

Itzultzaile automatikoak euskaraz, bilakaera eta prestazioak*

(Automatic translators in Basque, evolution and performance)

Rodríguez González, María del Mar*; Marauri Castillo, Iñigo**; Pérez Dasilva, Jesús Angel***
Univ. del País Vasco (UPV/EHU) Sarriena s/n. 48940 Leioa
*mariadelmar.rodriguez@ehu.eus;
**inigo.marauri@ehu.eus;
***jesusangel.perez@ehu.eus³

Jaso: 2016.11.16

BIBLID [eISSN 1988-3935 (2016), 15; 225-242]

Onartu: 2016.11.26

Ikerlan honek online aurki daitezkeen euskarazko itzultzaile automatikoak aztertu nahi ditu, haien ezaugarri nagusiak zehazteko xedez; halaber, ikerlanak nabarmendu egin nahi ditu zein diren erabiltzaileari parte hartzeko aukera ematen diotenak testuak aberastu eta makinak egindako itzulpenaren kalitatea hobetzeari begira. Lankidetzako plataforma adimentsuak dira, eragin sozial handia dutenak.

Giltza-hitzak: Itzultzaile automatikoak. Euskara. Ezaugarriak. Plataformak. Lankidetzak.

El objetivo de esta investigación es estudiar los diferentes traductores automáticos on line en euskera con el fin de determinar cuáles son sus principales características, así como destacar aquellas que permiten la participación del usuario en el enriquecimiento de los textos y mejora de la calidad de la traducción realizada por la máquina. Se trata de plataformas colaborativas inteligentes con un importante impacto social.

Palabras clave: Traductores automáticos. Euskera, Características, Plataformas. Colaboración.

L'objectif de cette recherche est d'étudier les différents traducteurs automatiques disponibles en ligne en basque afin de déterminer quelles sont leurs principales caractéristiques et de relever celles qui permettent à l'utilisateur de participer à l'enrichissement des textes et à l'amélioration de la qualité de la traduction réalisée par la machine. Il s'agit de plateformes collaboratives intelligentes à fort impact social.

Mots-clés : Traducteurs automatiques. Basque. Caractéristiques. Plateformes. Collaboration.

1. BILAKAERA ITZULTZAILEEN DISEINUAN ETA ERABILERAN

Gaur egun, app-en diseinua, elkarreragitea eta erabiltzaileen eta web-aplikazio dinamikoen arteko lankidetzak hain garrantzitsuak diren honetan, harrigarri samarra da online dauden euskarazko itzultzaile automatikoetan¹ apenas aurkitzea zenbait ezaugarri, hala nola erabiltzaileen arteko interakzioa; izan ere, itzultzailean terminoa sartu eta itxaron egiten da hark itzuli arte, lortutako emaitza hobetzeko aukerarik gabe. Hala frogatzen du online dauden euskarazko itzultzaile automatikoen artean ikertzaile-talde honek egindako ikerlanak. Bereizi beharra dago euskarazko edo euskararen itzultzaileen eta zuzentzaileen artean, kontuan hartu behar baita haietako batzuk prototipoak direla. Hiztegiak ere ikertuko dira. Jarraian, ikerlan honen oinarria diren onlineko itzultzaile automatikoak aipatzen dira: Xuxen.eus, Matxin onlineko itzultzailea, Euskalterm, Eusko Jauriaritzaren itzultzailea. euskadi.net, UZEI sinonimoen/antonimoen hiztegia, Elhuyar hiztegia, Googleko itzultzailea eta Openrand sistema. Ikerketa-esparrua² behin zehaztuta, haietako bakoitzaren ezaugarriak zehaztea komeni da, haien bilakaera eta prestazioak erakuste aldera.

Xuxen.eus: webgune horretan, Xuxen zuzentzaile ortografikoa erabiltzeko aukera izango dugu. Azken urteotan, zenbait aldaketa eta hobekuntza egin dira, bai Xuxen zuzentzaileak funtzionatzeko erabiltzen dituen datuetan, bai diseinuari lotutako alderdi batzuetan. Xuxen.eus baliabidearen bertsio berriak datu linguistikoen azken bertsioa jartzen du erabiltzailearen esku, Euskal Herriko Unibertsitateko IXA taldearen eta Elhuyarko Hizkuntza eta Teknologia unitatearen arteko elkarlanari esker sortutakoa.

Zuzenketa-atalean egindako hobekuntzak nabarmentzen dira. Lexikoari dagokionez, aurreko bertsioak baino 18.645 sarrera gehiago ditu, eta, halaber, aldatu egin da sistemak proposamenak eskaintzeko erabiltzen duen modua ere. Horrela, oker dagoen hitz baten ordez proposamenak eskaintzeko orduan, kontuan hartu dira euskaraz idazten dugunean egiten ditugun akats ohikoenak. Webgunean* (xuxen.eus) azaltzen duten bezala, irizpide hori erabiltzen da sistemak eskaintzeko proposamenetan, baina, gainera, proposamen horiek ordenatu egin daitezke Internetetik modu automatikoan eskuratutako euskarazko corpus baten datuak oinarri hartuta. Modu horretan, proposamenak horrela ordenatzeko aukera du erabiltzaileak, testu-eremuaren azpian ageri diren aukeren artean horri dagokiona aktibatuz. Baliagarria da irizpide hori erabiltzea proposamenak ordenatzeko; izan ere, gehien erabiltzen diren formak proposatzen ditu sistemak, gutxien erabiltzen direnen aurretik.

1. JAUREGUI, A. 2012. *EHUskaratuak. Corpus eleaniztun baten etiketatze automatikoa eta kontsulta-sistemaren garapena* [Tesis de máster]. Disponible en <http://ixa.si.ehu.es/master/master_tesiak>. Consultado el 20/07/2016.

2. LINDEMANN, D. (2014). Creating a German-Basque electronic dictionary for German learners. *Lexikos*, 24(1), 331-349.

Gaur egun, beharrezkoa da sisteman erregistratzea eta saioa hasia hiztegia erabiltzeko. Horrela, Xuxenek zuzenketa-pantailan ezezaguntzat jotzen dituen hitzak hiztegi pertsonalizatuan sartzeko aukera ematen da. Hiztegiak bezala, Xuxen.eus zuzentzaileak ere hobetsitako hitzak proposatzen ditu. Hobetsitako hitz horiek erabiltzailearentzako proposamenak baino ez dira, jakin dezan berak idatzitako hitza zuzena bada ere, badagoela hobetsitako forma bat. Kasu horretan, Euskaltzaindiaren Hiztegi Batuan oinarritu dira datu-basea sortzeko erabilitako datuak. Hortaz, Xuxen.eus zuzentzaileak proposatutako hitz hobetsiak Euskaltzaindiaren Hiztegi Batuan proposatutako berberak dira, eta «h» laburduraz markatzen dira.

Adierazitako prestazioez gainera, Xuxen.eus zuzentzaileak beste funtzionaltasun batzuk ere eskaintzen dizkio erabiltzaileari. Hona hemen garrantzitsuenak: hizkuntza detektatzea, Elhuyar hiztegiak kontsultatzea, ahots-sintesia, erabiltzailearen hiztegia kudeatzea eta TTS sistema bidez entzuteko aukera.

- **Hizkuntza detektatzea.** Sistemak biltzen dituen datuak analizatuta, ikus daiteke erabiltzaileek zuzentzeko sartzen dituzten testuak maiz ez direla euskaraz idatzita egoten. Xuxen euskarazko zuzentzailea izanik, sartutako testuak euskaraz egon behar luke. Hori kontuan hartuta, eta erabiltzailearen akatsa detektatze aldera, hizkuntza detektatzeko sistema batetik igarotzen da testua, hura zuzendu baino lehen.

Erabiltzaileak testu-eremuan sartutako testua euskaraz badago, sistemak testua zuzentzeari ekiten dio zuzenean. Baina sistemak sumatzen badu testua beste hizkuntza batean idatzita dagoela, hala jakinaraziko dio erabiltzaileari.



1. irudia: www.xuxen.eus

Onlineko Matxin itzultzailea: gaztelania-euskara (<http://matxin.elhuyar.eus>). Eleka Ingeniaritza Linguistikoa S.L. enpresak Matxin sareko itzultzaile automatikoaren –gaztelania eta euskara– bertsio hobetu batean lan egin zuen 2008an zehar, Elhuyar Fundazioarekin eta Euskal Herriko Unibertsitateko IXA taldearekin elkarlanean. Eleka 2003an hasi zen Openrad plataforma garatzen Galiziako eta Kataluniako enpresa eta unibertsitateekin batera; kode irekiko itzulpen automatikoak egiten ditu plataforma horrek, www.openrad.com webgunean. Matxin itzultzailearen (gaztelania-euskara) gaur egungo bertsioa plataforma horren parte da, eta zenbait itzulpen-teknika erabili dituzte lan horretarako. Doako sistema bat da, Openrad kode irekiko sistemaren parte dena; horrenbestez, edozein erabiltzailek bere beharretara egokitzeko aukera du inongo lizentziarik ordaindu behar izan gabe.



Matxin 2.0 - Euskarazko itzultzaile automatikoaren bertsio berriua

2. irudia: matxin.elhuyar.eus

Euskalterm (<http://www.euskara.euskadi.eus/r59euskalte/eu/q91Eus-TermWar/kontsultaJSP/q91aAction.do>). 1986an, UZEIk EUSKALTERM terminologia-bankua sortu zuen. Irabazi-asmorik gabeko entitate bat da, 1997an sortua, hizkuntzaren esparruan zerbitzuak eskaintzeko xedez. Horrela, 1986an, Eusko Jaurlaritzak, 204/1989 dekretuaren bidez, interes publikokotzat jo zuen: “interes orokorra bultzatzen du Euskal Autonomia Erkidegoan, ikerketa zientifikoan eskaintzen dituen zerbitzuengatik”. Esan liteke EUSKALTERM euskal terminologia-banku publikoa dela, UZEIk egindako bankuaren oinarriaren gainean sortua; izan ere, erreferentzia bat da euskarazko terminologiaren erabiltzaileentzat, terminologiaren arloan urte-tan egindako lana biltzeaz gainera, Terminologia Batzordeak onetsitako eta gomendatutako terminologia ere biltzen baitu. Hartara, terminologiaren erabileran orientabidea eta gomendioak ematea da haren helburu nagusia.

3. irudia: Euskalterm

Eusko Jaurlaritzaren itzultzailea (www.itzultzailea.euskadi.net). 2012ko otsailaren 28an, Eusko Jaurlaritzako Hizkuntza Politika eta Kultura Saileko sailburuordeak itzultzaile automatikoa aurkeztu zuen, 1.600.000 euroko inbertsioa egin ondoren. Hona hemen helburua: abian jartzea lan garrantzitsu batzuk eta aurretik zegoen sistemari itzulpenak egiteko bi helbide berri eranstea: euskara-gaztelania eta ingelese-euskara, baita Itzulpen Memorien Banku Publikoa ere.

Itzultzailea doako tresna bat da, gaztelaniatik euskarara itzultzen dituen testu ez espezializatuak eta webguneak –hazietan nabigatu ahala itzultzen ditu–. Halere, ez da sistema perfektu* bat (hala adierazi zuten Kulturako sailburu Blanca Urgell andreak, Hizkuntza Politikako sailburuorde Lurdes Auzmendi andreak eta itzultzailea garatzeaz arduratu zen Bartzelonako Lucy Software Ibérica enpresako zuzendariak 2012ko otsailaren 28an itzultzailea aurkezteko egin zen prentsaurrekoan).

Itzultzailea ezagutarazteko orduan, haren garapenean eta hobetze-lanean parte hartu zuten erakundeek argi utzi zuten laguntzeko tresna bat zela. “Itzultzaile automatikoak emandako euskarazko itzulpena erabili nahi izanez gero aurkeztu edo argitaratu behar den dokumentu edo testuren batean, ezinbestekoa da aurretik pertsona zuzentzaile batek aztertu eta zuzentzea itzulpen hori”. Eta ez hori bakarrik, Auzmendik azpimarratu zuenez, euskarazko itzulpena ahalik eta zuzenena izan dadin, ezinbestekoa da gaztelaniazko testua ongi idatzita egotea. Horrek esan nahi du ez dela akats ortografikorik egon behar, azentuak ere behar bezala jarri behar direla eta puntuazio-zeinuak ere egoki erabili beharra dagoela. Horretaz gain, sintaxiak zuzena izan behar du, eta ez oso konplexua.

Itzultzaile automatikoa

Testua Webguneak

Testu laburren itzulpena (muga:1.500 karaktere, testu formatuan)

Aukeratu hizkuntzak:

Erakutsi aukerak:

4. irudia: www.itzultzailea.euskadi.net

Euskara-gaztelania itzultzailea mugikorrean. Eusko Jaurlaritzaren app-a. Eusko Jaurlaritzak, itzultzailea.euskadi.net tresna aurkeztu eta handik bi urtera, aplikazio bat diseinatu zuen Android eta IOS sistema eragileak dituzten gailu mugikorretan erabiltzeko. Testu laburrak itzultzen ditu sistemak, gehienez 1.500 karakterekoak, baita webguneak ere. Aurreko kasuan bezala, aplikazioaren goi arduradunek ohartarazi zuten itzultzaile horrek, itzultzaile automatiko guztien antzera, muga handiak zituela. “Sekula ez du gizakia edo itzultzaile profesionala ordeztuko; ezinezkoa liteke. Hala ere, egunerokoan, euskara ikasi edo irakasteko eta, oro har, pertsonen arteko komunikazioan, tresna garrantzitsu bat izango da itzultzailea”. Hitz horiek erabiliz ezagutarazi zuten Baztarrikak lan berria. Google Play-n dago eskuragarri aplikazio hori, eta, IOS sistemako mugikorren kasuan, Apple Store-n.

UZEI sinonimoen/antonimoen hiztegia. Euskarazko lexikoa aberasteko baliabide bat da, testuingurura hobekien egokitzen den hitz zehatza aurkitzeko aukera ematen duena. Hitz batek adiera bakoitzean dituen sinonimoak kontsulta daitezke, edo euskalki zehatz batean, edota hizkuntza-erregistroei erreparatuta. Erabiltzaileak erabaki behar du sinonimo bakoitzaren egokitasuna, eman nahi dion erabilera zein den. Adibidez, euskalki batean edo beste batean erabiltzen den formak ñabardura desberdin bat emango du kasu bakoitzean. Horrela, “*emakume*” *emazteki* izango da lapurteraz/behe-nafarreraz, eta *andrazko* edo *andrakume* bizkaieraz. Emaidza ere desberdina

izango da erregistroaren arabera, jasoan edo lagunartekoa izan baitaiteke. “Pertsona hipokrita” bat *azalutsa* izan daiteke erregistro jasoan, edo *faltsua*, besterik gabe, lagunarteko hizkeran.

Kontuan hartu behar da hiztegi hori ez dela hiztegi orokor bat, eta, beraz, ez dituela biltzen hiztegi orokor baten sarrera guztiak, baizik eta sinonimoak dituzten haiek bakarrik. Eta gauza bera gertatzen da adierekin, sinonimoak dituztenak bakarrik bildu baitira hiztegi horretan. Guztira, sinonimoak dituzten 26.703 sarrera biltzen ditu hiztegiak, eta 30.628 adiera. Bestalde, esan behar da hiztegi hori etengabe eguneratzen dela, ez dela beharrezkoa itxarotea edizio berriak argitaratu arte, egunez egun eguneratzen baita bere datu-basea.

UZEIk, gainera, “Lexikoa eta hiztegiak” izendatu daitekeen esparruari lotutako beste prestazio batzuk ere eskaintzen ditu. Honako hauek dira:

- **TermiGai.** Testu-masa horietan termino-hautagaiak izan daitezkeen terminoak modu automatikoan identifikatzeko eta analizatzeko garatuta dago tresna hori, euskararen eta gaztelaniaren testuren corpusak oinarri hartuta. TerminiGai laguntza handia da terminologia-hiztegi berriak egiteko edo esparru zehatz baten terminologia eguneratzeko; izan ere, terminoetarako hautagaiak ematen ditu, bai euskaraz bai gaztelaniaz, testu-masa handiak modu automatikoan prozesatuz. Analisiaren emaitzak terminologoari laguntzen dio testu-corpus zabalak aztertzen.

- **Koloka.** Baliabide horrek modu automatikoan detektatu eta erauzten ditu hitz anitzeko unitate lexikoak (kolokazioak), bai euskararenak, bai gaztelaniarenak. Garrantzi handikoa da, bereziki euskararentzat, hitz bat baino gehiagoz osatutako unitateen erauzle automatikoa izatea, esaldiko hitz-ordenaren barnean konbinazio ugari izan baititzake hitz anitzeko unitate batek. Hori dela-eta, terminologian esaterako, terminologoarentzat ez da askotan erraza horrelako termino-hautagai guztiak sistematikoki detektatzea bere lanerako darabilen corpusean.

- **LEX2.** Hiztegiak modu erdiautomatikoan sortzeko tresna honek corpus paraleloak erabiltzen ditu testu elebidunak automatikoki detektatu eta erauzteko. Lex2 tresnak ez ditu bilatzen iturburu-hizkuntzako termino batek xede-hizkuntzan izan ditzakeen korrespondentziak. Izan ere, Lex2 tresnaren helburu nagusia ez da terminologiaren erabilera zuzena edo okerra baieztatzea. Lex2 tresnak, hortaz, modu independente batez bilatu eta erauzten ditu termino baterako hautagaiak hizkuntza bakoitzean, eta beste urrats bat eginda, estatistika-paralelismoak bilatzen ditu hautagaien arteko loturak edo erlazioak bilatzeko. Modu horretan, lexiko elebidunak erauzten ditu testu-masa handietan erabilitako terminoetatik abiatuta.

- **TEIS.** Lehen fase batean, terminoek corpusetan zuten ezarpena neurtzea izan zuen helburu; horretarako, zenbait euskarri digitaletan bildutako testuek osatutako corpusak aztertzen zituen. Bigarren fasean,

motorrak erabiliz, Internet corpus birtual baten antzera erabili zuen erabilera erreala neurtzeko testu-masak bildu behar izan gabe.

- **Hizperri.** UZEIk sistema erdiautomatiko bat diseinatu nahi zuen neologismoak detektatzeko, eta helburu horrekin garatutako ikerketen emaitza da tresna hori. Zehazki, hitz ezezagunak detektatzen ditu corpusean, eta analisi morfosintaktiko bat egiten du haien gainean; gero, horretatik abiatuta, egokiak izan daitezkeen lemak proposatzen ditu. Patroiak edo ereduak erabiliz, hitz anitzeko unitate lexikoak detektatzeko ere gauza da. Estatistika-azterketak oinarri hartuta zehaztu ziren patroi eraginkorrenak.



5. irudia: Uzei

- **Elhuyar hiztegia.** Luzea izan da Elhuyarren ibilbidea hiztegien esparruan; izan ere, <https://www.elhuyar.eus/> helbidean dauden doako hiztegiez gainera, beste produktu batzuk ere baditu. Haien artean, nabarmentzekoak dira enpresei eta erakundeei zuzendutako lexikografia- eta terminologia-zerbitzuak. Hona hemen Elhuyarrek eskaintzen dituen hiru hiztegiak: euskara-gaztelania-euskara hiztegia, euskara-ingelese-euskara hiztegia eta euskara-frantsesa-euskara hiztegia. Gainera, **Elhuyar Hiztegiak aplikazioa** sortu du tablet eta smartphoneentzat.

- **Elhuyar Zientzia eta Teknologiaren hiztegia** Elhuyar Fundazioa, gainera, espezializatuta dago zientzia eta teknologiaren arloko itzulpenetan, baita hizkuntzalaritzan ere, betiere esparru horren barnean. Zientzia eta teknologiaren hiztegi entziklopedikoa dugu hasiera batetik esparru horretan metatutako ezagutzaren emaitza; doan kontsultatu daiteke honako helbide honetan: zthiztegia.elhuyar.org/.

- **Lexikoaren behatokia.** Euskararako sortutako tresna bat da, Euskaltzaindiaren laguntzaz egina. Komunikabideek argitaratzen edo emititzen duten materialarekin corpus monitore bat eraikitzea da

helburua, hau da, hizkuntzaren erabilera hurbiletik jarraitzea ahalbidetuko duen corpus bat elikatzen eta automatikoki lematizatzen eta etiketatzen joatea.



6. irudia

Opentrad. Kode irekian oinarrituta dago, eta 2006an abiarazi zen sistema; adierazten zitzaizkion testuak itzultzen zituen, baita webguneak ere, Interneten nabigatu ahala. Bere prestazioen artean, nabarmentzekoa da Opentrad orrialdetik ere testuak itzultzeko aukera dagoela. Horrela, kode irekiko edozein garapen-enpresak integra dezake Opentrad bere sistemetan. C-n programatutako bi itzulpen-motor, abiadura handikoak, dira sistemaren oinarria: haietako batek hizkuntza hurbilak (gaztelania-katalana eta gaztelania-galegoa) direnekin lan egiten du transferentzia sintaktiko partzialaren bidez; eta besteak transferentzia sintaktiko osoa erabiltzen du elkarrengandik urrun dauden hizkuntza-bikoteetan, hala nola gaztelania-euskara bikotean. Sistema osatzen duten hizkuntza guztien datu linguistikoak –bai hizkuntza bakarrekoak, bai hizkuntza anitzekoak– XML formatuan biltzen dira, diseinu modular batekin; diseinu horrek datuen arteko elkarrengingarritasuna errazten du, baita programaren fluxuan integratzea ere.

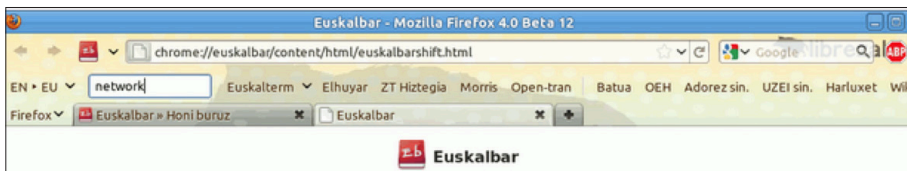
GOOGLE itzultzailea. Doako zerbitzu bat da, hitzak, esaldiak eta webguneak itzultzen dituena gaztelaniatik euskarara eta beste 99 hizkuntzatar.



7. irudia: Google itzultzailea

Euskalbar (<http://euskalbar.eu>). Nabigazio-barraren azpian instalatzen den sistema bat da, eta zazpi euskal hiztegi ere kontsultatzeko aukera ematen die internautei, bai euskaraz, bai gaztelaniaz. Juanan Pereira-k* sortu zuen Euskalbar (<http://www.euskalkultura.com/espanol/noticias/euskalbar-3-0-ayuda-para-traducir-entre-euskera-castellano-e-ingles-puede-descargarse-desde-su-web>), eta dohainik eta modu librean argitaratu zuen internet&euskadi-n. Harrezkeroztik zenbait hobekuntza egin dira, eta 3.0. bertsioa da azkenekoa.

Bertsio berri horrek honako hiztegi eta datu-base hauek kontsultatzeko aukera ematen du: Euskalterm, 3000 Hiztegia, Elhuyar, Morris, Open-tran, Hiztegi Batua, Adorez sinonimoak, UZEI sinonimoak, ItzuL posta zerrenda, Harluxet, Mokoroo, Intza, Eurovoc, Bergarako hiztegia, ZT Corpusa, CorpEus, Openrad eta XUXENWeb.



8. irudia: www.euskalbar.eu

2. DESAGERTU EGIN DIREN ITZULTZAILEAK

Azken urteotan, euskarazko itzultzaile batzuek funtzionatzeari utzi diote, nahiz eta ibilbidearen hasieran indartsu eta hazteko gogoz aritu ziren. Hala eta guztiz ere, zerbitzua utzi behar izan dute zenbait arrazoi tarteko. Honako itzultzaile hauek dira: Cervantes institutua eta Erdaratu.eu.

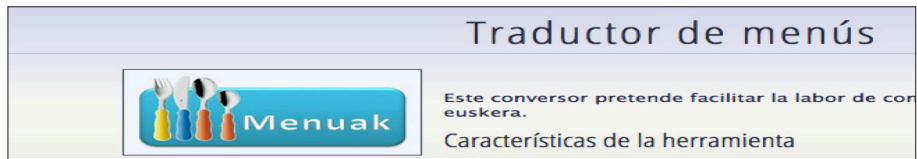
Cervantes institutuaren itzultzailea. Cervantes institutuaren hizkuntza-teknologiaren atariaren parte gisa, erakunde horrek itzulpen automatikoaren zerbitzu bat eskaini zuen 2005eko urtarriletik 2013ko abenduaren 15era arte, testuak eta web-orriak gaztelaniatik hizkuntza hauetara itzultzeko aukera ematen zuena: katalana, frantsesa, galegoa, ingelesa, portugesa, baita euskaraz ere, baina norabide bakarrean, gaztelaniatik euskarara.

Erdaratu.eu. 2008an, euskaratik gaztelaniara itzultzen zuen lehen itzultzailea sortu zen; software libreko proiektu bat zen (GPL lizentziarekin), Apertium proiektuan oinarrituta dagoena. Halere, artikulu hau egiten den unean³, erdaratu.eu orria ez zegoen erabilgarri, eta, hortaz, ezin izan dugu aztertu bilatzaile horrek eskaintzen dituen aukerak.

3. 2016. urtea

3. ITZULTZAILE TEMATIKOAK

Onlineko menu-itzultzailea euskaraz. 2008ko uztailean, Donostiako Udal Patronatuak informatika-tresna bat sortu zuen taberna eta jatetxeetako menuak eta kartak gatzelaniatik euskarara itzultzeko. Halako eskaintza egiten zuen lehen itzultzailea izan zen Interneten, eta, euskal ostalari eta merkatariak sortua bazen ere, edozeinek erabil zezakeen eta dezake. Guztira 1.057 termino biltzen ditu. Merkatari.net webgunean dago itzultzaile hori: www.merkatari.net. Itzultzaileak zenbait iturri erabiltzen ditu bere corpusa osatzeko: Merkatari.net webguneko Euskara Platerean ataleko menu-sorta eta hiztegia eta Gastronomiavasca.net eta sukaldean.com webguneetako funtsak.



9. irudia

Eitb-ren sukaldaritzaitzultzailea (<http://www.eitb.eus/eu/kultura/euskara/kontsultak/>) Eitb-ren sukaldaritzaitzultzailea HIKEA aplikazioaren eta Eitb-ren webgunearen bidez kontsultatu daiteke. 300 sarrera ditu, doakoa da eta edozein interesdunen esku dago. Hiztegiak Txoriene ETB1-eko sukaldaritzaitzultzailearen lana du oinarri, eta bere ardatza saio horietan prestatu diren plateren izenak dira. Hona hemen adibidea: “verdel-tartarra lima eta sojadun ozpin –oliotan” .



10. irudia: Eitb-ren sukaldaritzaitzultzailea

4. EROSKI CONSUMER KASUA, ELKARLANEKO PROIEKTU BAT

4.1. Eroski Consumer proiektuaren aurkezpena

1996an, Eroski enpresa sortu zen zenbait kontsumitzaile eta langileren ekimenez. Milaka kontsumitzailek eta ehun bat langilek kooperatiba bat sortzea erabaki zuten, helburu bikoitzarekin: kalitateko produktuak preziorik onenean lortzea eta kolektibo gisa zituzten interesak defendatzea. Urte hartan berean, Eroskik bere etekinen %10 gizarteari itzultzen hasi zen, Eroski Fundazioan zehar bideratuta. Fundazioaren ekimen sozialak hiru ekintzarantz handiren barnean biltzen dira, hain zuzen, honako hauen barnean: kontsumitzailearentzako informazioa eta prestakuntza; iraunkortasuna eta ingurumenarekiko konpromisoa; eta elkartasuna. Eta ekimen sozial guztien artean, Eroski Consumer informazio-proiektua da Fundazioaren adierazgarri nagusia. 1974ko martxoan, Eroski enpresa sortu eta handik bost urtera bakarrik, Eroski Consumer aldizkariaren lehen alea argitaratu zen, orduan Eroski izenarekin. Urte batzuk geroago Consumer izan zen, gero Consumer Eroski, eta gaur egun Eroski Consumer du izena.

Proiektu horrek hiru zati handi ditu: inprimatutako aldizkaria, www.consumer.es onlineko aldizkaria, eta bai paperean bai webgunean argitaratzen diren gidak eta eskolak. Hiru informazio-proiektu handi horien bidez, proiektuak bere asmoa betetzen du: kontsumitzaileari laguntzea haren egunerokoan eragina duten esparru guztietan, informazioa, prestakuntza eta parte-hartzea kontuan hartzen dituen lan baten bitartez. Eroski –laugarren banaketa-katea Espainian– entzuleria-liderra izan da azken urteotan, hasiera batetik egin duen apustua dela-eta: etekinen zati bat bideratzea gizarteari laguntzera. Kontu ugari irakatsiz gauzatzen da laguntza hori, adibidez nola elikatu modu osasungarrian, nola aurreztu krisialdian, zer eskubide dituzten herritarrek aseguruak kontratatzean, eta nola eduki elkartasun eta errespetu handiagoa ingurumenarekin, besteak beste. Hori guztia informazio zorrotza eta kontrastatua argitaratuz.

Kanal guztietatik, Elikadura eta Osasuna dira artikulu gehien biltzen dituztenak, hau da, ongizateari lotutakoak.



11. irudia

4.1.1. Consumer.es

Horrela, kontsumitzailearen webgunean –consumer.es–, erabiltzaileak informazio eguneratua aurkituko du interesatzen zaizkion gaiei buruz: elikadura eta osasun arloko aholkuak, etxeko ekonomia, 3.500 errezetatik gora ohar nutrizional eta dietetikoekin, bideo-errezetak, infografia erabilgarriak, bideoak eta app-ak modu grafikoa eta bisuarean informatzeko, baita gidaliburu praktikokoak ere.

Gidaliburu praktikoei dagokienez, esan daiteke une hauetan Eroski Consumer baliabideak 13 gidaliburu praktikoko dituela, Internetek dohainik kontsultatzeko moduan daudenak, baita zazpi eskola ere; eskola horien artean, bi nabarmentzen dira: haurren obesitatea prebenitzea (<http://obesidadinfantil.consumer.es/>) eta erretzeari uztea (<http://escuelas.consumer.es/web/es/dejardefumar/>).



12. irudia: consumer.es

4.1.2. Proyecto multilingüe

2007tik paperezko aldizkaria euskara plazaratzen da, gallegoz eta katalanez gain. Dena den, paperezkoa baino lehen, Interneten bazeuden Consumer Eroskiren edukinak euskaraz. Duela hamar urte 1998ko martxoan hasi zuen ibilbide hori aldizkariak. Aspalditik bazegoena orain paperera dator, hizkuntza ongi zainduta. Edukina berbera da baina euskaraz kontatua eta estilo jakin batekin: itzulpen hotzetatik aldegin eta interpretaziora jo. Horixe izan da erronka nagusia, hizkuntza ulerterreza lortzea, berez datorrena eta xamurra, hizkuntzaren tradizioa ahazten ez duena baina kazetaritza modernoaren moldeak ere buruan dituena. Azken hilabeteotan euskarazko edizioaren arduradunak izan duten lehen zailtasuna Consumer Eroskik bere dituen gaietan euskaraz dagoen tradizio eskasa izan da. Lexikoa falta da, batzutan. "Hirigune historikoen bizi baldintzak aztertu ditugu, eguzki betaurrekoen ezaugarri teknikoak buruz hitzegin dugu eta orbainik utzi gabe egiten diren ebakuntzen inguruan eta nola egin dugun, ba, esaldi luzeak atzean utzita, ahozko lengoia idatzia erabilita eta ahal izan den guztietan euskarak

eta euskal kulturak bere dituen esamoldeak, hitz joku eta irudiak erabilia", Agustin Marcaide, Eroskiko zuzendariaren hitzak.

Adibide moduan, "lanak erreta" titulua erabili da "Burn out" sindromeak jotako gaixoei buruz hitzegiteko. Eta titular politak bilatu dira, erakargarriak eta euskararen aukerak aprobetxatzen dituztenak. Esate baterako, eguzki betaurrekoak konparatu dira ale honetan eta garestienak ez direla hobeak izan da ondorio nagusia. Titulua beraz, hauxe: "modak itsutzen gaitu eta betaurreko txarrek". Jabe bat baino gehiago dituzten etxeei buruzko artikulua, halako kontratuen arazoak aztertzen zituen eta jarri zaion tituluak hitz jokua egiten du, halaxe: "Giltza gehiegi ate bakarrarentzat", Euskaraz idatzi nahi izan dute arduradunek, alegia, euskaldunentzat. Batzutan zaila izan da. Bela arinari buruzko infografia, esate baterako, euskaraz tradizio eskasa duen gaiari buruz aritzen da. Itsasoko hiztegia kontuan hartu behar izan eta era berean, hiztegia egin, kirol honi buruz azalpenak emateko modua bilatu, teknikoa bai, baina ulergarria.

4.2. Consumer.es itzultzailea

Onlineko baliabide hori sortu zenetik, milioika eduki, informazio, artikulua, erreportaje, infografia eta monografiko argitaratu izan dira, gehienbat gaztelaniaz. Halere, euskalduntze handiagoa lortu nahi duen proiektu baten parte bezala, 2016an, Eroski Fundazioak ekimen berri bat abiarazteko erabakia hartu zuen: consumer.es webgunean argitaratutako informazio guztiak modu automatikoan itzultzea. Hizkuntzek maila desberdina dutenez garapen teknologikoari dagokionez, proiektuan bi soluzio desberdin planteatu dira: gaztelania - euskara, Matxin bidez; eta gaztelania -katalana eta gaztelania - galegoa, Opentrand bidez, zeina Espainiako lau hizkuntza ofizialen (gaztelania, galegoa, katalana eta euskara) itzultzaile automatikoa baita, Euskal Herriko, Galiziako, Kataluniako eta Valentziako Erkidegoko unibertsitateek garatutakoa.

Itzultzaile automatiko hobetua da, software librean oinarritutakoa, eta aukera ematen dio consumer.es webguneari eduki guztiak euskaraz izateko lasterrago, edozein erabiltzailek haiek euskaraz kontsultatu eta irakurtzeko aukera izan dezan. Helburu horrekin, Eroski Fundazioa elkarlanean ari da Elhuyar Hizkuntza eta Teknologiarekin eta EHUrekin, IXA taldearen bitartez. Euskal gizarteari kode libreko itzultzaile automatiko hobetua eskaintzea da helburua; eragin sozial handiko itzultzailea, gainera, erabiltzaileen parte-hartzea bilatzen baitu emaitzak hobetzeko.

Alderdi garrantzitsuenak

- **Itzultzaile automatikoa.** Matxin baliabidean (Elhuyarrek eta IXA taldeak sortua) oinarritutako itzultzailea da, eta consumer.es webguneko eduki guztiak gaztelaniaz eta euskaraz lortzeko aukera ematen du.

- **Elkarlaneko proiektua.** Ezinbestekoa da erabiltzaile euskaldun guztien parte-hartzea, consumer.es webguneko edukien euskaratze automatikoaren emaitza hobetzeari begira. Erabiltzaileek horretarako propio diseinatutako plataforma bat izango dute erabilgarri, hau da, elkarlaneko interfaze bat sortuko da.

- **Sistema adimentsua.** Matxinen itzulpen automatikotik abiatuta eta erabiltzaileen zuzenketa erantsita, sistemak bildu egingo du informazioa bere akatsetatik ikasi eta erabiltzaileei kalitate handiagoko emaitzak eskaintzeko.



13. irudia

4.3. Hizkuntza-baliabideak

Jakin badakigu, zuzenketa egitean zalantza izatea oso ohikoa dela, eta, horregatik, zuzenketa-lanetan lagungarriak izango diren hainbat baliabide jarri ditugu erabiltzailearen eskura. Erabiltzaileak oso erraz kontsultatu ahal izango ditu baliabideok, Consumer webguneko edukia zuzentzen ari den bitartean.

4.3.1. Hiztegiak: orokorrak eta terminologiak

Testuak itzuli edo zuzendu behar direnetako ezinbesteko baliabide bihurtu dira hiztegiak eta korpusak⁴, eta horregatik zuzenketa egiteko pantailan bertan txertatu dira baliabideok: batzuk zuzenean kontsulta daitezke, esaterako Elhuyar hiztegia eta Consumer korpusa; beste baliabide interesgarri batzuetarako esteka zuzenak jarri dira.

Hauek dira zuzenketa-plataforma honetatik bertatik atzitu daitezkeen hizkuntza-baliabideak:

- Elhuyar hiztegiak. Euskara/Gaztelania –Español/Vasco hiztegi orokorra
- Consumer Korpusa. Consumer Eroski aldizkariaren edukietan edozein hitz bilatzeko aukera ematen dizu euskaraz, gaztelaniaz, galegoz edo katalanez.

4. BILBAO TELLETXEA, G. y MAKAZAGA EIZAGUIRRE, J.M. EHUskaratuak: un corpus de traducciones académicas en una lengua minoritaria, «Monumenta linguae Vasconum (IV)» (FFI2012-37696). pág. 324.

- Euskaterm. Euskal Terminologia banku publikoa. Terminoak euskaraz, gaztelaniaz, frantsesez eta ingelesez bilatzeko aukera ematen du.
- Gastronomía Hiztegia. Gastronomiazko Euskal Anairtearen euskara/gaztelaina hiztegia.
- Sukaldaritzazko hiztegia. Leioako Ostalaritza Eskolaren euskara/gaztelaina hiztegia.

4.3.2. Gramatika baliabideak

Proiektu honetan lantzen den informazioaren egitura eta ezaugarriak ikusita, beharrezkoa ikusi da gramatika-baliabide batzuk erabiltzailearen eskura jartzea. Euskaraz idaztera ohituta egon arren, sarritan izaten ditugu gramatika-arloko zalantzak, eta testuinguru honetara egokitzen diren baliabideen zerrenda egin da erabiltzaileei lagundu asmoz:

- Zenbait laburdura erabilgarri
- Zenbaki bereziak
- Puntuazio-markak
- Nola idatzi data eta idazkera
- Hitz elkartuen idazkera
- Deklinabidea
- Euskaltzaindiak araututako onomastika

4.4. Nola zuzendu testuak?

Consumer webguneko eduki bat zuzentzera sartzen garenean, pantaila hau ikusiko dugu:



Pantailaren goialdean, artikulua edo edukiaren inguruko informazioa ikusiko dugu, esaterako, jatorrizko testuaren hiz-kopurua. Informazio horrekin erabiltzaileari zuzentzeko testuaren bolumena zein den adierazi nahi zaio; hau da, errepasatu beharko testua luzea edo motxa den. Hizkuntza-baliabideen zerrenda ere goiburuan egongo da, eta uneoro egongo da bistan, erabiltzaileari zalantzak kontsultatzeko erraztasuna eskainiz.

Ezkerreko zutabearen zuzendu nahi dugun artikulua jatorrizko testua agertuko da (gaztelaniaz), eta, eskuinean, sistemak automatikoki itzulitako estua (euskaraz). Itzulpenaren jatorriaren arabera, jatorrizko testuaren atzealdeko kolorea ezberdina izango da:

- Kolore laranja, itzulpena itzulpen automatikotik ekarria izan dela esan nahi du.
- Kolore berdeak, itzulpena itzulpen-memoretatik (zuzendutako testuetatik) ekarria izan dela esan nahi du.
- Kolore horiak, testuaren %75a baino gehiago itzulpen-memoretatik (zuzendutako testuetatik) ekarria izan dela esan nahi du.

Eskuinaldeko, itzulitako testuaren gainean klik eginez testua zuzentzen has gaitezke. Itzulitako esaldietan gora eta behera mugitzeko sagua, geziak edo tekla-konbinazio bat (ALT+UP) edo (ALT+DOWN) erabil daitezke. Behin hurrengo itzulpen testua pasata, zuzendutako testua automatikoki gordeko da.

Itzulitako testu guztiak errepasatu ondoren, eta lana bukatutzat eman bada, “Amaitu” botoiari eman behar zaio. Momentu horretan, zuzenketa-lanak bukatutzat ematen dira, eta kudeatzaileak lana onartzeko zain geldituko da; izan ere, erabiltzaile batek eduki guztiak eskuz errepasatu arren, ez da jatorrizko webgunean modu automatikoan agertuko, kudeatzailearen azken oniritzia beharrezkoa izan baitu webgunean argitaratzeko. Baina, edozein arrazoigatik zuzenketa-lanak denbora batez utzi nahi baditugu, gerora artikulua behera itzultzeko amotan, “itxi” botoiari eman behar zaio. Honela, artikulua hori “Nire lanak” atalean geratuko da. Erabiltzaileak edozein momentutan berriro zuzenketa-lanekin jarraitzeko aukera izango du, tituluaren gainean klik eginez.

5. ONDORIOAK

1.- Aztertutako itzultzaileek, kontsultatutako hitzaren emaitza erakustez gainera, prestazio gehiago ere eskaintzen dituzte. Xuxen baliabidean, nabarmentzekoak dira zuzenketa-lanekin atalean egindako hobekuntzak. Horrela, pertsona batek oker idazten badu hitz bat, zenbait aukera biltzen dituen zerrenda bat eskaintzen dio itzultzaileak, betiere proposatutako terminoari lotutako akats ohikoenekin bat. Aurrerapen handia da hori, erabiltzaileak denbora aurrezten baitu, eta, gainera, emaitzaren kalitatea ziurtatzen da. Halere, ez du aurreikusten inongo elkarlaneko plataformarik, erabiltzaileen interakzioa lantzeko aukera ematen duenik, makinak egindako itzulpena hobetzeari begira.

2.- Eusko Jaurlaritzak, itzultzaile automatikoaren aurkezpenean, agerian utzi zuen jada tresna horrek ematen duen emaitza hobetu beharra. Horrek esan nahi du itzultzaile automatiko hori ez dela perfektua –ezta hurrik eman ere– eta beharrezkoa dela itzuli nahi den hitza gramatikalki perfektu idatzita egotea, eta ziurtatzea terminoaren erraztasuna. Hortaz, beharrezkoa da pertsona batek azken emaitza orraztea.

3.- Hiztegi guztiek bikain funtzionatzen dute, emaitza berdina eskainita. Halere, Elhuyarrek bakarrik eskaintzen du zerrenda zabalgarri bat, aukeratu-takoaren antzekoak direnak erakusten dituena emaitza ez badu berehalakoan aurkitzen.

4.- Makinak emandako erantzuna hobetzeko aukerari dagokionez, kontsumitzailearen webgunean –consumer.es– garatzen ari den hiztegia da halakorik eskaintzen duen bakarra, euskaldunei plataforma baten bidez haien harri-koskorra ekartzeko aukera ematen baitie itzulpena hobetze aldera. Horrela, onlineko itzultzaileak “patata ilea moztu” eskaini badu ere “pelar una patata” itzultzeko eskatu zaionean, itzulpen hori ikusten duen edozein pertsonak itzulpena hobetzeko aukera du “ilea moztu” terminoaren ordez “patata zuritu” jarriz eta aldaketa gordez; ia berehalakoan argitaratzen da emaitza berria. Elkarlaneko interfaze bat erabiliz gero, itzulitako terminoak hobetzeaz gainera, hizkuntzaren kalitatea kontrolatzen duen sistema bat ere lortzen da; gainera, ziurtatzen da Euskaltzaindiaren azken arauak betetzen direla, eta abian jartzen da gure hizkuntzarekin eta kulturarekin konprometitu dagoen euskaltzaleek osatutako komunitate bat.

6. BIBLIOGRAFIA

BILBAO TELLETxea, G. y MAKAZAGA EIZAGUIRRE, J.M. *EHUskaratuak: un corpus de traducciones académicas en una lengua minoritaria*, «Monumenta linguae Vasconum (IV)» (FFI2012-37696).

LINDEMANN, D. (2014). Creating a German-Basque electronic dictionary for German learners. *Lexikos*, 24(1), 331-349.

LETURIA, I. (2012). Evaluating Different Methods for Automatically Collecting Large General Corpora for Basque from the Web. In *COLING* (pp. 1553-1570).

GARCÍA DEL TORO, C. (2015). La investigación sobre la traducción entre catalán y español: estudios traductológicos. *Meta: Journal des traducteurs/ Meta: Translators' Journal*, 60(3), 557-575.

JAUREGI, A. 2012. *EHUskaratuak. Corpus eleaniztun baten etiketatze automatikoa eta kontsulta-sistemaren garapena* [Tesis de máster]. Disponible en <http://ixa.ssi.ehu.es/master/master_tesiak>. Consultado el 27/5/2014

<http://www.euskara.euskadi.eus/r59-euskalte/eu/q91EusTermWar/kontsultaJSP/q91aAction.do>

<http://matxin.elhuyar.eus/>

Rodríguez González. M. M. [et al.]: Itzultzaile automatikoak euskaraz, bilakaera eta prestazioak

<https://www.elhuyar.eus/es>

<http://www.opentrad.com/es/?>

**NOTA: Sailkapen honetan EHUren Egungo Euskararen Hiztegia ere aipatu behar da. Euskara Institutuak Laboral Kutxaren babesarekin garatzen dabilen proiektu honek euskera-euskera hiztegi-rik zehatzena da eta 60.000 sarrera eta 200.000 azpisarrera dauzka dagoeneko.*