertu; sentituta nahikoa da»

Enekoitz Telleria Sarriegi

'Victor de Diego on Trane, the music of John Coltrane' da musikari bilbotarraren azken proiektua, Coltrane handiari egindako omenaldia: «Gorputz osoa astintzen dit; sekulako inpaktua eragin zidan ariman».



Zortzi garro, eta hamaika galdera

Enekoitz Telleria Sarriegi

Galiziako enpresa batek munduko lehen olagarro etxaldea ireki asmo du Kanarietan. Zientzialariek animalia horren zikloa ixtea lortu dutenez, eskariari aurre egiteko modua ikusi dute horrela. Elkarte ekologistek diote «izugarrikeria bat» dela.



«Baliabide gutxirekin ere saio handiak egin daitezke»

Itsaso Jauregi

Hamaika telebista kateko 'Geltokia' programa aurkezten du Ganboak. Elikadura burujabetzari eta ekonomia sozialari buruzko saio bat da, eta errezeta batekin biribiltzen dute.

Astekaria

Asteko gai hautatuekin osatutako albiste buletina. Astelehenero, ezinbesteko erreportajeak, elkarrizketak, iritziak eta kronikak zure posta elektronikoan.



Andoni Imaz