

ITZIAR CORTÉS ETXABE  
Elhuyar Hizkuntza eta Teknologia



# ITZULPEN AUTOMATIKOAREN magia eta mugak

**D**efinizioz, hizkuntza batetik bestera itzultzen duen sistema informatikoa da itzulpen automatikoa; itzulpen-prozesuan zehar gizakiak parte hartu behar izan gabe ematen du itzulpena. Automatikoki sortzen den itzulpen hori testuak ulertzeko edo itzultzeko baliabide gisa erabil daitezke; hori izan daiteke itzulpen automatikoko sistemak eta hiztegiak sarritan nahastea eragiten duen arrazoia. Baina bi baliabide horiek badi-tuzte aldeak. Adibidez, hiztegieta bilaketak egiten ditugunean lortzen ditugun emaitzak profesionalek eskuz landuak dira. Itzulpen automatikoko sistemetan, aldiz, nahiz eta oinarri gisa erabiltzen diren datuak profesionalek landutakoak izan, erantzuna beti makinak berak sortzen du; automatikoki sortzen da. Hala ere, erabiltzaile arrunten ikuspuntutik alde nagusia hauxe da: itzulpen automatikoko sistemek esaldi osoak nahiz hitzak itzul ditzaket.

Itzulpen automatikoko sistema ezagunen artean Google Translator dago. Sistema horren bidez, 70 hizkuntzaren arteko itzulpenak egin daitezke automatikoki, eta, 2010etik, euskara ere hizkuntza horien artean dago. Baina zergatik ateratzen dira itzulpen xelebreak horrelako sistemetan?

Interneten gabiltzanean, nahiko ohikoa izaten da bilatzen ari garen informazioa beste hizkuntza batean aurkitzea, eta informazio hori ulertzeko itzulpen automatikoko sistema bat erabiltzea. Horregatik, itzulpen automatikoa erabiltzearen abantailak eta desabantailak zein diren jakin behar du erabiltzaileak. Horrelako sistema batek ez du beti itzulpen zehatza ematen, eta erabiltzailearen esku egoten da automatikoki lortutako itzulpen horrekin zer egin. Sarri gertatu izan da automatikoki lortutako emaitza horiek publizitatean nahiz bestelako karteletan ikustea.

Honako adibide hauek ezagunak ditugu gehienok: “Autobuses, mañana y tarde”, euskaraz “Autobusak bihar eta berandu”; “Combinación de trenes”, “Trenen azpikogona”; eta “Primera planta”, “Lehenengo landarea”.

Urtarrilean bertan, futbolaren munduan aski ezagunak diren Guardiola eta Iniesta haien jaioterriekin nahasten zituen Google-ren itzulpen automatikoko sistemak katalanetik ingeleserako itzulpena egitean. Zergatik sortzen dira, baina, horrelako itzulpenak modu automatikoan?

Itzulpen automatikoko sistemak bi multzotan sailkatu ditzakegu, haiek sortzeko erabiltzen diren tekniken arabera: estatistikan oinarritutako sistemak eta erregetan oinarritutako sistemak.

Lehen aipatutako Google Translate bera da estatistikan oinarritutako itzulpen automatikoko sistemen artean ezagunenetarikoa. Mota horretako sistemak testu-bildumak erabiltzaile sortzen diren eredu estatistikoetan oinarritzen dira. Adibidez, euskararen eta gaztelaniaren arteko sistema bat sortu nahi bada, era honetako testu-bilduma bat hartu beharko dugu oinarri: euskarazko esaldi bakoitzak gaztelaniazko itzulpena izan beharko dugu. Testu-bilduma horiek oinarri hartuta, eredu estatistiko batzuk sortzen dira, eta eredu horiek izango dira itzulpen automatikoko sistemaren muina.

Elhuyar Hizkuntza eta Teknologia unitatean ere baditugu horrelako sistemak, eta aurten horrelako sistema bat ezartzen ari gara MINHAPen (Espainiako Ogasun eta Administrazio Publikoetako Ministerioan). Webguneak gaztelaniatik euskarara eta ingelesera itzultzeko erabiliko dute sistema hori, eta automatikoki lortutako emaitzak gainbegiratu egingo dituzte eskuzko errepasso bat eginez.

Erregeletan oinarritutako sistemek, aldiz, oinarri linguistikoa dute; hau da, hiztegiak eta hizkuntzari lotutako erregelen moduko baliabideak dituzte oinarri. Normalean, estatistikoei baino emaitza zentzuzkoagoak ematen dituzten arren, sistema-mota horrek ere baditu zailtasunak eta emaitza arraroak. Hitz polisemikoek, adibidez, nahiko lan ematen dute itzultzeko garaian. Har dezagun, adibidez, gaztelaniazko 'tiempo' hitza, euskaraz hamabi adiera dituena (hiztegiak.elhuyar.org/es/tiempo): denbora, eguraldi...; itzulpen automatikoko sistemari adiera guztien artean egokiena zein den aukeratzeko erakutsi behar zaio.

Matxin da erregeletan oinarrituta gaztelaniatik euskarara automatikoki itzultzen

duen sistema baten adibidea. Elhuyar Hizkuntza eta Teknologia eta EHUko Ixa taldeak elkarrekin garatu dute Matxin sistema (<http://matxin.elhuyar.org>), eta, testu-hutsa itzultzeaz gain, formatu ezberdineko dokumentuak eta webguneak itzultzeko ere balio du. Hala ere, lehen esan den bezala, itzulpen automatikoa erabiltzen dugunean, ez dugu ahaztu behar automatikoki sortutako emaitza horiek gainbegiratu egin behar direla.

➤ *Horrelako sistema batek ez du beti itzulpen zehatza ematen, eta erabiltzailearen esku egoten da automatikoki lortutako itzulpen horrekin zer egin.*

Oinarri estatistikoa eta linguistikoa duten sistemen arteko aldea ikusteko, hona hemen pare bat adibide. Esaterako, "El hombre que vino era mi tío" esaldia honela itzultzen du Google Translatek: "Gizona etorri

zen nire osaba izan zen"; eta Matxin 2.0k, berriz, honela: "Etorri zen gizona nire osaba zen". Bestalde, "Han dicho que mañana va a hacer buen tiempo" esaldia "Bihar eguraldi azoka izango esan dute" itzultzen du Google Translatek; eta Matxin 2.0k, berriz, "Esan dute bihar denbora ona egitera doala".

Itzulpen automatikoaz hitz egiten dugunean, zenbaitek errezeloz begiratzen gaituzte. Gaia ulertzeak, ordea, baliabide haur erabiltzaileen ikuspuntua alda dezake. Itzulpena automatikoki egiten dela esatean, ez dugu esan nahi emaitza gisa itzulpen zuzen bat jasoko dugunik, hau da, itzultzaile profesionalen lana egingo duenik. Adierazi nahi duguna da automatikoki sortutako emaitza bat eskainiko duela, testu bat guk aukeraturako hizkuntza batetik beste batera bihurtuko duela. Itzulpen automatikoko sistemari buruzko iritzia eskatuko bagenu, sistema erabiltzen ari den erabiltzailearen araberako erantzunak jasoko genituzke, jakina. Horrenbestez, nahitaezkoa da horrelako baliabideak modu kontzientean erabiltzea. ●

