

Bost tesi berri euskara automatizatzeko

Informazioa erauztea eta itzulpen automatikoa ditu xede Donostiako Informatika fakultateko Ixa taldeak

Ainara Mendiola / Donostia

Donostiako Informatika fakultatean sorturiko Ixa taldeko bost kide euskararen sintaxia automatizatzeko tesiak lantzen ari dira, eta lehena joan den asteazkenean aurkeztu zuen Koldo Gojenola EHUko Informatika irakasleak. Euskararen automatizazioa da taldearen helburu nagusia, eta etorkizunean horretarako aplikazioak sortzea.

HAMABI urte dira Ixa taldea sortu zela, euskararen tratamendu informatikoa helburu. Hasieran, itzultzaile informatiko bat zergatik ez zuten egiten galdetzen zion jendeak taldeari, baina Kepa Sarasola Informatikan doktore eta Ixa taldeko kideak —Gojenolaren tesiaren zuzendari ere bada— EGUNKARIARI esan dionez, «etxea teilatutik hastea» izango zatekeen. Euskararen morfologia aztertzeari ekin zioten, eta sintaxia izan da bost tesi berrien lan esparrua. Lexikoa eta semantika alde batera utzi gabe, aurreko bi ildoak jorratu dituzte sakonen orain arte.

Morfologia, sintaxia, lexikoa eta semantika dira Ixaren ikergaiak, haiekin hizkuntza baliabideak eta hizkuntza lantzeko tresnak sortzeko eta azkenean aplikazioak erdies-teko. Nahiz Xuxen zuzentzaile ortografikoa den taldearen lan ezagunena, beste ekarpen batzuk ere egin dituzte: Euskarazko datu base lexikala (EDBL), analizatzaile morfologikoa, lematizatzailea... Lematizatzailea da, hain zuzen, Nerea Ezeizaren tesiaren ekarpen nagusia, egilearen beraren hitzetan.

Euskararen datu base lexikalaren laguntzaz egin dute lan Koldo Gojenolak eta beste lau doktore gaiek: Itziar Aduriz, Jose Mari Arriola, Nerea Ezeiza eta Izaskun Aldazabal. Gojenola eta Ezeizak alde teknikoa landu dute, eta hizkuntzalaritza besteek. Gainera, Ibon Sarasolaren *Euskal Hiztegia-z* baliatu dira, eta baita beste testu multzo batez ere.

EDBL eta EGUNKARIARI sarea dituen testuak erabili dituztela aitortu zuen Kepa Sarasolak. «Pena da gehiago ez edukitzea. Instituzioek euskararen korpus nazionala antolatu behar dute. Iturburu ikaragarria izango da ikerketarako», esan zuen.



Itziar Aduriz, Koldo Gojenola, Nerea Ezeiza eta Izaskun Aldazabal —Jose Mari Arriola falta da— ari dira euskal sintaxia automatizatzeari buruzko tesiak egiten.

Bostak bat

Tesiei dagokienez, koordinazio lana azpimarratu zuen Sarasolak. «Estrategia bat daukagu, eta estrategia horren barruan daude bost tesiak». Estrategia hori egokitzen du hizkuntza gutxituak lantzeko.

Hitz bakoitzak ematen

duen informazioa —lemak zein aurrizki, atzizki eta kasu markek— bildu du Gojenolak, Arriolak aditzen azpikategorizazioa egin du, Adurizek desanbiguzioa —hitz bakoitzaren analisi egokia lortzeko—, testu masa handietan bilaketak egiteko tresnak sortu Ezeizak eta Aldazaba-

lek guztien lana borobilduko du: «Aditzen azpikategorizazio informatikoa ondo aztertu, ikusteko erabilera batzuk okerrak direla eta zeintzuk diren aditz horiekin etorri behar diren kasuak. Informazio hori biltzea ezinbestekoa da aplikazio informatikoak egiteko», Sarasolaren arabera.

Urtebete barru inguru amaituko du tesia Izaskun Aldazabalek. Besteek, berriz, aurten bertan, eta Jose Mari Arriola irailaren 15ean da bere lana aurkeztekoa.

Hiztegi elektronikoa eta Wordnet

Euskal Hiztegia-rekin aditzen azpikategorizazioa egin du Arriolak, eta hiztegiak egiteko modua aldatzera bultzatzen du horrek, haren ustez: «Ondorio bat izan da hiztegiaren lan formalagoak, hobeto egindakoak, lor daitezkeela. Informatikaren bidez era taxutu eta formalean egitea hiztegiak eta ez, adibidez, prozesadore arrunt bat erabiliz». Hiztegi horrek bilaketa zehatzagoak egiteko aukera emango luke, esaterako; adibide jakin bat bilatu edo hitz bat edo jatorria.

Elhuyarren euskara-gaztelania hiztegia ere lantzen ari da Ixa taldea, eta baliteke udazkenerako edo datorren udaberriko aplikazio informatiko bat sortzea hitz bakoitzaren euskarazko edo gaztelaniazko ordaina automatikoki lortzeko.

Semantika alorrean *Wordnet* programaren euskarazko baliokidea egin nahi du Ixak; hots, euskarazko kontzeptuak hierarkikoki antolatu, gero dedukzioa egiteko.

'Cum laude' katearen lehen mailari

CUM LAUDE kalifikazioa eman zion, joan den asteazkenean, epaimahaiak Koldo Gojenola Informatika Fakultateko irakaslearen doktorego tesiari. *Euskararen sintaxi konputazionalerantz. Oinarriko baliabideak eta beren aplikazioa aditzen azpikategorizazio-informazioaren erauzketan eta erroren tratamenduan* izenburupean aurkeztu zuen lana Gojenolak. Epaimahaia ondokoek osatu zuten: Beñat Oihartzabal eta Patxi Goenaga euskaltzainek, EHUko Informatika irakasle Xabier Artolak, Informatikan katedradun Arantxa Illarramendik eta Andoni Sagarna injineruak. Goenaga izan zen epaimahai buru.

«Esango nuke euskararen tratamendu informatikoaren bide luzean (tesiak) aurrerapen handia dakarrela», adie-

razi zuen Xabier Artolak. Zuzentzaile gramatikala merkaturatzeko zein aukera ikusten duen galdetu zion Gojenolari, eta esan zuen «lan handia» emango duela.

«Egin ditugun tresnekin testu idatziak emanda eza-gutu ditzakegu dauden unitate sintaktiko nagusi batzuk, esaldiak edo izen sintagmak. Eta hori egiteko gauza bagara, aplikazioak landu daitezke, adibidez azterketa linguistikoak egiteko», adierazi zion Gojenolak EGUNKARIARI. Errore sintaktikoen tratamendua izan da taldean landu duten beste alorretako bat, eta Nerea Ezeizaren arabera, bakarrik %2-3koa da tresna horiekin errorea egiteko probabilitatea.

Koldo Gojenolak zuzentzaile gramatikala egiteko oinarriak jarri dituela esan



PABLO S. QUIZA

zuen Sarasolak, tesiaren zuzendariak. Kepa Sarasolaren arabera, «oinarriak jarrita daude, eta hemendik bi-hiru urtera atera daiteke gauza politen bat; itzulpen automa-

tikoak egiteko horrelako tresnak funtsezkoak dira, hizkuntzak irakasteko ere bai. Eta informazio bilaketa eta erauzketarako lagungarria izango da».