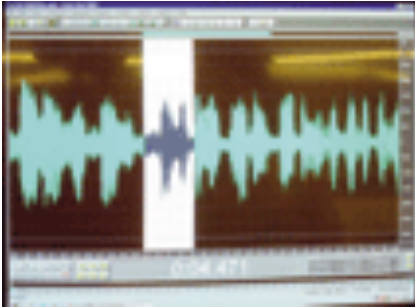


GIZAKI ETA MAKINEN ARTEKO KOMUNIKAZIOA

Xabier Telletxea

2003-05-18



Hizkuntza teknologiak funtsezkoak dira informazio eta komunikazioaren gizartean. Egun, badira zenbait hizkuntza aplikazio eskuragarri, hala nola, ortografia eta estilo zuzentzaileak, on-line hiztegiak, itzulpen laguntzaileak, Internetarako bilatzaileak, hizketa testu bihurtzen duten sistemak, testuak irakurtzen dituztenak... Sistema hauen eta gizakiaren arteko komunikazioa lantzen duen zientziari deritzo ingeniarietza linguistikoa, pertsona eta makinaren arteko komunikazioari, alegia.

Euskal Herrian badira ingeniarietza linguistikoa lantzen duten zenbait enpresa eta talde, besteak beste, Elhuyar, IXA, Eleka, Code & Syntax, VicomTech, Hizkia, Plazagunea, Ametzagaiña eta Aholab. Unibertsitate mailan ere, esaterako EHUKo Zientzia Fakultatean, Bilboko Industri Injineruen Eskolan eta Gasteizko Injineru Teknikoen Eskolan, bada honetan jarduten duenik.

Ingeniarietza linguistikoaren munduan lantzen diren esparru ezberdinak eta Euskal Herrian duten trataera eta ikerlana ezagutu xedez, hiru enpresetara jo dugu, hain zuzen ere Eleka, Aholab eta Ametzagaiñara. Teknologia linguistikoetan espezializatuak diren Eleka enpresan esaterako, Josu Waliño ikertzailearekin hitz egin dugu. Teknologia linguistiko estrategikoak garatzeko eredu aitzindaria da Eleka. Urteak daramatza ingeniarietza linguistikoaren arloan ikertzen, behar dokumentalei eta linguistikoei aurre egiteko teknologiak eta irtenbideak garatzen. Unibertsitateetan garatzen diren teknologietan oinarrituta eta produktuen gaineko lizentzien bidez bezeroen eskakizunen arabera tresnak garatzea da haien eginkizun nagusia.

Bi talde elkartzuz sortu den enpresa proiektua da Eleka: IXA Taldea eta Elhuyar Fundazioa. Errendimendu Handiko Taldea izendatu du Eusko Jauriaritzak EHUko Informatika Fakultateko IXA Taldea. Talde horrek 15 urte eman ditu ingeniari-tza linguistikoa-ren arloan. Hainbat produktu garatu dituzte, esaterako, Xuxen (Worden txertatutako euskarazko zuzentzaile ortografikoa), lematizazioan oinarritutako Interneteko bilatzaileak eta hainbat tresna linguistikoa. Bestalde, Elhuyar Fundazioak hizkuntzaren esparruan hogeita hamar urteko esperientzia du. Denbora honetan zehar, Euskal Herriko hizkuntzalaritzaren arloan enpresa aitzindaria bihurtu da Elhuyar.

Josu Waliñoren ustez, ingeniari-tza linguistikoa hizkuntza eta informazioa informatikaren bidez tratatzea da. Bi arlo ezberdintzen dira zientzia honetan: Testuaren tratamendu informatizatua, hau da, lengoia naturalaren prozesamendua eta bestetik, ahotsaren tratamendua. Azken honen barruan bi arlo bereizi behar dira, ahotsaren sintesia, hots, ordenagailuak hitz egitea, eta ordenagailuak zuzentzea edo transkribatzea. Waliñoren aburuz, ezagutzaren tratamendua da ingeniari-tza linguistikoa-ren helburua, maila guztietan. Informazioa ezagutza da, era testualean nahiz ahotsaren bitartez eman daitekeena. Beraz, informazio hori jaso eta ezagutza tratatzeko aukera izatea da xedea. Honekin lortzen dena da pertsonak egun behartuta egitera gauden hainbat lan ordenagailuek era automatikoan egitea; itzulpena esaterako. Itzulpen automatikoa-ren helburua sistema bat testu bat bere testuinguruan ulertzeko gaitzatea eta itzultzea izango litzateke. Ahots errekonozimenduan beste horrenbeste izango litzateke, ezagutzaren tratamendua.

Nazioartean informatikaren lehen pausoekin hasi ziren ingeniari-tza linguistikoa lantzen. Euskal Herri mailan aldiz, Informatika Fakultateko IXA taldea hasi zen arlo hau jorratzen duela 15 urte. Beraz, EHU-n hasi zen ingeniari-tza linguistikoa lantzen eta ikertzen. Ondoren, IXA taldeak eta Elhuyar Fundazioak Eleka sortu zuten ingeniari-tza linguistikoa enpresetara eramateko. Memento honetan, zortzi laguneko ikerketa talde batek euskarazko zuzentzaile gramatikala lantzen dihardu. Urtebete barru amaitua egotea espero dute. Orain arte Elekaren testuaren tratamendu informatizatua landu badu ere, Elhuyarrek, Aholabekin, IXA-rekin, Robotikerrekin eta VicomTechekin bat eginik, ahots errekonozimendua landuko du lehen aldiz. Hain zuzen ere, ahotsaren bitartez

euskarazko bilaketak egiteko proiektua dute. «Hizking 21» izeneko hiru urteko proiektu honen helburua euskara ingelesaren mailara iritea da.

Euskara informatikoki lantzeko zailtasunak

Euskarak informatikoki landua izateko bi zailtasun nagusi ditu. Alde batetik morfologia, hau da, hizkuntzaren berezitasun linguistikoa. Hizkuntza osatzeko garaietan, adibidez, zuzentzaile ortografiko bat egiten bada, gaztelaniar hiztzerrenda mugatu bat nahiko da. Aitzitik, euskaraz zerrenda hori mugagabea bihurtzen da hitzen aldaera posible guztiekin eta deklinabideekin. Beraz, arau linguistikoetan oinarrituz funtzionatu behar da. Bestalde, ingelesa eta gaztelaniarrekiko, oinarritzko baliabideetan oso atzeratua dago euskara. Garrantzitsua da ingeniari-tza linguistikoetan lan egiteko corpus handiak izatea, bai ahotsean eta bai testuan. Josu Walfiñok dioen moduan, "uzta jasotzeko eroin egin behar da eta guri ereteke lur hori falta zaigu. Lur hori eraikitzen dihardugu orain, ondoren aplikazio indartsuak lortzeko. Zailtazari gabe, ingelesa da teknologia hauetan hobereen kokatuta dagoen hizkuntza".

Ahots errekonozimendua

Ahots errekonozimenduan gehiago sakondu nahian, Gasteizko Ingeniari Teknikoen Eskolan Sistemen Ingeniaritza eta Automatika sailko koordinatzaile eta Aholab ikerkuntza taldeko kide den Karmele Lopez de Ipiñaren jarduera diugu. "Euskal Herrian nahiko aberatsa da ingeniari-tza linguistikoan ahotsaren errekonozimendua, Unibertsitate munduan nahiz enpresa munduan", baieztatu digu Lopez de Ipiñak. Euskal Herrian, besteak beste, IXA taldea eta Aholab taldeak dihardute ahots errekonozimendua lantzen. IXA taldeak lengoai naturala lantzen du gehienbat eta Aholab-ek, aldiz, hizketaren tratamenduan dihardu. Azkeneko honetan bi arlo ezberdindu beharko litzaiteke: batetik, ahots bihurtzea eta bestetik, ahots ezagutzea. Lehenengoa konputagailuak hitz egiteko izango litzateke, bestea ordea hiztunak egiten duen mezua ulertzeko. Enpresa munduan, lehen aipatu legez, Eleka enpresak dihardu honetan. Telefonikak, Euskaltelek, Fagorrek... ere zenbait lan txiki egin dituzte arlo honetan.

Katalunian kokatuta dagoen ATLAS (Applied Technologies on Language and Speech) enpresak esaterako, euskaraz ataza sinpleak ezagutzeko sistema garatu du. Honekin lortzen dena da konputagailuari hitz egitea, berak mezu bat ulertzea, transkribatu edota hitz gako batzuen bitartez bilaketa batzuk egin edota gailu bat aktibatzea. Aholab taldea berriro hizpide hartuta, testuak transkribatzeko prototipo bat lantzen ari dira, ikerkuntza mailan betiere. Talde honek euskara ere badu erreferente, bestelako aplikazioak ezagutzeko zenbait ataza landu dituzte, besteak beste domotikan aplikatzeko, elbarrientzako aplikazioak... Memento ez erkomertzialik ez badute ere, ikerkuntzan zenbait aplikazio garatu ditu Aholab taldeak. Finantziario aldetik, tresna hauek garatzeko datu base onak behar direnez inbertsio handiak egin behar direla jakinarazi digu Karmele Lopez de Ipiñak. "Datu guztiak bildu, etiketatu eta tratatzeko baliabideak behar dira eta finantziario ezinbestekoa da. Nire ustez erakundeak sartuko dira finantziario honetan eta hiru urte barru egoera aldatzea espero dut. Izan ere, honek beste sistema batzuk garatu eta aurrerapen, ikerkuntza eta baliabide gehiago edukitzea ekarriko bailuke. Memento honetan bilaketak egiteko finantziario eta denbora ez da hain handia. Urte honetan behingoz zerbait operatiboa eta erabilgarria edukitzea espero dugu".

Hizkuntzaren tratamenduari dagokionez, bat dator Aholabeko kidea Josu Waliñorekin, hots, euskara informatikoki landua izateko zailtasun nagusia hizkuntzaren egitura morfologikoa dela adierazi digu: «Zailtasun honen aurrean beste eredu batzuk aurkitu behar izan ditugu, esaterako turkiera. Ahotsa eta hotsak grabatzeko ez dugu batua bakarrik erabili, euskalki guztiak baizik, batez ere hots ezberdinak egon daitezkeelako».

Naturaltasuna da sistema hauen helburua. Orain arte eginiko sistemak deserosoak izan dira erabiltzailearentzat. Memento honetan, sistema eleanitzak, erosotasuna eta erabilgarritasuna dira nazioarte mailan gehien lantzen ari diren esparruak. Horretarako, baliabideen optimizazioa hartzen da kontuan. Euskal Herrian ordea, helburu horiek murriztu egiten dira, helburu "errealagoak" finkatuz. "Ikerkuntza taldean nazioarteko kongresuetara mugitzen gara, ezinbestekoa baita kanpoko jendearekin harremanak mantentzea sistemak garatzeko eta eurengandik ikasteko. Askotan euskarak baliorik ez duela iruditzen bazaigu ere, gure lanak nazioartean erakustean, euskararen egitura berezia duenez, jendea asko hurbiltzen zaigu. Guk esaterako, turkiera eta japoniera ditugu eredu bezala, erabiltzen duten egitura antzekoa baita", azaldu digu Karmele Lopez de Ipiñak. Bere esanetan, gu ez gaude nazioartekoen mailan, finantziazioa eta baliabideak falta baitzaizkigu. Dena den, egun, erakunde eta enpresen aldetik interes handia omen dago.

Ahots errekonozimenduaren barnean, robotizazioa eta domotizazioa ere aurkitzen dira. Batak robotek agindupean gizakiak esanikoa burutzea bilatzen du eta besteak berriz, etxeko gailuek gure ahotsaren aginduak betetzea. Sistema hauek errazagoak dira, hitz sinpleak direlako. Arazoak ordea, aplikazioa garatzeko modua bilatzerakoan sortzen dira. Hizketa ezagutza ez da oso konplexua, zailtasuna arlo teknikoan dute gehienbat.

Etorkizunari begira, "nazioartean dituzten tresnak Euskal Herrian edukitzea izango litzateke ikerlarion ametsa", adierazi digu Aholabeko kideak. Tresna ugari daude, telefono bidezko aplikazioak esaterako. Elkarrizketa sistema batzuk beste hizkuntza batzuetan garatuta daude. Katalunian esaterako, eguraldia kontsultatzeko zerbitzu hau eskaintzen da. 906 789 987 zenbakira deitu eta sistema automatiko batek hartzen du deia. Lengoaia naturalaz baliaturik eguraldiari buruz informazioa eskatu dezakezu. Sistema hau Kataluniako Unibertsitate Politeknikoko TALP ikerketa taldeak garatu du. Antzeko sistemak

garatu dituzte Telepizza eta Renfek, besteak beste.

Hizkuntza idatziaren tratamendua

Euskal Herrian ingeniari-tza linguistikoa lantzen duten beste enpresek esparru honen inguruan zer egiten duten jakin nahian, teknologia ikertzen eta produktuak garatzen diharduen Ametzagaña taldeko Josu Landa ikertzailearekin mintzatu gara. Indexatzaile, bilatzaile eta lematizatzaile propioa lortu ditu talde honek. Industria saillean hiru urteren buruan memoriaren oinarrituriko Ametra izeneko itzulpen-erako laguntzaile bat lantzen dihardutela jakinarazi digu. "Hizkuntza idatziaren azterketan dagoen teknologia eta teknologia horrek ematen duen maila nahiko eskasa da». Hori gainditzeko eman beharko litzatekeen urrats kualitatibo nahiko urrun ikusten du Josu Landak, baina bere aburuz horrek ez du esan nahi bitartean gauza interesgarriak eta baliagarriak egin ezin direnik. «Kalitatezko itzultzaile automatiko aplikaziorik ez dago egun. Jendea oso zorrotza da eta itzultzaile automatiko bati itzulpen onak egitea eskatzen dio. Horrek ordea zapuzketa dakar, ondorioz jendeak horretan sinesteari uzten dio. Horixe bera gertatu zen duela hamar bat urte ahotsaren ezagupenarekin. Jendea produktu horrekiko zapuztu egin zen, uste baino emaitza kaxkarragoak eman baitzituen. Egun, oso diktatu programa onak daude baina ez dira erabiltzen". Horretaz gain, jendearen beharrak zeintzuk diren jakitea oso zaila dela jakinarazi digu Landak. Hori dela-eta, garapen eta produktuak euren ustez interesgarria izango litzatekeenaren bila egiten dituztela adierazi digu.

Garapen eta tratamendu egokiak lortzeko beraz, gai honetan ikerle eta aditu direnek gizarte eskariari begiratu behar zaiola diote. Sistema eta aplikazio horiek lantzeko eta helburu hori lortzeko ordea, ikerketa taldeek, industriak eta erakunde ofizialek koordinatu egin behar dute. Noiz izango dugu posible teknologia guztiak lengoia naturalaren bitartez maneatzea? Finantziarioak eta dedikazioak izango dute horren erantzuna. Hori da euskarak teknologien munduan bere lekua izateko eta gure hizkuntzari esparru honetan etorkizun bat ziurtatzeko modu bakarra.