

OpenTrad, Babelgo dorreari aurre egiten

Galarraga Aiestaran, Ana

Elhuyar Zientziaren Komunikazioa

Euskara, katalana, gaztelania eta galiziera; lau hizkuntza horiek bat egiten dute sistema batean: OpenTrad itzulpen automatikoko sistema. Testuak eta web orriak gaztelaniatik euskarara, galegora eta katalanera automatikoki itzultzeko aukera ematen du, bai eta galegozko eta katalanezko testuak gaztelaniaz jartzekoa ere. Gainera, kode irekian garatu dute sistema. Elkar ulertzeko, oztoporik gabe.

EL PERIÓDICO DE CATALUNYA BI HIZKUNTZATAN KALERATZEN DUTE EGUNERO: gaztelaniaz eta katalanez. Hori egiteko, ez dute langile-kopuru bikoitza, edo langileek ez dute ordu-kopuru bikoitza ematen lanean, beste egunkariekin alderatuta. Haien sekretua itzultzaile automatiko bat da. Kazetariak gaztelaniaz idazten dute egunkaria, eta, segidan, itzultzaile automatikoak katalanez jartzen du. Hainbat zuzentzailearen artean testua zuzentzen dute, eta prest dago kalera ateratzeko, gaztelaniazkoarekin batera.



A. GALARRAGA

Itzultzaile automatikoen balioa erakusteko adibide esanguratsua da *El Periódico de Catalunya*. Egunkariak duen itzultzailea, gainera, ez da gaztelaniatik katalanera itzultzeko dagoen bakarra, beste hainbat adibide ere badaude. Esate baterako, Alacanteko Unibertsitateak interNOSTRUM sortu zuen Caja de Ahorros del Mediterráneoentzat. Bi noranzkoetan egiten du itzulpena, eta,

orain, izen bereko web gunean doan erabiltzeko aukera ematen dio edonori. Hori bai, gehienez ere 16.384 karaktereko testuak onartzen ditu.

Horretaz aparte, estatu espainiarrean badago galegotik gaztelaniara itzultzeko itzultzaile automatiko bat, baina oso produktu itxia eta mugatua da. Eta euskaraz, zer? Orain arte, ezer gutxi.

EHuko Informatika Fakultateko IXA taldekoak ari ziren ingelesetik euskarara itzultzeko sistema automatiko bat garatzen, baina ez zuten nahi bezain azkar egiten aurrera.

Horixe zen egoera duela bizpahiru urte. 2004an, baina, OpenTrad garatzeko egitasmoa abiatu zen. Izan ere, IXAkoek ezagutzen zituzten interNOSTRUM garatu zuten ikertzaileak, eta Eleka Ingeniaritza Linguistikoak eta IXAk elkarrekin egiten dute lan. Galizian antzeko lanetan ari zirenekin batu ziren, eta kode irekiko itzulzaile automatikoa sortu nahian hasi ziren, Espainiako Industria, Turismo eta Komertzio Ministerioaren diru-laguntzari esker.

Elekako Iñaki Arantzabalen esanean, bi mailatako helburuak definitu zituzten hasieratik: “batetik, galegoa-gaztelania eta katalana-gaztelania bikoteentzat itzulzaile automatiko ona, azkarra eta kode irekikoa lortu nahi genuen,

eta, bestetik, gaztelaniatik euskarara itzultzeko prototipoa. Kontuan izan behar da hizkuntza guztien abiapuntua ez zela berdina: gaztelania-katalana bikotea nahiko aurreratua zegoen, eta, beste muturrean, gaztelaniatik euskarara automatikoki itzultzeko, ia dena egiteko zegoen”.

“hizkuntza erromanikoen arteko itzulpena askoz ere errazagoa da, euskara tartean denean baino”

Hurbilekoak, azaletik

Horretan, noski, eragin handia du hizkuntzen artean dagoen aldeak. Begibistakoa da elkarrengandik askoz ere

Partaideak eta lan eta arduren banaketa

Eleka Ingeniaritza linguistikoak, S.L.: koordinatzailea eta sistema guztien integrazioaren arduraduna.

Alacanteko Unibertsitateko Transducens taldea: Apertium motorra garatzea.

EHuko IXA taldea: Matxin motorra garatzea.

Kataluniako Unibertsitate politeknikoak: gaztelaniaren analisisa eta katalanaren baliabide linguistikoak.

Vigoko Unibertsitatea: galegoaren baliabide linguistikoak.
Imaxin Software enpresa: galegoaren atalaren egiaztapena.

Elhuyar Fundazioa: euskararen atalaren egiaztapena eta euskarazko baliabide linguistikoak.

hurbilago daudela gaztelania, galegoa eta katalana, euskara horietako edozeinetatik baino. Ondorioz, askoz ere errazagoa da hizkuntza erromanikoen arteko itzulpena egiteko sistema on bat lortzea, euskara tartean denean baino.

Hori dela eta, OpenTradek itzulpen automatikoko bi motor ditu, bat hizkuntza erromanikoen arteko itzulpenetarako, Apertium, eta bestea gaztelaniatik euskarara itzultzeko, Matxin.

Biak ere hizkuntzaren erregelatan oinarritzen dira. Izan ere, itzulpen automatikoa egiteko hainbat bide daude, baina bi dira nagusiak: aurretik itzulita dauden testu-bildumetan, hau da, corpusetan, oinarritzen direnak, eta erregela linguistikoetan oinarritutakoak –hitzen ordena esaldian, deklinabidea, aditzak...–.

IXAko Iñaki Alegriak azaldu duenez, “erregela linguistikoetan oinarritutako sistemek hiru fasetan egiten dute lana. Aurrena jatorrizko testuaren analisi sintaktiko-morfologikoa egiten dute, gero beste hizkuntzarako transferentzia gauzatzen dute, eta, azkenik, bigarren hizkuntza horretan sortzen dute testua”.

Hizkuntza erromanikoen arteko itzulpena egiteko, transferentzia azalekoa izanda ere, emaitza onak lortzen dira. InterNOSTRUMek horixe egiten du,



Adibide hauetan ikusten den bezala, testuak itzultzeko ez ezik, aukeratutako hizkuntzen nabigatze ere balio dute itzulzaile automatikoei.

Īñaki Alegria, IXA taldeko koordinatzailea: “Arazo handiena anbiguotasuna da”

EHUko Informatika Fakultateko IXA taldeak izan du gaztelaniatik euskarara itzultzeko motorra garatzeko ardura. Ondorioz, ondo baino hobeto dakite zein diren zailtasun nagusiak.

Īñaki, zuei egokitu zaizue lan zailena?

Egia esan ez da erraza, batez ere hizkuntza erromanikoen arteko itzulpenarekin alderatuta. Katalanak gu baino aurreratuago zeuden arlo horretan. Bazuten oinarri sendo bat, eta, gainera, nahikoa dute azaleko transferentzia egitearekin, ahaidetutako hizkuntzen arteko itzulpena egiten ari baitira.

Gure kasua oso bestelakoa da. Egia esan, gu ere ez gara hutsetik abiatu; urteak daramatzagu gai hau ikertzen eta lantzen, eta, gainera, FreeLing-en ere oinarritu gara.



E. CARTON

Zer da FreeLing?

FreeLing gaztelaniarako analizatzaile bat da, Kataluniako Unibertsitate Politeknikoak sortua. Analizatzaile horrek zuhaitz sintaktiko osoa egiten du, euskara itzultzeko ez baita nahikoa azaleko analisisa. Esaldiaren barruko elementuen ordena oso bestelakoa da, eta, horregatik, zuhaitz osoa egin behar du, gero transferentzia egin eta esaldia euskaraz zuzen eraiki ahal izateko.

Sintaxiak edo joskerak ez ezik, lexikoak ere emango dizkizue buruko minak...

Noski. Beste hizkuntzak antzekoak dira, baina euskararen eta gaztelaniaren artean askoz ere adiera anbiguo gehiago daude. Gaztelaniazko hitz batek euskaraz bi adiera edo gehiago izan ditzake, eta arazoa da aukeratu egin behar dela. Zailtasun lexiko-semantikoa dago, beraz. Eta gero dago zailtasun morfologikoa; alegia, gaztelaniazko preposizioetatik, kasua aukeratu behar da euskaraz.

Adibide bat ematerik bai?

Betiko adibidea jarriko dizut: *zapatos de piel de seńora*. Larrua norena da? Makina batentzat, larrua emakumearena izan daiteke; hori arazo sintaktikoa da. Gainera, zer da *de* hori: nondik, nongo, noren, aposizioa? Euskararen kasuarekiko, beraz, anbigua da. Horretaz aparte, *piel* izan daiteke *larrua* edo *azala*, eta makinak *azala* aukeratzen badu, gaizki dago.

Horiek dira arazo handienak. Ahaidetuta dauden hizkuntzen artean, arazo horiek askoz ere arinagoak dira, baina guri lan handia ematen digute. Oraindik lanean ari gara arazo horiek eta beste batzuk konpontzeko, baina oinarria behintzat sortu dugu, eta aurrera jarraitzeko asmoa eta gogoa dugu.

eta, hain juxtu, hortik abiatu dira Apertium motorra garatzeko. Nolabait esateko, Apertium interNOSTRUMen bertsio hobetua da kode irekian.

Hori irabazi dute batez ere katalanek, kodea irekia izatea alegia. Horretaz gain, OpenTradek erabat berezita ditu

kodea eta baliabide linguistikoak. Horri esker, sistemak erraztasun guztiak ematen ditu elkarrekintzarako eta erabiltzailearen beharretara egokitzeko baliabideak txertatu ahal izateko. Sistema prest dago aberasteko eta hobetzeko egin nahi diren aldaketa guztiak onartzeko.

Orain ere, Apertiumek ez du soilik transferentzia sintaktikoa egiten. Horretaz gain, baditu hainbat ‘iragazki’ itzulpena fintzeko. Esate baterako, gai da hizkuntza batean berezkoak diren hainbat egitura detektatzeko eta haren baliokidea emateko beste hizkuntzan. Ondorioz, kalitate handiagoko itzulpena lortzen da. Esaterako, gaztelania-katalana bikotearen itzulpena % 95eko fidagarritasuna du; hau da, itzulitako ehun hitzetik bost besterik ez daude oker.

*“Matxinekin
emaitzak hain onak
ez diren arren,
hasierako helburua
bete dute”*

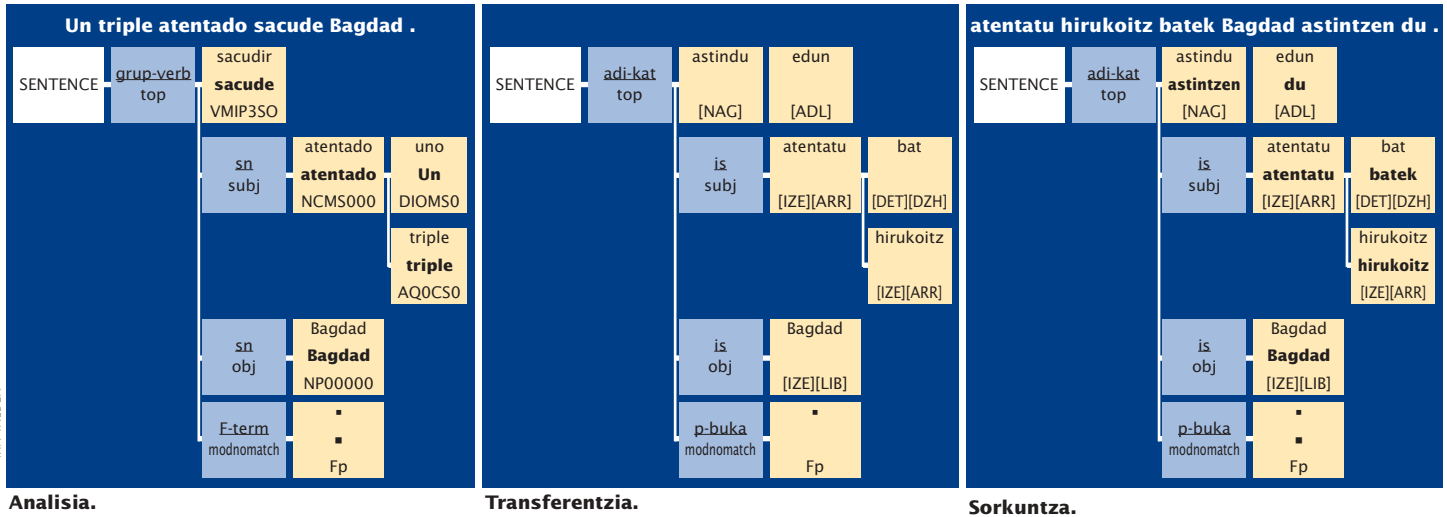
Urrunekoak, sakonean

Alabaina, Apertiumek ez du balio gaztelaniatik euskarara itzultzeko. Hizkuntzak hain desberdinak izanda, azaleko transferentzia sintaktikoa ez da nahikoa. Esaldien egitura ere erabat aldatzen da, eta, beraz, sakoneko analisi sintaktiko-morfologikoa egiten duen motorra behar da, menpekotasunen zuhaitza eraiki, transferentzia egin, eta testua euskaraz sortzeko gai dena. Horretarako sortu dute Matxin.

IXAkoek aitortu dutenez, Matxin garatzeko “lan gogorra” egin behar izan dute, eta emaitza ez da Apertiumek hizkuntza erromanikoen arteko itzulpenean ematen duena bezain ona. Nolanahi ere, hasieran jarritako helburua lortu dute, azpiegitura sortzea baitzen asmoa.

Bestalde, itzulpen automatikoa garatzean itzulpenaren kalitatea kezka nagusietakoa izan bada ere, sistemaren abiaduraz ere arduratu dira. Alderdi horretatik, gustura daudela onartu dute. Izan ere, horrek aukera ematen du web guneetan barrena ibil-

Erregela linguistikoetan oinarritutako sistemaren adibidea



Analisia.

Transferentzia.

Sorkuntza.

tzeko itzulitako hizkuntzan. Arantzabalek emandako adibide baten arabera, jatorrizkoa gaztelaniaz izanda, badago katalanez eta galegoz nabigatzea gipuzkoa.net-en, OpenTrad-en bidez.

Aurrera begira

Orain arte, galegoa-gaztelania eta katalana-gaztelania bikoteentzat bi noranzkoetan itzultzen duen sistema automatiko ona eta azkarra lortu dute, baita gaztelaniatik euskarara itzultzeko prototipoa ere. Beraz, Elekako arduradunaren hitzetan, "helburua lortu dugu".

Baina ez dute hor gelditzeko asmorik. "Hobetzen eta osatzen jarraitzeko asmoa dugu. Emaizak hobetzeko modu bat alor jakinetara mugatzea da.

Alor bakoitzean berezko hizkera erabiltzen da, eta orokorrean arituta baino anbiguotasun-arazo gutxiago izaten dira. Hortaz, kalitatea handitu egiten da itzulzailea alor batera egokituta, dagokion hiztegi terminologikoa txertatuz adibidez". Horrekin guztiarekin, fidagarritasuna hobetzea espero dute.

"itzulpenaren kalitateari ez ezik, sistemaren abiadurari ere erreparatu diote"

Horretaz gain, erregela bidezko teknologiari bestelakoak gehitzeko asmoa dute; zehazki, corpus paraleloak erabili nahi dituzte. "Horrela, esaldi bat itzuli nahi bada, aurrena begiratu du ea lehendik itzulita dagoen esaldi hori bera, edo antzeko zerbait ba ote dagoen. Aurretik zerbait egonez gero, hortik abiatuko da itzulpena egiteko. Aldiz, lehendik ez badago antzeko ezer, erregela bidezko teknologia erabiliko du".

Hobetzeaz eta osatzeaz aparte, euskaratik gaztelaniara itzultzeko itzulzaile automatikoa sortu nahi dute. Horri esker, kanpokoek euskaraz sortzen denaren berri jasotzeko aukera izango lukete. Ingelesetik euskarara ere itzuli ahal izatea da etorkizunerako beste helburuetako bat.

Aurrerapauso horiek emateko, Eusko Jaurlaritzaren laguntza izango dutela espero du Arantzabalek. Hain zuzen ere, duela urte batzuk Eusko Jaurlaritzak Kataluniako enpresa bati aginduz itzulpen automatikoko prototipo bat garatzeko. Orain, berriz, OpenTrad da sistema aurreratuena estatu espainiarrean. Horregatik dio Arantzabalek: "gure sistema bultzatzeko konbentzitu nahi genuke Eusko Jaurlaritza. Gure ustez, behintzat, ezin da kanpoan geratu".

- <http://www.opentrad.net>
- <http://apertium.sourceforge.net>
- <http://matxin.sourceforge.net>



Emaizak hobeak lortzen dira itzulzailea alor batera egokituz gero. Izan ere, alor bakoitzean berezko hizkera erabiltzen da, eta ez dira izaten hainbeste anbiguotasun-arazo.