

EUSKARAZKO ERRORE SINTAKTIKOAK DETEKTATZEKO ETA ZUZENTZEKO BALIABIDEEN GARAPENA: DATAK, POSTPOSIZIO-LOKUZIOAK ETA KOMUNZTADURA

Tesiaren egilea: Maite Oronoz Anchordoqui

Unibertsitatea: Euskal Herriko Unibertsitatea

Saila: Lengoia eta Sistema Informatikoak

Tesi-zuzendaria: Arantza Díaz de Ilarraza Sánchez eta Koldo Gojenola Gallettebeitia

Tesiaren laburpena:

Euskararen egoera dela-eta (estandarizazio-prozesuan dago, aberatsa da euskalkiei dagokionez eta erdara indartsuek inguratzen dute), ohikoa da erabiltzaileok Euskaltzaindiak ezartzen dituen arauetatik aldentzen diren hizkuntza-egiturak erabiltzea. Egitura zuzenetatik, arauetatik, desbideratzen diren egitura horiei errore deitu diegu. Tesi-lan honen helburua da euskarazko errore sintaktikoak detektatzeko eta zuzentzeko baliabideak garatzea. Errore sintaktikoak automatikoki detektatzeko eta zuzentzeko sortu ditugun sistemak, erabilgarriak izango dira euskararako gramatika-zuzentzailean eta ordenagailuz lagundutako hizkuntzen irakaskuntza adimenduneko sistemetan (OLHIA).

Tesi-lanean, datetan, postposizio-lokuzioetan eta, batez ere, subjektu, objektu eta zehar-objektu funtzioa betetzen duten osagaien eta aditzaren arteko komunztaduran gertatzen diren erroreak detektatu ditugu. Errore horien bila corpus erroredunak zein zuzenak automatikoki analizatzeko IXA taldeko (<http://ixa.si.ehu.es>) analisi sintaktikorako katea erabili dugu, eta, informazio linguistikoa erreprezentatzeko, XML bidezko anotazio-amaraura.

Erroreak detektatzeko eta zuzentzeko erabiltzen diren teknikei gainbegiratu zabala egin ondoren, hizkuntza-ezagutzan oinarria duten teknikak (sinbolikoak) erabiltzea erabaki dugu (detektatze-lanetan teknika enpirikoek baino doitasun handiagoa lortzen dute, eta, gainera, OLHIA sistemetan garrantzitsuak diren errore-mezuak lantzeko egokiak dira). Horrela, testuinguru lokaletan (jarraian dauden bospasei hitzeko leihoetan) gertatzen diren erroreak detektatzeko, hizkuntzaren tratamendu automatikoaren munduan ezagunak diren bi tresna erabili ditugu: Xerox Finite State Tool (XFST) datetako erroreak lantzeko eta Murriztapen Gramatika (MG) postposizio-lokuzioen erabilera okerrak tratatzeko. Azken bi kasuetan, erroreen zuzenketa ere egin dugu. Testuinguru zabaleko erroreak detektatzeko, komunztadura-erroreak adibidez, esaldiaren analisi-zuhaitza arakatzea beharrezkoa iruditzen zaigu, eta, horretarako tresna aproposik topatu ez dugunez, *Saroi* tresna sortu dugu. *Saroi* tresna orokorra da, erroreen tratamendua egiteko ez ezik, analisi-zuhaitzetako egituren gaineko kontsultak egiteko eta zuhaitz horietan egitura linguistikokoak bilatzeko ere balio baitu.

Lan honetan zailtasun ugari topatu dugu. Batzuek zerikusia dute analisi automatikoarekin (analisiko geruzetan zehar metatutako akatsek eta tresna automatikoen estaldura partzialak emaitzak kutsatzen dituzte, adibidez); beste

oztopo batzuek hizkuntzaren izaerarekin dute zerikusia (hizkuntzaren anbiguotasunak egitura batzuk erroredunak ote diren ezagutzea zaila egiten du, edo, errore sintaktikoak lantzeko, ezagutza semantikoa behar izan dugu). Azkenik, egitura sintaktikoen —erroredunen zein zuzenen— deskribapen formaletan zehaztapen falta ere nabaritu dugu. Hala ere, emaitza onargarriak lortu ditugu.