



03

APLIKAZIOAK ETA AZTERKETA TEORIKOAK

>>>

KATALANAREN ITZULPEN AUTOMATIKOA

Mikel L. Forcada

Alacanteko Unibertsitatea

Helbide elektronikoa: mlf@dlsi.ua.es

Itzulpen automatikoan, testu idatziekin egiten da lan eta, bereziki, **testu informatizatu**ekin, hau da, euskarri informatiko batean gordeta dauden testu-dokumentuekin, esate baterako, testu-tratamenduko programekin sor edo edita daitezkeen dokumentuak. *Automatika* da *informatika-sistemek* egiten dutelako, hau da, software egokia instalatuta daukaten ordenagailuek. Honela definitzen da **itzulpen automatikoa**: *sorburu-hizkuntzan* idatzitako testu informatizatu bat eraldatzea (informatika-sistema bat erabiliz) *xede-hizkuntzan* idatzitako beste testu informatizatu bat lortzeko; *itzulpen landugabea* deritzogu azken testu horri.

Itzulpen automatikoak baditu zenbait **muga**. Oro har, oso desberdinak izaten dira itzulpen automatikoko sistemek egindako itzulpen landugabeak eta profesionalek egindakoak; gainera, gerta liteke itzulpen automatiko horiek egokiak ez izatea zenbait komunikazio-helburutarako. Hainbat faktorek eragiten dute desegokitasun hori, hala nola: giza testuen *anbiguotasuna* (esanahi bat baino gehiagoko hitzak dituzte edo egitura sintaktiko bat baino gehiagoko esaldiak), sorburu-hizkuntzaren eta xede-hizkuntzaren artean dibergentzia sintaktikoak egotea eta abar. Arazo horiei aurre egiteko erabiltzen diren metodoek, oro har, erabat sinplifikatzen dute itzulpen-prozesua. Sinplifikazio horiei esker, alde bate-

tik, erregela mekaniko errazak egin daitezke itzulpen automatikoko sistema azkarrak eta trinkoak arrazoizko denbora-tarte batean sortzeko; baina, bestetik, lortzen diren irtenbideak ez dira onenak.

Zer dakar itzulpen automatikoak? Muga horiek ikusita, espero dezakegu itzulpen automatikoak gu itzulpenaren zati mekaniko (edo "mekanizagarri") horretatik askatzea, baina, itzulpen automatiko hori ona izan arren, ez du testua ulertuko; ez ditu beti zuzen konponduko anbiguotasunak eta testuak ez ditu xede-hizkuntzaren aldaera jator batean sortuko.

Bi aplikazio-talde handi daude itzulpen automatikoan:

Lehenengo taldea **asimilaziorako** aplikazioek osatzen dute, hau da, itzulpen automatikoa erabiltzen dutenak beste hizkuntza batean idatzitako dokumentuak oro har ulertzeko (adibidez, Interneten argitaratutako testuak). Txat bateko elkarrizketen itzulpena da asimilaziorako itzulpen automatikoaren beste adibide bat; txat horretan parte hartzen duen pertsona bakoitzak bere hizkuntza erabil dezake eta beste parte-hartzaileen ekarpenak irakur ditzake bere hizkuntzara itzulita. Aplikazio mota horietan, oso azkar itzuli behar da, hobe mementoan bertan, eta zuzenean erabiltzen da, landugabe; batzuetan, ez da osorik irakurtzen eta normalean, ez da gordetzen irakurri ondoren. Itzulpen automatikoaren aplikazio hori ez dago itzulpengintzako profesionalen mende.

Bigarren taldean daude **banaketarako** aplikazioak: aplikazio horietan, itzulpen automatikoa tarteko pausoa da argitaratuko edo banatuko den dokumentu bat xede-hizkuntzan egiteko. Beraz, zirriborro gisa egindako itzulpena gorde egiten da, berrikusi eta zuzendu (*posteditatu*) egin behar baitu aditu batek. Sinplifikatuz, esan dezakegu posteditzioa behar duen itzulpen automatikoa izango dela itzulpen profesionalaren alternatiba, baina baldin eta haren guztizko kostua itzulpen profesional tradizionalarena baino txikiagoa bada. Batzuetan, posteditzioan aurrezteko, lehenago, aurreedita egin daiteke automatikoki itzuliko den jatorrizko testuaren zati bat; hala, alde batera utziko dira erabiltzen ari den itzulpen automatikoko sistemaren arazo ezagun horiek.

Itzulpen-teknologiak bi talde handi ere badaude. Duela 50 urte inguru egindako saioetatik 90eko hamarkada arte, *erregeletan oinarritutako itzulpen automatikoa* deritzona izan da itzulpen automatikora gehien hurbildu den sistema: informatikariek eta itzulpengintzaren arloko adituek osatutako taldeek hiztegi elektronikoak konpilatzen dituzte, analizatzaile morfologikoak eta sintaktikoak programatzen dituzte, transformazio gramatikalerako erregelak definitzen dituzte eta abar. 90eko hamarkadaren hasieratik, gero eta gehiagotan erabiltzen da *corpusean oinarritutako itzulpen automatikoa* (testuarena) deritzona: itzulpen automatikoko programek "itzultzen ikasten dute" (adibidez, eredu estatistiko konplexuak erabiliz) ehunka mila testu elebidunetako corpus erraldoietatik abiatuta; horietan, hizkuntza bateko ehunka mila esaldi lerrokatu dira beren itzulpenarekin.

Honela definitzen da itzulpen automatikoa: sorburu-hizkuntzan idatzitako testu informatizatu bat eraldatzea (informatika-sistema bat erabiliz) xede-hizkuntzan idatzitako beste testu informatizatu bat lortzeko; itzulpen landugabea deritzoguz azken testu horri.

Eta nola erabiltzen da katalanerako itzulpen automatikoa? Hizkuntza hori idazten den ia lurralde guztietan beste hizkuntza bat dagoenez (ofiziala da) eta ezagutzen ditugun arrazoiak tarteko menderatzailea denez, zaila da katalanez irakurtzen dakien eta beste hizkuntza batean irakurtzen ez dakien jendea aurkitzea (jende askori, ordea, kontrakoa gertatzen zaio).

Baina katalanerako itzulpen automatikoaz hitz egin behar zen; hizkuntza horri “valentziera” ere esaten zaio, besteak beste. Baina, eztabaidak alde batera utzita, eta **teknologia alorretik aztertuta**, besteak beste, Castellán, Sabadellen, Ordinan, Perpinyán, Manacorren, Mequinensan eta l’Alguerren bizi garen gehienok hitz egiten dugun hori edo, hobe, itzulpenez ari garenez, **ondo idazten saiatzen garen** hori izango da katalana. Informatikan, *ISO-639-2 jarraituz, *ca* kodea ematen zaio horri eta, funtsean, hizkuntza bakar gisa erabili behar da.

Eta nola erabiltzen da **katalanerako itzulpen automatikoa**? Hizkuntza hori idazten den ia lurralde guztietan beste hizkuntza bat dagoenez (ofiziala da) eta ezagutzen ditugun arrazoiak tarteko menderatzailea denez, zaila da katalanez irakurtzen dakien eta beste hizkuntza batean irakurtzen ez dakien jendea aurkitzea (jende askori, ordea, kontrakoa gertatzen zaio). Horregatik, asimilaziorako aplikazioak ia beti egiten dira katalana sorburu-hizkuntza dela kontuan izanda: katalanez dagoen dokumentua ulertzen ez duenak itzulpen automatikoa erabiltzen du ulertzeko. Banaketarako aplikazioak, ordea, nahiko ohikoak dira, sarritan itzuli behar baitira dokumentuak katalanera legeak agintzen duelako edo enpresa pribatuetako administrariak edo bezeroek hala eskatzen dutelako. Testu asko itzultzen da automatikoki katalanera eta berrikusi egiten dira (edo berrikusi egin behar lirateke) argitaratu aurretik.

Beste hizkuntza **gatzelania** denean, pentsatu behar litzateke itzulpen automatikoak arazorik ez lukeela izango, ez baitira oso desberdinak katalana eta gatzelania. Egiaz: sistema helduenek, gutxi gorabehera, hamar hitzetik bakarra zuzendu beharra duten testuak egiten dituzte.

Anbiguotasunak dira arazo garrantzitsuenak. Batetik, **lexikozko anbiguotasunak** daude, hitz jakinei eragiten dietenak: “nadie conocía el destino del avión secuestrado” gatzelaniazko esaldian ez dakigu nola itzuli behar den destino: *destí* (patua) edo *destinació* (-ra doan). Izan ere, hizkuntza guztiek dituzte hitz anbiguo oso erabiliak. Gatzelaniaz *como* hitza honela itzul daiteke: *menjo* (aditza), *com* (juntagailua) edo *com a* (preposizioa); *para* hitza, honela: *para* edo *atura* (aditza), *per* edo *per a* (preposizioa); eta *una* hitza honela itzul daiteke: *una* (determinatzailea) edo *uneixi* (aditza). Badaude anbiguotasun sintaktikoak edo egiturazkoak ere: “se comió el postre enfadado” esaldian, argi eta garbi dago platera ez dagoela haserre, baina substantiboaren eta adjektiboaren komunztadura dela eta, sintaktikoki onargarria da interpretazio hori eta itzulpen automatiko-programa batek “es va menjar les postres enfadades” (*postres* femenino singularra da katalanez) itzuliko luke; aukera horrek oso ondo funtzionatuko luke “se comió el postre azucarado” eta antzeko esaldietan.

Itzulpen automatikoa erabiltzeak **ondorio positiboak eta negatiboak** izan ditzake. Batetik, katalanerako itzulpen automatikoak erraztu egin dezake katalanez testuak egitea, dokumentuak, webguneak eta abar. Jatorrizko informazioa ere, beti, katalanez egon ahal izango litzateke

itzulpen automatikoari esker; izan ere, katalanez hitz egiten ez dutenek irakurtzeko bitartekoak egongo lirateke. Baina, bestetik, gerta liteke batzuek pentsatzea “ez dutela katalana ikasi beharrik” testuak katalanez egiteko (gezurra: beti zuzendu beharko dira testuak) edo hizkuntza horretan argitaratzen den informazioa ulertzeko; gainera, beste hizkuntzaren kalkoak eta zuzendu gabeko akatsak dituzten testuak ugartu daitezke, postediziorik ez dagoelako (baina, sukaldeko labanarekin ere, jende ugari erail dute eta ez dut uste labana horiek erabiliko ez ditugunik horregatik).

Zein da **katalanerako itzulpen automatikoaren** eskaintza? Oso zabala eta, kasu batzuetan, ez ditu katalana eta gaztelania soilik lotzen, bai eta katalana eta ingelesa, frantsesa, alemana, okzitania edo esperantoa ere, besteak beste:

- SALT, Valentziako Generalitateko itzultzailea; Josep Lacreuk diseinatu zuen; dagoeneko, 3. bertsioa dugu eta, oker ez banago, hamar urte bete behar ditu. Gaztelaniatik itzultzen du katalanaren Valentziako aldaerara, eta, alderantziz; doan deskarga daiteke leku askotatik eta Windows sistema eragilean erabil daiteke. Software bikaina da lexiko aldetik duen estaldura dela-eta, eta erabiltzaileak erabiltzen duen bitartean katalanaren aldaera valentziarra ikasteko sortu zen.
- Ara deritzon programa eta SALT oso antzekoak dira teknologikoki, baina programa hori erabiltzeko ordaindu egin behar da. Arak (www.ara-autotrad.com) Katalunian erabiltzen duten katalanaren aldaeran idatzitako testuak egiten ditu.
- InterNOSTRUM.com itzultzailea Caja Mediterráneo eta Alacanteko Unibertsitateak finantzatu dute eta artikulu honen egilean zuzentzen du; linean erabiltzeko prest dago. Gaztelaniatik Katalunian erabiltzen den katalanaren aldaerara eta Valentziako unibertsitateetan erabiltzen duten aldaerara itzultzen du, eta katalanaren hainbat aldaera “ulertzen” ditu gaztelaniara itzultzean. Segur aski, interneteko itzulpen-programa erabiliena da.
- Apertium.org webgunea kode irekia erabiltzen duen (software librea)¹ itzulpen automatikoko plataforma da; interNOSTRUM itzultzailetik eratorri zen. Alacanteko Unibertsitateko talde berak abiarazi zuen eta mundu osoko programatzaileek eta hizkuntzalariek elkarlanean sortu zuten. Linean erabil daiteke eta Linux sistema eragilean instala daiteke. Plataforma horretan erabiltzeko prest daude, besteak beste, katalanaren eta gaztelaniaren, frantsesaren, ingelesaren, okzitaniaren eta esperantoaren arteko itzultzaile automatikoak.
- Translendum (www.translendum.com) enpresa bat da eta, besteak beste, hizkuntza hauen arteko itzulpen automatikoko sistemak merkaturatzen ditu: katalanaren (Kataluniako aldaera soilik) eta gaztelaniaren, ingelesaren, alemanaren, frantsesaren eta Araneko okzitaniaren arteko itzulpenak. Doan proba daiteke linean.

Anbiguotasunak dira arazo garrantzitsuenak. Batetik, lexikozko anbiguotasunak daude, hitz jakinei eragiten dietenak. Badaude anbiguotasun sintaktikoak edo egiturazkoak ere.

Zein da katalanerako itzulpen automatikoaren eskaintza? Oso zabala eta, kasu batzuetan, ez ditu katalana eta gaztelania soilik lotzen, bai eta katalana eta ingelesa, frantsesa, alemana, okzitaniara edo esperantoa ere, besteak beste.

Esan dezakegu valentziera maila bikainean dagoela itzulpen automatikoko teknologien arloan, eta segur aski potentzial hori ez dela erabat erabiltzen oraindik.

- AutomaticTrans (www.automatictrans.es) itzultzaileak gaztelanieren eta katalanaren (Kataluniako aldaera soilik) arteko itzulpenak egiten ditu; *El Periódico de Catalunya* egunkariaren edizio elebiduna egiten duen itzultzailearekin du zerikusia.
- SisHiTra (“itzultzeko sistema hibridoa”) da gutxien ezagutzen den itzultzailea. Valentziako Unibertsitate Politeknikoak sortu zuen finantzazio publikoa erabiliz eta, prhltedemos.iti.es/%7Esisshitra/ webgunean sartuz gero, zuzenean erabil daiteke. Valentziako aldaeran oinarritzen da. Emaitzak oso onak dira.

Sistema hauek guztiak, agian azkena izan ezik, erregeletan oinarritzen dira. Bada beste sistema esperimental bat ere corpusean oinarritzen dena. *Periódico de Catalunya* egunkariko testu elebidunekin trebatu da eta Valentziako Unibertsitate Politeknikoak sortu du. Probak egiteko (esaldi batekoak), <http://dcomgp05.gnd.upv.es/WebTrans.debug/root> webgunean sar daiteke, baina oraindik ez da erregeletan oinarritutakoak bezain eraginkorra.

Zein da sistema onena? Nik guztiak probatzeko aholkatuko nuke, baina, zeregin jakin baterako zein sistema komeni den erabaki aurretik, itzulpenaren kalitateaz gain, honako hauek ere hartuko nituzke kontuan, besteak beste: dokumentuaren formatua errespetatzen duten ala ez, testuaren tamainarekin mugarik dagoen eta prezioa.

Zer egingo dugu programek “Bartzelonako katalana” soilik ulertzen edo sortzen badute? Horrelako programetan, arazoak izan ditzakegu, ez dutelako ulertzen edo sortzen katalanaren aldaera bat. Aldaera ulertzen ez badute, emaitzak hobetu daitezke testu orijinala aurreidatuz, sistemak erabiltzen duen aldaerara egokitzeko (adibidez, sistemak Valentziako formak ulertzen ez baditu, *este* guztien ordez, *aquest* idatz daiteke; *-uïska* bukaeraren ordez, *-ueixi*, eta abar); aurreidazio hori automatizatu egin daiteke, adibidez, testu-tratamenduko programaren makroaginduak (“makroak”) erabiliz askotan azaltzen diren hitzak ordeztzeko. Programek aldaera egokia sortzen ez badute, irtenbidea errazagoa da: gehiago posteditatzea (bestela ere, posteditatu egin behar genuke); posteditazio hori ere automatizatzeko aukera izango genuke neurri batean, aurrekoak bezalako aldaketak programatuz, baina alderantziz.

Azkenik, esan dezakegu valentziera maila bikainean dagoela itzulpen automatikoko teknologien arloan, eta segur aski potentzial hori ez dela erabat erabiltzen oraindik.■

OHARRA

1. Programa bat librea dela esaten da, baldin eta (a) edonork edozein helburutarako libreki erabil badezake, (b) edonork libreki kopia eta bana badezake, (c) hala nahi duenak aldatu egin badezake, hobetzeko edo beste helburu baterako egokitzeko, eta (d) hala nahi duenak aldatutako bertsiok bana baditza ke. Azken bi baldintzak betetzeko kodea eskuratzeko aukera eduki behar denez (programatzaileak idatzita bezalaxe), *kode irekiko programak* ere esaten zaie. Software librearen inguruan garatzaile boluntarioen komunitateak sor daitezke, hura hobetzeko eta zabaltzeko; horixe ari da gertatzen Apertiumekin.