

Alicia Perez

Itzulpengintza automatikoan aditua

«Kostuak jaisten saiatu behar dugu, itzultzaileentzat erremintak sortuz»

EHUko Donostiako
campusean Sistemen
Ingeniaritza eta Automatika
Sailean irakasle da. Leioan
berriz, Elekrika eta
Elektronika Sailean ikerlari.
Harrituta dago kazetariak
bere tesiari buruz galdetu
diolako. Lehen ere hurbildu
zaizkio hedabideetatik eta
interes hori arraroa egin zaio.

ONINTZA IRURETA AZKUNE
Argazkiak: Dani Blanco

ESAN DIOGU, bere alorra ulertzeko zaila den arren, itzulpengintza automatikoa badela gai interesgarria, hizkuntza gutxituentzat adibidez. *Speech and text translation involving basque language: application of stochastic finite-state transducers* izeneko tesia egin zuen, zer ikertu duen galdetu diogu: “Makinak itzulpenak egiteko metodo desberdinak ditu. Nik metodo estatistikoa darabilt. Euskaraz zentratu naiz, baina berez, estatistikak edozein hizkuntzatik edozein hizkuntzara egitea ahalbidetzen du. Metodo horrek hitzak etiketak edo zenbakiak bailiran tratatzen ditu, lehenengo hitza a hizkuntzan eta bigarren hitza b hizkuntzan. Dena dela, okerrago edo hobeto



“Hogei urte pasa dira eta ez dugu lortu hizkuntza batetik bestera natural eta zuzen itzultzea. Apurka ari gara”.



“Estatistikak, euskara eta gaztelaniaren arteko itzulpenetan, azkenean txapuza egiten du, oso-oso hizkuntza desberdinak baitira. Zerbait egiten du, baina gutxi, oraindik bere ahalmena nahiko murrizta da”.

ibiliko da sistema, hizkuntzek euren artean duten hurbiltasunaren edo urruntasunaren arabera. Estatistikak, euskara eta gaztelaniaren arteko itzulpenetan, azkenean txapuza egiten du, oso-oso hizkuntza desberdinak baitira. Zerbait egiten du, baina gutxi, oraindik bere ahalmena nahiko murrizta da.

Estatistikarenaz gain, badira beste metodo batzuk itzulpengintzan, eta nahastu egin ohi dira.

Bai, hori da gaur egungo joera.

Baina estatistika aukeratu duzu.

Bai. Ametzagaña Taldea itzulpen

memoriekin ari zen eta estatistika sartu nahi zuen proiektuan. Eurek lagundu ziguten itzulpen unitateak –izen sintagmak, adibidez– identifikatzen.

Beraz, alde linguistikoa eta estatistikoa nahastu dituzu.

Itzulpen unitateak identifikatu nahi genituen, baina metodo linguistiko-en bidez. Metodo estatistikoen bidez esaldiko subjektua, aditza eta bestelako elementuak identifikatu daitezke. Alabaina, metodo estatistikoei bere logika erabiliz egiten dute zatiketa hori: iruditzen zait honaino dela partiketa. Ondo,

baina guk partiketa metodo linguistikoak erabiliz egitea nahi genuen, alegia, unitateak segmentatzea metodo linguistikoaren bidez.

Eta horretarako Euskalmeteko eguraldi iragarpenen lagina erabili zenu-ten.

“Zerua oso hodeitsu egongo da arratsalde partean kostaldean eta mendietan aldiz...”. Horrelako laginak hartu genituen. Alor mugatu baterako aplikazioa da nik landutakoa, eta helburua da aplikazio eremu apurka-apurka handitzea. Baina, hau hasiera baino ez da.

Hori galdetu nahi nizun. Batzuetan badirudi itzulpen automatikoa asmatuta dagoela, hedabideek kontatzeko duten moduagatik, adibidez.

... agerian dago ezetz, bestela denok erabiliko genuke mugikorren! Ez, ez dago asmatuta.

Nola transmititzen duzue egiten duzuen lana, noraino iritsi zareten azaltzea, adibidez?

Aspalditik ari gara itzulpengintza automatiko estatistikoan. 90eko hamarkadan AEBetan IBMko taldea sortu zen, eta horrek eman zien buelta aurreko metodoei. Hogei urte pasa dira eta ez dugu lortu hizkuntza batetik bestera natural eta zuzen itzultzea. Apurka ari gara uztartzen metodo linguistikoak eta estatistikoak.

Euskararekin bezala beste hizkuntze-kin ere berdin aritzen dira lanean?

Azken finean, adimen artifizialeko ezagumenak aplikatzen dira, metodo orokorrak dira aplikatzen direnak problema jakin baterako. Nahiko gatz da gure problema zeren ez da matematika, hizkuntza da, oso aberatsa.

Buelta eman diozu, hizkuntza arazoa dela esan beharrean aberatsa dela diozu.

Erronka da, ea zer ekarri ahal diodan hizkuntzalaritzari nire metodo-en bidez.

Itzulpen “perfektua” noiz izango dugun imajinatzen al duzu?

Ez nuke jakingo aurretik, adibidez, hainbeste urtetan maila ertain

neko itzultzailea izango dugu... Google-ek badauka berea, nahiko ona da, baina alor guztietarako aplikagarria al da? Oraindik ez. Hala ere, zientziaren bilakaera oso-oso azkarra da. Garai batean, bospasei urtean inork ez zizun tesia zapaltzen, gaur egun zure gaia urte pare batean zaharkituta geratu da, jende piloa ari da gai horretan.

Jendeak espero du itzultzaile perfektua aurkitzea Interneten, adibidez.

Bai, eta gero bazoaz probatzera eta txarto egiten du, ezta? Biltzarretan eta aipatzen dena da: zuk zer nahiago duzu, Taiwanetik erositako garbigailuaren eskuliburua taiwanera perfektuan egotea edo euskara xumean? Nik nahiago dut ezagutzen dudana hizkuntza egotea, akatsen bat badago ere nahiago dut ulertzea, beste hizkuntza batean perfektua izatea baino. Agerian dago itzultzaileen beharra oso handia dela. Kostuak jaisten saiatu behar dugu, itzultzaileei lagunduko dien erremintak sortuz.

Lagungarri beraz.

Euskalmeteko lagina, euskaraz eta gaztelaniaz geneukan, baina publikatzeko ingelesez behar genuen zerbait. Kontratatu genuen itzultzaileak 14.000 esaldi itzultzeko milioi bat pezeta kobratu zigun, ez da asko. Gure sistemak, entrenatuta zegoenean, 1.500 esaldi zuzentzeko segundo batzuk behar izan zituen. Itzultzaileari pasa nizkion esaldiak eta esan nion, “zuk egin duzun lana nireak segundo batzuetan egin du”. Noski, berak bezain ondo ez zuen egin makinak, baina kalitatea begiratzeko esan nion. Harri eta zur geratu zen, “ez dira zuzenak, baina ulergarriak dira” esan zuen. Aitortu beharra dago segundo batzuetan 1.500 esaldi itzultzea ondo dagoela. Gero, itzul-

tzaileari gutxiago kostako zaio esatea ondo ala gaizki dagoen, zerotik itzultzen hastea baino.

Itzulpen automatikoak egiteko corpus bolumen handiak behar dira eta euskarak ez ditu, era berean dauzka-gun itzulpen memoriak ez ditugu elkarbanatzen. Ondo ari al naiz?

Nondik aterako ditut laginak? Euskaraz publikatutakoa, doan, erraz eta formatu egokian lortzeko moduan... baldintza larregi dira, zaila da.

“Biltzarretan-eta aipatzen dena da: zuk zer nahiago duzu, Taiwanetik erositako garbigailuaren eskuliburua taiwanera perfektuan egotea edo euskara xumean?”

Tesirako enpresa batekin jarri zinen harremanetan. Ohikoa da hemen alor akademikoa eta enpresa elkarlanean aritzea?

Nahiko banatuta egoten dira. Unibertsitatean, batzuetan, ez dakigu ebatzitako problemak praktikan zelan jarri. Gero hor dago enpresa mundua gu gure lekuan jartzeko, “horrek ez du ezertarako balio, gaur egungo gizarteak behar duena hau da”. Oso tesi gutxi daude enpresarekin lotuta. Unibertsitatean ahaztu egiten gara gizartearen beharrez, enpresak berriz lurrean kokatuta daude. Gure proiektua sortu zen enpresa batek eskatuta.

Euskara askoz zailagoa da itzulpengintzarako, berezitasun linguistikoak dituelako, ala hori ez da egia eta arazoa corpus falta da.

Batetik, euskara oso bitxia da mor-

fologikoki, atzizkiak jartzen ditu bata bestearen atzetik eta hitzaren esanahia askotxo aldatzen da. Bes-tetik, sintaxia ere oso bitxia da, gaztelaniaz egiten dena da subjektua, aditza eta osagaiak. Euskaraz ohiko egitura da subjektua, osagaiak eta aditza, eta aditzaren aurretik garrantzitsuena dena. Itzulpengintza egin behar dugunean, bi gauza hartu behar ditugu kontuan: bata, esanahia batetik bestera trasladatzea, eta bestea, gero hitzen orden egokia aukeratzea. Katalanaren eta gaztelaniaren artean ordena nahiko parekoa da, sistemari ez zaio hainbeste kostatzen ordena hori ulertzea, baina sistema estatistikoentzat kristo bat da urrutiko lerrokadurekin lan egitea [euskara eta gaztelania kasu]. Erronka da desberdinak diren hizkuntzen arteko itzulpena egitea.

Euskara eta beste hizkuntza batzuk egongo dira antzeko egoeran, ezta?

Alde horiek kontuan hartuta bai, baina euskarak gainera, baliabide linguistiko gutxi dauka, hiztun gutxi dauka. Baliabide gutxi izanda kalitate bereko gauzak lortzen saiatzea beste erronka bat da.

Alegia, berezitasunetan finlandiera pare da euskara, baina kopuruan finlandierak irabazten dio.

Noski. Orain biltzarretan esaten ari direna da metodo bereziak behar ditugula hizkuntza murriztuak bultzatzeko.

Singapurren egon naiz duela gutxi. Egunkarietan parte batzuk txineraz zeuden eta beste batzuk ingelesez. Itzulpengintza egin nahi baduzu hor ez daukazu corpus paralelorik, alegia, hau honen itzulpena da. Gara egunkariaren kasua bera da.

Erronka da hizkuntza batean zein bestean hurbilpen handia izango duten edukiak aurkitzea. ■

Babeslea: HABE