

Zubi bat ulermenera .net

Mikel Peruarena :

Okerragoa izan zitekeen. Hobeagoa izan zitekeen. Nondik begiratzen den. Denetariko iritziak bildu ditu Eusko Jaurlaritzak aurreko astean aurkeztutako gaztelania-euskara itzultzaile automatikoak (*itzultzailea.euskadi.net*). Proiektuan esku hartu duten Joseba Abaitua Deustuko Unibertsitateko irakasleari eta Juan Alberto Alonso Lucy Software enpresako arduradunari, asetzeko modukoa iruditzen zaie. Abaitua: «Espero nuen baino emaitza hobek izan ditu». Opentradek eta Googlek baino emaitza hobek sortzen dituela uste du. Alonso: «Espero zitekeena kontuan izanda, emaitza ona da». Kepa Sarasola IXA taldeko kide eta Euskal Herriko Unibertsitateko irakasleak ñabardura erantsi du: «Emaitzak ez dira hain txarrak. Perfektua, ez da. Baina erabilgarria da». Baliagarritasunari eman dio garrantzia, pentsatuz zuzenean argitaratzeko moduko itzulpen automatikorik euskaraz sortzea ia ezinezkoa dela gaur-gaurkoz. Karlos Del Olmo itzultzaileak zalantza gehiago ditu: «Emaitza ez da hain ona. Emaitza sendoagoak lortzeko, corpus sendoagoa behar da».

Adituak bat datoz: itzultzaile automatiko guztiek muga argiak dituzte. Euskararentzat, bi dira zailtasun handienak: inguruko hizkuntzekin dituen ezberdintasun morfologiko eta gramatikalak batetik; itzulitako testuen corpus txikia edukitzea bestetik. Gainera, muga teknologikoak ere badira; ordenagailu bidezko itzulpen automatikoak egiteko programek bi teknologia erabiltzen dituzte nagusiki: erregela gramatikaletan oinarritzen da bat, eredu estatistikoan bestea. Biek dituzte arazoak itzulpen «perfektuak» lortzeko. Jaurlaritzak aurkeztutako itzultzaileak (Lucy) eta IXA taldeak 2006an aurkeztutakoak (Opentrad) eredu gramatikala dute oinarri; Googlerenak, eredu estatistikoa.

Itzultzaile estatistikoetan, itzuli nahi dena erreferentziako testu corpusetan bilatzen du makinak; corpus horietan aurkitutako erreferentzien arabera, makinak estatistikoki ebazten du itzulpen egokiena. Ondorioa argia da: zenbat eta corpus txikiagoa, orduan eta itzulpen kaskarragoa. Ingelesarentzat, Googlek mila milioi hitz baino gehiago izango ditu; 10.000 nobela, denak itzulita. «Euskara oso urruti dago kopuru horietatik, eta konparazio estatistikoari kalte egiten dio horrek, konparatzeko eredu gutxi daudelako», Kepa Sarasolaren hitzetan.

Itzultzaile horiek *n-gram* izeneko algoritmo baten arabera *ikasten* dute, itzulpenak hobetzeko, erabili ahala. Esaldiak bi edo hiru hitzeko multzotan zatitu, eta corpusarekin erkatzen ditu programak. Horren mugak azaldu ditu Joseba Abaituak: «Gertatzen da euskarak egitura oso ezberdina duela. Eta itzulpeneko elementuen ordena asmatzea oso zaila da itzultzaile automatikoentzat». Ordena bakarrik ez, hitz bakoitzaren deklinazioa ere ebatzi behar du makinak. Gorka Labaka IXA taldeko ikerlariak tesia aurkeztu zuen, duela bi urte, Googleren teknologia euskararen itzulpenera egokitzeko proposamenak eginez. Norabide horretan lanean ari da IXA, pentsatuz sistema estatistikoak ekarpen inportanteak egin diezazkiokeela itzulpengintza automatikoari.

Itzultzaileen 'ulermen' mugak

Aldiz, eredu gramatikaletan oinarritutako itzultzaileei etorkizun lausoagoa ikusten diete adituek. Corpusik baliatu gabe, itzuli nahi den esaldia gramatikalki aztertzen dute itzultzaile horiek, hitzak euskarara itzultzen dituzte, eta arau gramatikalen arabera erabakitzen dute euskarazko itzulpenaren ordena eta deklinazioa. Hortaz, esaldiak *ulertzeko*, anbiguitateak argitzeko, deklinazioa eta ordena ebazteko... arazoak dituzte. Muga horiez aspaldi ohartu zela azaldu du Abaituak: «1990eko hamarkadarako argi ikusi nuen erregelen bidez aurrera egitea oso konplexua izango zela». Horregatik, Jaurlaritzaren itzultzaileak erregeletan oinarrituta egiten duena asko dela uste du: «Erregela bidezko sistementzat, muga bat jarri du Lucyk. Hori gairiditzezotan, sistema hibridoak garatu beharko dira». Juan Alberto Alonsok ere sistema hibridoetan ikusten du etorkizuna, eredu gramatikala eta eredu estatistikoa, biak baliatuko dituzten itzultzaileetan.

Lucyk ez du sistema hibridoa, baina hobekuntzak egin daitezke, Alonsok dioenez: «Memoria banku bat txertatuko balitzai, aukera egongo litzateke itzulpenak hobetzeko». Jaurlaritzak asmoa badu, izan, Memoria Banku Publiko bat sortzeko, eta Lucyk baliatu ahal izango luke. Memoriaren Banku Publiko hori Jaurlaritza ari da egiten, eta Lucy garatu dutenek ez dute haren berri zehatzik oraingoz. Sarasolak eta Del Olmok ere ez dakite bankuak zer ekar dezakeen, zehazki zer izango den ez baitu azaldu Jaurlaritzak.

Corpusa handitu behar

Nolanahi ere, eredu gramatikaletan oinarritutako itzultzaileekin konfiantza handirik ez du Del Olmok: «Sistema hauek behar dute hizkuntzalari talde izugarri handi bat atzean, lematizazio oso zehatza egiteko». Eta hori inbertsioa da, bai denboran, bai dirutan. Aukeran, inbertsio hori corpusa handitzen egin behar litzatekeela uste du Del Olmok. Sistema estatistikoentzat eta hibridoentzat baliagarriago izango litzateke corpusa handitzea: «Apustu estrategiko bat egin behar da, corpusaren kantitatea eta kalitatea, biak handitzeko». Corpusa sendotu eta zabalduz gero, enpresak «berez» gerturatuko direla uste du, programak eta probak egitera. Memoria banku horretan hezur-haragizko itzultzaileek ere parte hartu behar luketela uste du Del Olmok, «aurrerapausoa» erabatekoa izan dadin.

Abaitua bat dator corpusa handitzearen beharra nabarmentze horretan: «Corpusa da orain garrantzitsuena. Dauden baliabide guztiak eskura jarri behar dira». Beraz, Jaurlaritzari eskatu dio datu publikoak askatzeko politikarekin jarraitzeko, baita hizkuntz baliabideen arloan ere: «*Open data* horri jarraituz, baliabide linguistikoak eskura jar ditzatela, horren gainean probak eta prototipoak egiteko». Lucy itzultzaileak baditu, esaterako, itzulpen erregela batzuk —gazelaniazko genitiboak *-ren* edo *-ko*, nola itzuli behar diren zehazten dutenak, erlatiboaren itzulpenari buruzkoak...—, hiztegi elebiduna...

«Eduki horiek formatu berezietan daude, baina, askatzen badira, posible da berrerabiltzea, eta interesgarria da hori egitea, besteek ere erabil ditzaten». Jaurlaritzaren beste baliabideak batzuk ere — itzulpen bankuak, Euskalterm...— askatzea komenigarri jo du, gainera, Abaituak. IXA taldeak eta Eleka enpresak Opentrad garatzeko erabilitako baliabideak askatzea ere interesgarria litzatekeela erantsi du.

Inbertsioaren egokitasunaz

Itzultzaile automatikoa sortzeko prozesua luzea izan da. Jaurlaritzak 2007an abiatu zuen; bost urte igaro eta 553.000 euro ordaindu eta gero iritsi da produktua —aurrez 50.000 euro gehiago gastatu ziren, bost enpresari eskatutako prototipoetan—. Nahiz eta produktuari kalitaterik ukatu ez, inbertsioaren egokitasunari buruz sortzen dira galderak adituen artean. «Hau egin zenean ez zen kontuan hartu

bazegoela proiektu bat, software librean oinarritua gainera», adierazi du Sarasolak, Openradi erreferentzia eginez. Del Olmok ere kezka azaldu du: «Horrenbesteko inbertsioa egin behar zen, merkatuan ja bazegoen zerbait aurkezteko?».

Itzultzailea sortzeko prozesua hasi zenean, duela bost urte, Jaurlaritzak aditu talde bat elkartu zuen. Haien artean ziren Del Olmo eta Abaitua. «Batzuek adierazi genien hobe zela corpus sendo bat sortzea eta administrazioak itzultzaileen sare bat sortzea, itzulpen corpus publiko potolo bat egiteko», azaldu du Del Olmok. Abaituak ere esan zuen itzulpen sistema bat egiten hasteak ez zuela zentzu askorik, tresna horien mugak ezagututa. Baina Jaurlaritzak tresna sortzeko gogo handia zuela zehaztu du, eta lehiaketa publikoa atera zenean Lucy Software enpresako kideekin, UZEIrekin eta Gizarbitekekin aurkeztea hobetsi zuen Abaituak, bere ekarpena egin nahian. IXA taldea ere aurkeztu zen lehiaketara, baina Opentradek ez zuen batere punturik eskuratu, software librean oinarritzen zelako.

Lucy Software Alemaniako enpresa baten filiala da, eta alde horretan ikusten du desabantaila handiena Sarasolak: «Alemaniako enpresaren morrontza ekarriko du, eta etorkizuneko garapenak enpresa horrekin etorriko dira». Bertan enpresarik eta teknologiarik ez balego bezala jokatu du Jaurlaritzak, Sarasolaren ustez. Etorkizunerako estrategia ondo pentsatu behar dela uste du, bestalde: «Aukeratu behar da zer estrategia nahi den: inizatiba hemendik eramatea, lizentzia irekiekin; edo Alemaniako enpresa batekin, eta lizentzia itxiekin».

IXA taldeak egindako lanaren balioa onartu du Abaituak, baina kritika horiekin ez dago ados. Bertako enpresa eta ikerlariak proiektuan parte hartu dutela adierazi du. Lucy Software enpresak abantaila bat zuela uste dute, gainera, Alonsok eta Abaituak: «Gaztelaniatzko analisi sintaktikoa egiteko tresna oso ona du. Eta oso inportantea da jatorriko hizkuntza ondo aztertzea».

Lucy Software enpresak badu esperientzia Valentziako eta Kataluniako administrazioekin. Gaztelaniaren eta katalanaren arteko itzulpenetan «kalitate oso altua» lortu dutela azaldu du Alonsok. Baina bestela da euskararekin. «Euskarak ez du parekotasunik. Euskararekin, edozein hizkuntza pare askoz konplexuagoa da. Ez zailagoa, diferentea baizik». Denbora gehiago behar da lanerako, eta «emaitza ez da hain ona», Alonsok aitortu duenez. Hala ere, kontentu dago Lucyrekin: «Tresna bat da, ezin zaio eskatu eman ezin duenik».

Testuak, beti orraztuta

Iragarri ere hala egin zuten. Lorea Bilbaok, 2008an Hizkuntza Politikarako Sailburuordetzako aholkulari zenak, hala esan zuen: «Badakigu eskainitakoa ez dela izango itzulpen perfektua, baina itzulpenerako laguntza bat izango da, itzulitako testuari buruzko ideia orokor bat eskainiko duena». Eta aurreko astean honela aurkeztu zuen Jaurlaritzak: «Askotan, euskarazko itzulpena zuzena izango da, edo ia zuzena, estilo kontuak alde batera utzita; beste batzuetan, ostera, euskarara egindako itzulpena zati batean bakarrik izango da zuzena». Eta oharra erantsi zuen: «Itzulpen automatikoko beste edozein sistema bezala, laguntza tresna da, eta ezin du inolaz ere giza itzultzailea ordezkatu».

Oharrak ohar, itzultzaile automatikoa mezu publikoetan erabili da dagoeneko, eta orraztu gabe. SOS Deiak zerbitzuko sare sozial bateko kontuan hiru mezu agertu dira, hirurak Lucyrekin itzulita, eta orraztu gabe. Azkena, atzo: «N-1 Istripua 433 pk-a zentzua Tolosa Irun. Zauritutako eramandako motor-gidaria Clinica Asunción. Bideko sarearen gaineko eraginik gabe». Horiek ziren itzultzailearen arriskuak, eta horretaz ohartarazi zuten Jaurlaritzako ordezkariak aurkezpenean.

Lehen egunetan erabiltzaileek egindako probak ere sonatuak izan dira. Alonsok azaldu du konpondu egingo dituztela akats batzuk —*egun on, gero arte...* oinarritzako esapide batzuentzako itzulpen egokiak sartuko dituztela adierazi du—. Itzultzaile automatikoekin halako «txisteak» beti ateratzen direla esan du Sarasolak. Baina ez dela hori garrantzitsuena zehaztu du: «Ulermena laguntzea da garrantzitsuena. Itzulpen ez-perfektu batek lagunduko dizu testuak zer esan nahi duen ulertzen. Ez du balio argitaratzeko, baina baliagarria da».

Orain, Lucyren beste adarrak garatzen ari dira. 2013ko udarako, euskara-gaztelania eta ingelesa-euskara itzultzaileak prest eduki nahi dituzte. Jaurlaritzak 1,6 milioi euroko inbertsioa egin du itzultzaile horietan eta Memoriaren Banku Publikoan. Deustuko Unibertsitateak, UZEIk eta Lucy Softwarek jarraitzen dute lan horretan. Euskaratik gaztelaniarako itzultzailea egiteko, euskararen aztertzaile gramatikala garatu behar dute, eta horretan izango dute lan gordinena. «Seguru asko, orain arte egin dugun zailena izango da», aitortu du Alonsok: «Baina, aldi berean, oso interesgarria da alde profesionaletik». Enpresako langileen inplikazioaren erakusgarri, Alonsok adierazi du Bartzelonako Euskal Etxean euskara ikasten aritu zirela.

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



[Martxan da Gasteizko tranbiaren Salbururako luzapena](#)

[Edurne Begiristain](#)

Lehen eguna du gaur tranbiaren adar berriak. ELAk eta LABek protesta egin dute ETSko langileek hitzarmenik gabe jarraitzen dutela salatzeke.



«Zuk %200 eman behar duzu beti»

Arantxa Iraola

Konta daiteke, batetik, Alaitz Ziarsoloren istorioa; haren gorreriarena, trabena. Bestetik, Noelia Dacostarena; paralsiarena, gutxiespen guztiena. Biak batera kontatuta, ordea, gardenago ikusten da dena: gizartea, elbarritasunak tarteko, motza da sarri, eragozpenez josia. Baita EHUn ere.



Grebara deitu dituzte Arabako landa eremuko osasun zentroetako langileak hilaren 27rako

Edurne Begiristain

ELAK deitu du lanuztera. Osakidetzaren lan berrantolaketak osasun arretaren kalitatea eta langileen baldintzak okertuko dituela salatu dute.



Moto gidari bat hil da Luzaiden, bidetik aterata

Berria

Errepidetik atera, eta zirkulazio seinale bat jo du. Ipar Euskal Herrirako noranzkoan zihoan, beste bi moto gidarirekin batera.

Astekaria

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

