











 Gorago

 Zeu berriemaile

 Bestelakoak
 Enpresa
 Herri-administrazioa
 Hizkuntza
 Informatika
 Internet
 Irakaskuntza
 Kultura
 erabili.com



Hemen zaude: [Sarrera](#) » [Sare-aldizkariak](#) » [Berri berriak](#) » Euskaraz egiten diren akatsak aztertu dituzte, zuzentzaile automatikoetan eta euskara ikasteko programetan aplikatzeko

Berri Berriak

Euskaraz egiten diren akatsak aztertu dituzte, zuzentzaile automatikoetan eta euskara ikasteko programetan aplikatzeko

2010-03-16 / 08:45 / EHU

 **INFORMATIKA**

EHUko Informatika Fakultateko IXA taldeak urteak daramatza ikerketa-lanetan, euskararen onurarako lirakeen sistema (erdi)automatikoak garatzeko. Sistema horien artean egongo lirakeke, esaterako, euskarazko akatsen tratamendu automatikoa eta ordenagailuz lagunduta euskara ikastea ahalbidetzen duten bitartekoak.

Larraitz Uriak IXA taldeko kideak bi sistema horiek garatzeko zenbait oinarri ezarri ditu EHUn aurkeztu duen doktore-tesian, euskarazko akats eta desbideratzeen analisirako irizpide batzuk finkatuta.

[Euskarazko errorearen eta desbideratzeen analisirako lan-ingurunea. Determinatzaile-errorearen azterketa eta prozesamendua](#) izena du Uriaren doktore-tesiak. Akatsak eta desbideratzeak ondo bereizi nahi izan ditu lehenik eta behin, eta hori da ikerketaren ekarpenetako bat. Akatsak ortografia edo gramatika bezalako kontuei lotuta daude. Desbideratzeak, berriz, gramatikalki zuzenak baina testuinguru jakin batean desegokiak diren hitzak dira; erregistroari edo euskalkiari lotuta daude. Etorkizuneko sistema automatikoei bi horiek desberdintzea da asmoa, eta bereizketa garrantzitsua da, beraz.



Euskarazko akats eta desbideratzei buruzko adibide eta xehetasunak biltzen hasi dira bi datu-basetan, eta horren berri eman du Uriak bere tesian. IXA taldeak jarri ditu martxan, eta bi aplikazio zehatzetarako egokitu. Euskararen akatsen tratamendu automatikoa (zuzentzaileak, dialektoen markatzaileak...) garatzeko behar den informazioa biltzeko da lehena, eta ordenagailuz lagundutako hizkuntza-i(ra)kaskuntzarako tresnak sortzeko datuak jasotzeko bigarrena. Bi ikerketa-alor horiek uztartzea erabat ezohikoa da, baina Uriak dioenez euskararen akatsen tratamendu automatikorako balio duten datuetako asko erabilgarriak dira ordenagailuz lagundutako hizkuntza-i(ra)kaskuntzarako, eta alderantziz. Horixe da lan horrek aurkezten duen nobedade eta ekarpen nagusietako bat.

Akatsen detektagailua garatzeko ezinbestekoa

Tesiaren beste ekarpenetako bat dagoeneko abian den corpusa da, datu-basearen oinarria, alegia. Bertatik ari dira ateratzen akatsen eta desbideratzeen lehen adibideak, horiek detektatzeko gai den sistema bat garatzeko ezinbestekoak. 113.290 hitzeko corpusa osatua dago jada, hainbat mailatako euskara-ikasleen testuen bildumarekin. Euskara teknikoko ikasleen eta hitzun arruntzen testu batzuk ere txertatu dira bilduman. Lehen urrats honekin, analisiarekin hasteko informazio garrantzitsua finkatu ez ezik, corpusa osatzeko irizpideak ere definitu dira.

Corpusetan dauden adibideen etiketatzea da hurrengo urratsa. Doktore-tesi honetan, eta

ikerketaren abiapuntu gisa, mugatzaileei lotutako akatsak etiketatu dira gehienbat. Euskaraz mugatzaileei dagokienez akatsak egitea ez da oso ohikoa, baina aldi berean egiten denean oso akats larria izaten da. Horregatik, adibide egokitzat jo du Uriak lehen proba moduan. Hala ere, gerora akats eta desbideratze guztiak detektatzeko gaitasuna garatzea da asmoa. Etiketatze-prozesurako EtikErro editorea izan du lagun, IXA taldeak berak sortua. Akatsak etiketatzeaz gain, etiketatutako adibideak datu-baseetara esportatzen ditu, hizkuntza-azterketa egiteko informazio linguistikoa eta guzti.

Etiketatzearen ondoren datorren sailkapen-fasean ere ekarpen garrantzitsua egin du ikerketak. Sailkapenaren egitura nagusia definitu du, bereziki mugatzaileei buruzko akatsei dagokien kategoria landuz. Aurreko fase horiek beteta ekin zaio bi datu-baseak osatzeari. Adibide eta informazio linguistiko bera gordetzen dute biek, baina desberdintasunak ere badituzte. Euskarazko akatsen tratamendu automatikorako datu-baseak informazio teknikoa ere badu. Ordenagailuz lagundutako hizkuntza-i(ra)kaskuntzari dagokionak, bestalde, informazio psikolinguistikoa biltegitratzen du.

Tratamendu automatikoaren lehen emaitzak

Uriak, IXA taldearekin batera, eginak ditu lehen probak, bitarteko horiek erabilia akatsen tratamendu automatikoak ematen dituen emaitzei dagokienez. Mugatzaileekin egindako akatsetarako bereziki egokitutako teknika eta erregela-multzoa erabilia, tratamenduaren doitasuna neurtu du. Hau da, programa informatiko baten bidez tratamendua zenbateraino den eraginkorra ikusi du. Doitasuna % 45,5ekoa besterik ez zen izan hasiera batean. Hala ere, etiketatu gabeko beste akatsak aurrez kenduta, "zarata" desagertu egin zen, eta doitasuna % 80ra igo zen. Horrenbestez, finkatutako irizpideek eraginkor izateko zantzuak erakutsi dituzte. Corpusa handituta, eraginkortasuna handiagoa izango dela uste du Uriak. Bere tesiaren ekarpena lehenengo urratsa besterik ez da, etorkizuneko erronka bati begira.

Egileari buruz

Larraitz Uria Garin (Hernani, 1977) Ingeles Filologian eta Lehen Hezkuntzako irakasle-ikasketetan lizentziatua da. Igone Zabala Unzalu Euskal Filologia Saileko irakaslearen eta Montse Maritxalar Anglada Informatika Fakultateko irakaslearen zuzendaritzapean egin du tesia. Gaur egun, ikertzailea da EHUko IXA taldean eta Baionako Unibertsitateko IKER taldean.

 [Inprimatu](#)

[Erantzun](#)



Euskara Patronatua
Tel.:943-814518
Faxa: 943-811947
erabili@erabili.com

