

# Itzultzaile automatikoa: mehatxua ala aukera?

---

ITZIAR CORTÉS ETXABE ETA  
SAROI JAUREGI AIESTARAN (ELHUYAR)

Teknologia berriek eraldaketa dakarte, eta horren adibide da itzulpengintza. Aspaldi hasi ginen itzulpenak modu profesionalean egiten; teknologia baliatuz, ordenagailuz lagundutako itzulpenak egiten hasi ginen. Gaur egun ere hala dihardugu, software gero eta ahaltzuagoekin eta hizkuntza-baliabide gero eta aberatsagoekin.

Hizkuntza-teknologiak etengabe garatuz eta eraldatuz doaz, eta horiek gure eguneroko bizitzan txertatzen ari gara, ia konturatu ere egin gabe. Gizarte osoaren esku dauden baliabi-deak dira, eta, helburu ezberdinak lortzeko erabiltzen ditugun arren, denontzat dira ezagunak. Agian, itzulpen automatikoa da ezagunena. Nork ez du noizbait erabili Google Translate, edo nork ez du *sartu* ulertzen ez duen testu bat itzultzaile automatikoan? Euskararako, akaso, zaila izan da emaitza onak lortzea halako teknologiak erabiliz; euskarak ba omen ditu hainbat zailtasun eta muga hizkuntza-teknologiak garatzeko orduan...

Azken hilabeteetan izan ditugun berriek, ordea, besterik diote: euskararen eta gaztelaniaren arteko itzulpen automatikoko sistema on bat sortzea posible da, eta horren adibide dira modela.eus webgunean erabiltzaileari eskaintzen zaizkion sistemak. Segundo gutxi batzuk baino ez dira behar paragrafo oso bat itzultzeko, euskaratik gaztelaniara edo alderantziz. Gainera, dokumentuak itzultzeko ere balio du, behin zerbitzua kontratatuta.

Orain arte, erabili izan ditugu itzulpen automatikoko sistemak beste hizkuntza-bikote batzuetarako, dela testu bat ulertzeko, dela itzulpen bat egiteko laguntza gisa, baina ez genuen espero kalitate-maila bertsuko sistemarik euskararako. Eta hain ona al da gaur egun itzulpen automatikoak eskaintzen duen kalitatea? Itzulpen hau makina batek egin duela esango bagenu, sinetsiko zenuke?

Jatorrizkoa:

*Duela 20 urte eman zuen Teknopolis saioak bere lehen emanaldia, ETBn. Geroztik, astero izan da Elhuyarrek ekoiztutako saio hau ETBko lehen bi kateetan, helburu jakin batekin: zientzia- eta teknologia-gaiak gerturatzea ikus-entzuleei, modu entretenigarrian, bertako zentroei eta ikertzaileei emanaz lehentasuna.*

Itzulpena:

*El programa dio su primera actuación en ETB hace 20 años. Desde entonces, el programa ha sido producido por Elhuyar en las dos primeras cadenas de ETB con un objetivo concreto: acercar los temas científicos y tecnológicos a los telespectadores, de una manera entretenida, dando prioridad a sus centros e investigadores.*

Elhuyarrentzat ezaguna da itzulpen automatikoaren arloa, eta teknologia horren bilakarearen lekuko gara. Sistema batekin baino gehiagorekin lan egin dugu, hainbat ikerketa-ildo jorratu ditugu, eta emandako jauzia nabarmena da. Hasieran sortu ziren sistemak erregeletan oinarritutakoak ziren; hainbat hiztegi eta hizkuntza-baliabide batera lanean jarri ematen zuten itzulpena sistema horiek. Ondoren, estatistikan oinarritutakoak etorri ziren, testu-bilduma elebidun erraldietatik sortuak; metodo estatistikoak oinarritzat hartuz, itzulpenak sortzeko gai ziren. Azkenak, gaur egun hizkuntza-bikote gehienetarako erabiltzen ditugunak, sare neuronaletan oinarritzen dira, adimen artifizialean.

Ikus dezagun adibide batekin nolakoa izan den bilakaera hori.

Jatorrizko esaldia, albiste batetik ateratakoa, gaztelaniaz:

*La Hacienda de Gipuzkoa deja exentas de IRPF las bajas de maternidad y paternidad.*

1. itzulpena, erregeletan oinarritutako sistema batek emandakoa (Matxin, 2009):

*Hacienda De Gipuzkoak salbuetsiak uzten du PFEZ-ren amatasuneko baxuak eta aitatasuna.*

2. itzulpena, estatistikan oinarritutako sistema batek emandakoa (Elhuyarren sistema bat, 2016):

*Gipuzkoako Ogasunaren salbuetsita uzten PFEZ amatasun- eta aitatasun-bajak.*

3. itzulpena, sare neuronaletan oinarritutako sistema batek emandakoa (modela.eus, 2018).

*Gipuzkoako Ogasunak PFEZtik salbuetsita utzi ditu amatasun eta aitatasun bajak.*

Sistema batetik bestera dagoen aldea begi-bistakoa da. Agian, adibide gisa erabilitako esaldia errazegia zen? Hona beste adibide bat, konplikatuagoa:

Jatorrizkoa:

*Las bases reguladoras y la convocatoria de 2018 de subvenciones destinadas a ayuntamientos de Gipuzkoa para la realización de proyectos en los municipios dentro del programa Bizikidetza Lantzen 2018/2019 fueron aprobadas mediante Acuerdo del Consejo de Gobierno Foral de 26 de junio de 2018 (BOLETÍN OFICIAL de Gipuzkoa n.º 128, de 4 de julio de 2018).*

Itzulpena:

*2018/2019 Bizikidetza Lantzen programaren barruan udaletan proiektuak egiteko Gipuzkoako udalei zuzendutako 2018ko diru-laguntzen oinarri arautzaileak eta deialdia 2018ko ekainaren 26ko Foru Gobernu Kontseiluaren Erabakiaren bidez onartu ziren (Gipuzkoako Aldizkari Ofiziala, 128. zk., 2018ko uztailaren 4koa).*

Akatsak akats, automatikoki sortu da euskarazko esaldia, segundo pare batean.

Emaitza horiek ikusita, ez al da espero izatekoa itzultzaile profesionalak ere baliabide hori erabili nahi izatea? Itzulpen-lan bat egiteko orduan, itzulpen-memoria bat edo glosario bat hautatzen den gisara, ez al da zentzuzkoa azken belaunaldiko itzulpen automatikoko sistema bat ere lotzea norberaren OLI tresnari (Ordenagailuz Lagundutako Itzulpen-tresna: Wordfast, SDL Trados Studio, MemoQ...)?

### **Itzulpen automatikoa itzulpen-zerbitzu profesional batean**

Itzulpen automatikoak gaur egun ematen dituen emaitzak ikusita, ondoriozta dezakegu itzulpenen merkatua eraldatzera doala. Orain arte beste hizkuntza-bikote batzuekin lan egiten zuten itzulpen-merkatuetan gertatu den gisara, gurera ere badator iraultza; alegia, teknologiak aurrerapausoak ematen dituen heinean, itzulpengintza eraldatzen ari da. Zenbat aldiz pentsatu dugu itzultzaile profesionalen lana kolokan dagoela eta desagertu egingo dela? Akaso, itzulpen-memoriekin lanean hasi ginenean ere hori bera pentsatu genuen, baina hemen jarraitzen dugu, lanean, talde-lanean, teknologiarekin batera gure lan-jarduna eraldatzen.

Demagun itzulpen automatikoa erabili nahi dugula itzulpen-zerbitzu batean, eta itzultzaile profesional baten eguneroko lanean baliabide gehigarri gisa txertatu. Imajinatuko dugu gure ohiko OLI tresnan guztiz integratuta dagoen baliabide bat dela, eta gure lana egiteko modua optimizatzeke sistema bat dela. Beraz, segmentuz segmentu itzultzerakoan, esaterako, modu honetara egin dezakegu lan:

1. Segmentu bat *irekiko* dugu.
2. OLI tresnak itzulpen-memorian bilaketa egingo du, eta, antzekotasun-atalase baten arabera, emaitza bat(zuk) ekarriko d(it)u itzulpen-memoriatik.

3. Kontsultatutako itzulpen-memori(et)an emaitzarik ez badago, hau da, sistemak itzulpen-memori(et)an antzeko esaldirik topatu ez badu, itzulpen automatikoko zerbitzura joko du OLI tresnak, eta haren emaitza erakutsiko dio itzultzaileari zuzenean; hots, automatikoki sortutako itzulpen bat.

Nola eragingo dio itzultzaile profesionalaren eguneroko lanari horrelako baliabide bat OLI tresnan integratuta edukitzeak? Itzulpen-proposamen baten gainean egingo du beti lana, eta inoiz ez du segmentu hutsik izango.

Teknologia hau izaten ari den eboluzioa ikusita, pentsatzekoa da gero eta presentzia handiagoa izango duela gure egunerokoan. Gizakion lana ordezkatu al du? Elhuyarren argi daukagu ezetz: adimen artifizialak eta ikasketa automatikoko teknikak bultzada nabarmena eman diete itzulpen automatikoaren kalitateari eta erabilerari, bai, baina ezin dute zuzenean ordezkatu itzultzaile profesional baten lana. Teknologia eta baliabide horiei etekina ateratzen ikasi behar dugu, denok, baita itzultzaileok ere; baina dena ez da ezkurra zerriaren askan.

### Itzulpen automatikoaren eragina itzulpengintzan

Euskararen eta gaztelaniaren arteko itzulpen automatikoko sistemak espero gabeko emaitzak ematen ari dira, eta horrek emozioa, zirrara, harridura sortu du guregan... baina ziurgabetasuna eta errezeloa ere ekar lezake, ezin baitugu neurtu teknologia horiek zer bilakaera izango duten, eta, *magia* balitz bezala, gero eta emaitza hobek ematen baitituzte. Beste aurrerapen batzuekin gertatu den moduan, ohitzea izango al da kontua?

Itzulpen automatikoa denon eskura dago, Internet eskuragarri duen edonork erabil dezake, eta erabiltzailearen erantzukizuna da emaitza hori zertarako eta nola erabili. Denon eskura badago, itzultzaile profesionalen eskura ere badago teknologia hori.

Itzultzaile profesionalen kasua berezia da, ordea: sistematikoki gordetzen dute beren eguneroko lana, OLI tresnen bidez, itzulpen-memoria batean (edo batzuetan). Nolabait, lan egiten duten dokumentu bakoitzeko esaldi bakoitza eta haren itzulpena gordetzen dituzte, txukun-txukun, eta hori da, hain zuzen, gaur egungo itzulpen automatikoko sistemen lehengaia. Hizkuntza-teknologiaren arloan, baliabide preziatua dira itzulpen-memoriak.

Eskura ditugun itzulpen-memoretatik «ikasten» dute sistema automatikoek, eta «ikasitako» horren araberrako proposamenak egiten dituzte. Esaterako, Euskaltzaindiak hartutako azken erabakietako bat izan da *dirulaguntza* hitz bakarrean idaztea; lehen, *diru(-)laguntza* zeukan arautua. Horrek esan nahi du itzultzaile automatikoak ikasteko erabilitako testuetan bi aukera hauek ageri direla nagusiki: *diru-laguntza* eta *diru laguntza*. Beraz, sistemak hortik ikasi du.

Jatorrizkoa: *El sistema actual de subvenciones ...*

Itzulpena: *Gaur egungo diru-laguntzen sistema...*

Itzultzaile automatikoen erabiltzaileon lana da jasotzen ditugun proposamen horiek zuzentzea, adibide horrek argi adierazten duen moduan.

Garrantzitsua da kalitate oneko «lehengai»ekin elikatzea gure sistema automatikoak, gero eta proposamen hobeak (eta gaurkotuagoak) egin ditzaten. Alegia, itzultzaile automatikoek egindako proposamenak bere horretan sartzen baditugu gure itzulpen-memoretan, zuzendu gabe, itzulpen-memoria horien kalitatea ez da izango orain artekoena bezain ona, eta gero memoria horiek erabiltzen baditugu etorkizuneko itzulpen automatikoko sistemari «irakasteko», emaitzak okerrera egingo du. Azken batean, hizkuntza aterako da galtzaile.

Itzultzaile profesionalak hartzen dituzten erabakiek hizkuntza-teknologietan duten eragina ikusteko, hona beste adibide bat.

Jatorrizkoa: *Zalantzarik baduzu, deitu erizainari.*

Itzulpena: *Si tiene alguna duda, llame a la enfermera.*

Jatorrizkoan *erizain* ikusita, zergatik proposatu du sistemak *enfermera* eta ez *enfermero*? Arrazoiak bera da: ikasteko erabilitako corpusetan *enfermera* hitza lau aldiz gehiago agertzen da *enfermero* baino. Bere horretan onartzen badugu automatikoki sortutako testu hori, ez dugu konponduko genero eta estereotipoen arazoa etorkizuneko hizkuntza-teknologien emaitzetan.

## **Aurrera begirakoak**

Itzulpen automatikoa erabili ala ez, sortzen dugun edukiaren erantzule gara; are gehiago, hurrengo urteetan sortuko diren hizkuntza-teknologietan eta hizkuntza-baliabideetan eragina izan dezakete gaur sortutako testuek; teknologia horien adibide bat besterik ez da itzulpen automatikoa.

Orain arteko hizkuntza-teknologiak erregeletan oinarritutakoak ziren: hizkuntzalariek hizkuntza deskribatzen zuten kontzienteki, eta informazio egituratu hori nolabait martxan jartzen aipatutako teknologiak sortzeko egin beharreko lana. Gaur egun, ordea, dagoeneko sortuta dauden testuetan oinarritzen dira. Horrek esan nahi du gaur hartzen diren erabakiek eragina izango dutela etorkizuneko hizkuntza-teknologietan; esaterako, zuzentzaile automatikoetan. Teknologia horiek denon eskura egongo dira, gure ordenagailu eta aplikazioetan, eta, zuzentzaile horiek testu okerretatik «ikasten» badute, «zuzenketa» okerrak egingo dituzte. Beraz, gaur hartzen diren erabakiek eragina izango dute etorkizuneko hizkuntzan, oro har.

Galdera bat du izenburutzat artikulua honek, eta erantzuna da bietatik duela. Itzultzaile automatikoko sistemak aukera bikaina dira nola erabili asmatzen badugu, baina mehatxu bihurtu daitezke egiten dituzten proposamenak zalantza jarri gabe onartzen baditugu eta zabaltzen badira. Etorkizuneko corpusen kalitatea eskastuko luke horrek, eta, ondorioz, baita corpus horien gainean eraikitzen diren teknologia eta baliabideen kalitatea ere (hiztegiarena, besteak beste). Eta, azken-azkenean, hiztunek erabiltzen duten hizkuntzaren beraren kalitatea ere eskastuko litzateke, «ereduzko» corpus «ez-eredugarri»en eraginez.

### Resumen

Durante muchos años, las tecnologías lingüísticas no han cumplido las expectativas de los vascohablantes. Frente a otras lenguas más «fuertes», teníamos grandes limitaciones e inconvenientes para conseguir corpus, para procesar automáticamente los textos orales, para conseguir traducciones automáticas aceptables. El origen de esas dificultades eran, por un lado, las características propias del euskera y, por otro, la escasez de materias primas para las tecnologías lingüísticas. Sin embargo, en los últimos años hemos avanzado mucho en este tipo de tecnologías y, en especial, en la traducción automática. Utilizando tecnologías basadas en redes neuronales, hemos obtenido resultados de gran calidad en los sistemas que traducen del euskera al castellano y del castellano al euskera. Esto, evidentemente, provoca cambios en la actividad de los traductores profesionales: oportunidades y amenazas. Son dos extremos a tener en cuenta, porque condicionarán la lengua del futuro.

### Résumé

Pendant de nombreuses années, les technologies linguistiques n'ont pas répondu aux attentes des basco-phones. Par rapport à d'autres langues plus « puissantes », nous étions très limités et nous nous sommes confrontés à des obstacles pour obtenir des corpus, pour traiter automatiquement des textes oraux, pour obtenir des traductions automatiques acceptables... Ces difficultés découlaient, d'une part, des caractéristiques de la langue basque et, d'autre part, de la pénurie de matières premières pour les technologies linguistiques. Cependant, au cours des dernières années, nous avons réalisé de grands progrès dans le domaine de la technologie linguistique, en particulier dans le domaine de la traduction automatique. En utilisant des technologies basées sur des réseaux neuronaux, nous avons obtenu des résultats de très grande qualité dans les systèmes de traduction du basque vers l'espagnol et de l'espagnol vers le basque. Cela entraîne évidemment des changements dans l'activité des traducteurs professionnels. Changements qui sont des opportunités mais aussi des menaces : deux extrêmes à prendre en considération, car ils conditionneront la langue du futur.

### Abstract

For many years, linguistic technologies have fallen short of the expectations of Basque speakers. In comparison with other “stronger” languages, we have suffered considerable limitations and obstacles in obtaining a corpus in order to automatically process oral texts, in order to achieve correct automatic translations, etc. The characteristics of the Basque language itself contributed to these difficulties, as did the scarcity of primary materials for the linguistic technologies in question. Nevertheless, in recent years we have advanced significantly in language technologies and especially in automatic translation. Using technologies based on neural networks, we have obtained high quality results in systems that translate from Basque into Spanish and vice versa. This of course entails changes in the activity of professional translators: opportunities and threats. Both must be taken into account because they will determine the language of the future.