

ITZIAR CORTÉS ETXABE

Elhuyarreko itzulpengintza-teknologiaren arduraduna



# BEGIRADA BAT

## itzulpen automatikoko sistemen bilakaerari

**I**tzulpen automatikoa teknologia ezaguna dugu. Interneten nabigatzen dugunean, sarri erabiltzen dugu atzerriko hizkuntzetan idatzitako testuak ulertzeko, edo, zergatik ez, bai eta itzulpenak egiteko laguntza gisa ere.

Askotariko zerbitzuak topatzen ditugu Interneten, gehienak doakoak, eta, itxuraz berdinak diruditen arren, teknologia ezberdinetan oinarritutako tresnak izan daitezke, edo denboran zehar moldatzen joan dira, eta erabiltzaileontzat gardena izan da aldaketa hori. Itzulpen automatikoko sistema bat martxan jartzeko, beraz, hainbat aukera daude: erregeletan oinarritzen diren sistemek (RBMT) hizkuntza-baliabideak edo -ezagutza behar dituzte; estatistikan oinarritutakoek (SMT) dagoeneko itzulita dauden testuen bildumak behar dituzte, eta teknika estatistikoak erabilia, nolabait, itzultzen ikasteko gaitasuna dute; eta, azkenik, NMT sistemak edo neurona-sareetan oinarritutakoak daude, azken boladan oso ezagun egin direnak.

Giza garunaren funtzionamendua imitatzen dute neurona-sareetan oinarritutako sistemek, eta informazio-bilduma handietatik sortzen dira, baina estatistikan oinarritutako sistemak baino konplexuagoak dira. Garuneko neuronaren funtzionamendua imitatzen saiatzen den eredu matematiko batean oinarritzen da teknologia

hori (*deep learning* delakoaren barnekoa), eta, ezagutu zenetik urte batzuk igaro diren arren, orain ari da indarra hartzen. Izan ere, horrelako sistemak martxan jartzeko, datu-bilduma handiez gain, konputazio-ahalmen handiko gailuak behar dira, eta gaur egun ditugun ordenagailu eta txartel grafikoekin lortu da halako sistemak modu eraginkorrean sortzea.

**“Euskarara itzulpenak egiten dituen neurona-sareetan oinarritutako sistema bat martxan jartzea da helburua”**

Itzulpen automatikoko sistemak aldatzen joan dira denboran zehar, eta horren adibide ezaguna dugu Google: erabiltzaileok konturatu gabe, eredu berrietara egokitzten joan da, eta erabiltzen ditugun itzultzaileetako batzuk, dagoeneko, neurona-sareetan oinarritutakoak dira. Ez dezagun pentsa, ordea, teknologia hori Interneteko erraldoien esku bakarrik dagoenik; geurean ere ari gara arlo horretan ikertzen, eta euskararekin lanean dihardugu. Oraingoz, ikerketa-mailako esperimentuak egiten gabiltza, baina jasotako lehen emaitzak itxaropentsuak dira. Denbora

gutxi berru euskarara itzulpenak egiten dituen neurona-sareetan oinarritutako sistema bat martxan jartzea da helburua.

### CONSUMER.EUS WEBGUNEAN, ELE BIETAN

Matxin (<http://matxin.elhuyar.eus>) aitzindaria da gaztelaniatik euskarara itzulpenak egiten. 2007an doktoretza-tesi batetik sortu zen itzulpen automatikoko sistema bat da (Aingeru Mayor Martinez, UPV/EHU), euskararako lehen itzultzaile automatikoa, eta, ordutik, garai berrietara egokitzen ari da pixkanaka, hizkuntza-teknologiak hain garrantzitsu diren aro digital honetan. Erregeletan oinarritzen da; beraz, hizkuntza-baliabide bidezko ezagutza dauka, eta horri esker egiten ditu itzulpenak. Zehatzago esateko, gaztelaniazko testuak ulertzeko behar duen jakintza dauka, eta, nolabait, haiek euskarara bihurtzekoa. Horretarako, besteak beste, hiztegiak eta erregela sintaktikoak erabiltzen ditu.

Azken hiru urteotan, [Consumer.eus](http://Consumer.eus) proiektuan lanean gabiltza Eroski Fundazioa, EHUko Ixa ikerketa-taldea eta Elhuyar. Orain arte gaztelania hutsean irakur zitezkeen edukiak automatikoki itzultzen dira euskarara, eta Matxin erabiltzen da lan hori egiteko; itzultzaile sendoa da, software librean oinarritua, hizkuntza-baliabideak erraz egokitzeko aukera ema-



ten du, eta, haiek aplikatu ondoren, uanean bertan nabaritzen da emaitzen hobekuntza. [Consumer.eus](http://Consumer.eus) webguneko edukiak itzultzeko, elikadura-arloan jarri dugu arreta, eta arlo horretara egokitu ditugu itzultzaile automatikoak erabiltzen dituen hizkuntza-baliabideak. Egindako moldaketen eragina berehala ikusi da: baliabideak egokitzearekin batera, sistemaren kalitatea hobetu egin da, elikadurari buruzko testuetan, betiere.

Jakinekkoa da, ordea, automatikoki sortutako itzulpenen kalitatea ez dela beti esperotakoa izaten, edo, behintzat, ez dela zuzenean argitaratzeko modukoa izaten. Horregatik, proiektu horretan bertan, itzulpenak moldatzeko aukera eskaintzen zaio irakurleari; edonork har dezake parte proiektuan. [Consumer.eus](http://Consumer.eus) webgunek euskarazko artikulua eta errezeptak zuzentzeko aukera ematen du, eta egindako lan hori gorde egiten da. Zertarako? Bildutako datuekin itzultzailea hobetzeko. Jasotzen den informazioarekin eta ikasketa automatikoa erabiliz, itzulpen automatikoko sistema berri bat martxan jarriko da, eta elikadura-arloko testuak itzultzen espezialista izango da.

## ITZULPEN AUTOMATIKOAREN ERRONKAK

Itzulpen automatikoko sistemetan aurrea egitea izugarriko erronka da, eta emaitzen kalitatea hobetzeko bidea esperimentuz beteta dago: erregeletan oinarritutako sistemekin lanean hasi ginean, hizkuntzalarien eskuza lana ezinbestekoa zen. Itzultzailearen jatorri-eta helburu-hizkuntzak ezagutu behar zituen hizkuntzalariak, eta hizkuntza arteko zubia egiteko erregelak sortu behar zituen (morfologia- eta sintaxi-mailan, esaterako). Horrelako sistema bat hutsetik hastea, beraz, lan eskerga da. Alabaina, estatistikan edo ikasketa automatikoan oinarritzen diren sistemak sortzeko ez da, nahitaez, ezagutza linguistikorik behar (eredu simple bat sortzeko, behintzat).

Azken urteotan sarri entzuten dugu *Big Data* kontzeptua: eskura ditugun datu-multzo erraldoiei erreferentzia egiteko erabiltzen da. Haietatik informazioa erauzten da, eta, kasuan kasu, dagokion azterketa egiten da. Artikulu honetan aipatutako ikasketa-teknikak, izan ere, ez dira itzulpen automatikoaren kasuan bakarrik erabiltzen; *Big Data* kontzeptua

beste ikasketa- edo azterketa-prozesu batzuetarako ere erabiltzen da. Badirudi, beraz, erraza dela itzulpenak gauzatzeko sistemak sortzea datuak eskura ditugunean, baina, kasu horretan ere, mugak topatzen ditugu.

Ikasketa automatikorako edo sakonerako (*machine learning*, *deep learning*) datu-multzoak prestatzea, sarritan, ez da erraza izaten. Kontuan izan behar dugu eredu matematikoak sortzeko aurrez itzulitako esaldien multzo bat behar dela: itzulpenek kalitatezkoak izan behar dute, eta datu-multzoak handia izan behar du. Teorian, sistema sortzeko zenbat eta datu-multzo handiagoa erabili, orduan eta emaitza hobea emango du itzultzaileak.

Badirudi, ordea, goia jo dela estatistikan oinarritutako sistemekin: nahiz eta ikasketarako erabiltzen den datu-multzoa handitu, itzultzailearen emaitza ez da proportzio berean hobetzen. Horregatik ari gara neurona-sareetan oinarritutako sistemetan ikertzen, muga hori gainditu, eta itzulpen automatikoaren alorrean aurrera egiteko. ●