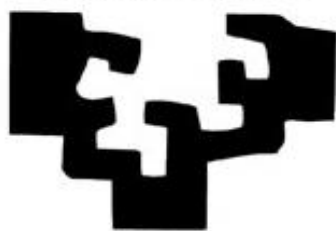


eman ta zabal zazu



Universidad
del País Vasco

Euskal Herriko
Unibertsitatea

GRADU AMAIERAKO LANA

ITZULPENGINTZA ETA TEKNOLOGIA

EUSKARAZKO POSTEDIZIOA: AURREIRITZIAK, APLIKAZIOA ETA
GOGOETAK ITZULPENGINTZA ETA INTERPRETAZIOKO IKASLEEN ARTEAN

Irati Irurzun Biain

Itzulpengintza eta Interpretazioko Gradua

2017/2018 ikasturtea

Tutorea: Elizabete Manterola Agirrezabalaga

AURKIBIDEA

IRUDIEN AURKIBIDEA	4
GRAFIKOEN AURKIBIDEA	5
LABURPENA	6
SARRERA.....	8
ZER DA POSTEDIZIOA?	10
Postedizio arina.....	11
Postedizio osoa	11
ITZULPEN AUTOMATIKOA ETA POSTEDIZIOA EUSKAL HERRIAN:	
TESTUINGURUA	13
ZERGATIK GRADU AMAIERAKO LANA POSTEDIZIOARI BURUZ?	18
METODOLOGIA ETA PROBAREN DISEINUA.....	20
Prestakuntza	21
Parte-hartzaileak biltzea.....	26
Esperimentua edo ataza	26
IKERKETAREN MUGAK	30
EMAITZEN AZTERKETA: LEHEN GALDETEGIA, ITZULTZAILE	
AUTOMATIKOAREKIKO (AURRE)IRITZIAK ETA PARTE-HARTZAILEEN	
OHITURAK EZAGUTZEA.....	31
1. Inoiz IA erabili duen parte-hartzaile kopurua.....	31
1.1. Zein IA erabili izan dute gehien parte-hartzaileek ES-EU konbinazioan? ..	32
1.2. Zer erabilera eman diote parte-hartzaileek IAri?	33
1.3. ES-EU bikotean IA ez erabiltzeko arrazoiak	34
1.4. IA beste hizkuntza-bikote batzuetan	35
2. Lagungarri izan liteke IA itzulpen profesional bat egiteko?	36
3. Postedizioa: produktibitatearen gaineko iritzia	38
POSTEDIZIO PROBAN ATERATAKO DATU OROKORRAK.....	40

EMAITZEN AZTERKETA: BIGARREN GALDETEGIA, PROBA ONDORENGO ONDORIOAK ETA IRITZIAK EZAGUTZEA	43
1. Itzultzaileen iritzia itzulpen kalitatearen inguruan	43
2. Itzultzaileen aurreikuspena itzultzeko edo posteditatzeko behar izan duten denboraren inguruan	45
3. Itzultzaile automatikoaren erabilgarritasuna ataza honetako itzulpen-prozesuan	46
4. Zer beste baliabide erabili dituzte parte-hartzaileek itzulpen-prozesuan?.....	47
5. Parte-hartzaileen iruzkinak eta gogoetak (hautazko galdera).....	48
ONDORIOAK	49
BIBLIOGRAFIA	51
ERANSKINAK	53
I. ERANSKINA: LEHENENGO GALDETEGIA	53
II. ERANSKINA: BIGARREN GALDETEGIA.....	54
III. ERANSKINA: JATORRIZKO TESTUA	55
IV. ERANSKINA: ERREFERENTZIAZKO ITZULPENA.....	56
V. ERANSKINA: ITZULPEN AUTOMATIKOA (MATXIN).....	57
VI. ERANSKINA: ITZULPEN AUTOMATIKOA (GOOGLE TRANSLATE)	58
VII. ERANSKINA: ITZULPEN AUTOMATIKOA (EUSKO JAURLARITZA)	59

IRUDIEN AURKIBIDEA

Irudia 1 Modela Proiektuaren demoa	16
Irudia 2 Metrika automatikoen emaitzak.....	23
Irudia 3 TAUSen proiektua sortzeko.....	24
Irudia 4 TAUSen proiektua sortzeko.....	25
Irudia 6 Posteditatzeko interfaza	28
Irudia 7 Itzultzeko interfaza.....	28
Irudia 9 Probako emaitza orokorrak	40
Irudia 10 Probako parte-hartzaile bakoitzaren emaitzak	42

GRAFIKOEN AURKIBIDEA

Grafikoa 1 Erabili duzu inoiz itzultzaile automatikoa?	31
Grafikoa 2 Zein IA erabili duzu?.....	32
Grafikoa 3 Zertarako erabili izan duzu IA?.....	33
Grafikoa 4 Zergatik ez duzu inoiz euskarazko IA erabili?.....	34
Grafikoa 5 IA beste hizkuntza-bikote batzuetan	35
Grafikoa 6 Lagungarri izan liteke IA itzulpen profesional bat egiteko?	36
Grafikoa 7 Posteditatzen hutsetik itzultzen baino denbora gehiago edo gutxiago behar da zure ustez?.....	38
Grafikoa 8 Noiz izan da hobe, zure ustez, itzulpenaren kalitatea?	43
Grafikoa 9 Noiz behar izan duzu denbora gehiago?.....	45
Grafikoa 10 Erabilgarria izan da IAren laguntza?.....	46
Grafikoa 11 Bestelako baliabideak.....	47

LABURPENA

Euskarazko postedizioa izan da lan honen ardatza. Gure xedea euskarazko postedizioa zertan den aztertzea eta etorkizunean itzultzaile profesional izango direnen artean horren inguruko gogoeta bultzatzea da. Zehazki, erabiltzaileen pertzepzioan jarri dugu fokua. Horretarako, proba bat diseinatu dugu eta UPV/EHUko Itzulpengintza eta Interpretazioko graduako hiru eta laugarren mailetakoko 22 ikaslerekin gauzatu dugu. Alde horretan da berritzailea gure ikerketa, aurretik eginiko ikerketetan inoiz aztertu ez den kolektibo bati jarri baitiogu arreta. Hasierako sarreraz eta ondorioez gain, hiru atal nagusitan antolatu dugu lan hau.

Lehenik, gaiaren inguruko azalpen teorikoa eskaini dugu. Postedizioaren definizio bat ematen saiatu gara eta postedizio jarduera baldintzatzen duten faktoreak zeintzuk diren ikusi dugu. Euskal postedizioa zertan den ere aztertu dugu eta orain arte gaiaren inguruan egin diren azterketak aipatu ditugu.

Bigarrenik, alderdi teorikoa landu ondoren, lanaren metodologia deskribatu dugu xehe-xehe. Lan honetan bi informazio-iturri erabili ditugu nagusiki: parte-hartzaileek proba aurretik eta proba ondoren erantzun beharreko galdetegi bana eta postedizio ataza. Horretarako eman dugun urrats bakoitza azaldu dugu atal honetan, probaren prestakuntzatik hasita, parte-hartzaileak biltzera eta esperimientua gauzatzera arte.

Hirugarrenik, probatik ateratako emaitzak aztertu ditugu. Hirugarren atal honen barruan hiru azpiatal bereiz ditzakegu: lehen galdetegitik ateratako ondorioei dagokien atala, probako produktibitatearen azterketari dagokion atala, eta bigarren galdetegitik ateratako ondorioei dagokiena.

Ondorio nagusiak aipatu ondoren, lanaren marko teorikorako oinarri hartu ditugun lanak bildu ditugu bibliografia atalean, eta lanerako funtsezko izan zaizkigun dokumentu batzuk jarri ditugu eskuragarri eranskin gisa: parte-hartzaileek erantzun beharreko bi galdetegiak, proban itzuli beharreko testua gazteleraz, metrika automatikorako erabilitako erreferentziazko itzulpena eta hiru itzultzaile automatikoen itzulpena.

Azkenik, garrantzitsua da azpimarratzea gure ikerketa honen emaitzak ezin direla behin-betikotzat jo, 22 ikasleko lagin txiki batekin egin baitugu proba eta ezin baitira, inondik inora ere, guk ateratako emaitzak graduako ikasle guztietara orokortu, are gutxiago itzultzaile guztietara. Hala eta guztiz ere, euskarazko postedizioaren inguruko ideia

orokor bat egiteko balioko digu proba honek, eta interesgarria litzateke arlo honetan ikertzen jarraitzea, euskal itzulpengintzaren oraina eta geroa deskribatzen lagunduko duelako eta itzultzaileen jardunaren eboluzioa hobeto ulertzen lagunduko duelako.

SARRERA

Globalizazioa eta nazioartekotasuna inoiz baino garrantzitsuagoak diren garai honetan, testuak itzultzeko beharra areagotzen ari da etengabe. Behar horrekin batera, itzulpengintzari lotutako teknologia ugari ari dira garatzen, itzulpen-lana errazago, azkarrago, eta, beraz, efizienteago bihurtze aldera.

Informatikaren garapenak errotik aldatu du itzulpengintza-jarduera: testu-prozesatzaileei esker itzulpenean nahi beste egokitzapen egin ditzakegu, internetek hiztegiak, corpusak, tesaurusak eta bestelako baliabideak jartzen dizkigu eskura, eta itzulpen-memoriak OLI tresnetan txertatzeari esker aurreko itzulpenetan egindako lanaz balia gaitezke, eta, beraz, alferrikako lana saihestu. Bestela esanda, eraginkortasuna hobetu.

Hala ere, badago zenbait hizkuntzarekin lan egiten duten enpresa askok egiten duten hautu bat itzulpen-prozesua azkartze aldera, baina euskarazko itzulpen-fluxuetan oraindik txertatu ez dena: itzulpen automatikoa (IA) eta itzulpen-memoriak uztartu eta posteditatzea.

Postedizioa graduan landu ez dugunez, lan honetarako dokumentazioa maila ugartan egin behar izan dugu: alderdi akademikoari begira, postedizioaren inguruko informazioa bildu dugu artikuluak eta gaiaren inguruko bestelako dokumentuak kontsultatuz. Alderdi profesionalari begira, postedizioak merkatuan duen presentzia gertutik ezagutu dugu. Horretarako, ISEA Berrikuntza eta Ekintzailetasun Zentruko Igor Ellakuriarekin eta gaia lantzen ari den Elhuyarreko Itziar Cortesekin bildu ginen. Igor Ellakuriak aipatu zigun gaur egun itzultzaile automatikoa eta postedizioa uztartzea oso ohikoa dela mundu mailako itzulpen enpresetan. Hala ere, euskararen kasuan oraindik garapen bidean dagoen berrikuntza bat dela esan zigun, eta oraindik ez dela euskararekin lan egiten duten itzulpen enpresetan txertatu. Itziar Cortesi esker Consumer proiektua gertutik ezagutu eta postedizioan sakondu genuen. Azkenik, ikerketari begira, zer motatako azterketa egin behar genuen aztertu dugu, artikulua akademikoak irakurrita eta Nora Aranberri EHUKo irakasle eta ikertzailearekin hitz eginda.

Hau izango da gure lanaren egitura: Gaiaren muinari heldu aurretik postedizioari buruzko oinarrizko informazioa aurkeztuko dugu: zer den, zer faktorek baldintzatuko duten postedizio jarduera eta zeintzuk diren postedizio motak. Jarraian, euskal postedizioaren hurbilpen bat eskainiko dugu, arreta berezia eskainiko diogu orain arte

arlo honetan egin diren proiektuak edo aurrerapausoak azaltzeari eta gaur egun ematen zaion erabilera ezagutarazteari.

Testuingurua kokatu ondoren, lan hau egitera zerk bultzatu gintuen azalduko dugu. Horretarako, postedizioa landu duten eta eredu izan ditugun ikerketa batzuk aipatuko ditugu. Gure ikerketak berezitasun bat izango du kontsultatu ditugun beste ikerketen aldean: etorkizunean itzultzaile izango diren ikasleek itzultzaile automatikoarekiko duten ikuspuntua aztertzea.

Behin alderdi teorikoa landuta, lanaren muinari helduko diogu. Itzulpengintza eta Interpretazioko graduiko ikasleekin egindako proba bat izan da gure aztergaia lan honetan. Lehenik, metodologia eta probaren garapena zein izan zen argituko dugu urratsez urrats, eta ondoren probatik ateratako ondorio nagusiak aztertuko ditugu.

ZER DA POSTEDIZIOA?

Itzulpen automatiko bidez burututako lan bati gizakiak egokitzenak egitea da posteditzioa, azken produktuak esperotako eta bezeroarekin adostutako ezaugarriak bete dituzan; kalitatea, esaterako (TAUS *Translation Technology Landscape Report*: 2016). Posteditzioaren abantaila nagusia denbora eta, horrenbestez, dirua aurrezteko aukera ematea da, denbora berean lan-karga handiagoa itzultzen laguntzen baitu. Eta hori da, hain zuzen ere, mundu mailan hainbat alorretan gailentzen ari den teknika honen arrakastaren arrazoia.

Posteditzioan jarduteko itzulpena ebaluatzeko gai izan behar da eta itzultzaileak zein zuzentzaileak dira horretarako ondoen prestatuta dauden profesionalak. Hala ere, posteditzioan lan egiteko behar diren trebetasun bereiziak kontuan izanik, ezinbestekoa izango da etorkizunean posteditzioan arituko diren profesionalen gidalerro batzuk eskaintzea, posteditzio jarduerak bere-bereak dituen trebetasun horiek lantzeko.

Bi faktoreren arabera izango dira posteditzioan aritzeko oinarritzko gidalerroak eta profesionalen egin beharreko esfortzua: IAren kalitatea batetik, eta posteditatutako lanaren esperotako kalitatea bestetik (TAUS *Machine Translation Post-editing Guidelines*: 2016).

Alde batetik, itzultzaile automatikoak itzulitako testu baten kalitateak posteditatzeko behar den denboran eragingo du, posteditore batek denbora gehiago eman beharko baitu kalitate txarreko testu bat zuzentzeko eta egokitzeko kalitate onargarria duen bat zuzentzeko eta egokitzeko baino (Garmendia et al.: 2017). Koehn eta Germann-ek (2014) itzultzaile automatikoak posteditzio prozesuan zer nolako eragina duen aztertu zuten. Lau itzultzaile automatikorekin proba egin ostean, itzultzaile automatiko onenaren produktibitate-irabaziak txarrenarenak baino %20 altuagoak zirela ondorioztatu zuten. Horregatik da hain garrantzitsua kalitate oneko itzultzaile automatikoa aukeratzea, posteditzioan eragingo baitu zuzen-zuzenean. Hainbat faktorek baldintzatzen dute itzultzaile automatikoaren kalitatea: hizkuntza-konbinaketak (bi hizkuntzen arteko antzekotasun maila), hizkuntza bateko corpusa eratzeko eskura daukagun testu kopuruak, ikerketari eta tresnen garapenari eskaintzen zaion esfortzuak eta horretara bideratzen diren inbertsioek, eta abar. Faktore horien guztien arabera izango da hobe edo kaskarragoa itzultzaile automatikoko sistema bat. Zenbaitetan, IAren kalitate txarra dela medio, posteditatzen hutsetik itzultzen baino denbora gehiago

eman beharko du posteditoreak, beraz, posteditore profesionalaren esku dago kasuan kasuko hautuak egitea eta produktibitatearen mesedetan IA alde batera uztea, lagundu baino oztopatu egiten badio.

Itzultzaile automatikoaren kalitateaz gain, lehenago aipatu dugun bigarren faktoreak ere eragingo du testu bat posteditatzeko beharko den denboran: bezeroak azken produktuaz eskatzen duen kalitateak (Garmendia et al.: 2017).

Espero den kalitatearen edo emaitzaren arabera, bi postedizio mota bereizten dira:

Postedizio arina edo *light postediting*: Izenak adierazten duen bezala, ahalik eta zuzenketa gutxien egingo ditu posteditoreak, ahalik eta denborarik laburrenean kalitate onargarria lortzeko helburuarekin. Postedizio mota hau, batez ere, barne-erabilerarako testuetan erabiltzen da, ulergarritasuna eta denbora garrantzitsuagoak baitira kalitatea baino. Batez ere ulergarritasuna oztopatzen duten akatsak zuzentzean datza, eta estilo kontuak alde batera uzten dira (Garmendia et al.: 2017).

Postedizio osoa edo *full postediting*: Giza-itzulpen baten kalitatera hurbiltzea du helburu, eta horretarako edizio-maila gorena lortu behar du eta errore mota guztiak eta estiloa zuzendu behar ditu posteditoreak. Kontuan izan behar dugu postedizio osoa produktiboa izan dadin postedizioan emandako denborak hutsetik itzultzen emandakoak baino txikiagoa izan behar duela (Garmendia et al.: 2017).

Jakina da itzultzaileek baliabide informatiko ugari erabiltzen dituztela haien lana errazteko eta azkartzeko. Itzulpengintzak eta teknologiak lotura zuzena dutela graduko lehen mailatik ikusi dugu, eta itzultzaileentzat erabilgarri diren tresna ugari ere ezagutu ditugu: corpusak, thesaurusak, kontsulta-guneak, itzulpen-memoriak, testu-prozesatzaileak... Horietako asko praktikan jartzeko aukera izan dugu graduan zehar, baina bada bat graduan landu ez duguna eta beste hizkuntza-bikote batzuekin lan egiten duten itzulpen-enpresa askok mundu mailan lan-fluxuan txertatzen dutena, emaitza onak ematen dituelako: itzultzaile automatikoa.

Lan honetan aztertu nahi izan duguna da zertan den, gaur egun, euskarazko IA, eta euskarazko postedizioaren erabilera zer nolakoa den. Horretarako, Itzulpengintza eta Interpretazioko graduko hiru eta laugarren mailako ikasleen pertzepzioari jarri diogu arreta bereziki. Izan ere, itzulpen alorrean espezializatzen ari dira, haiek dira epe

laburrean itzultzaile profesional izango direnak, baina oraindik ez dute merkatua barrutik ezagutzen.

ITZULPEN AUTOMATIKOA ETA POSTEDIZIOA EUSKAL HERRIAN: TESTUINGURUA

Espainia mailan eginiko inkesta baten emaitzek diote Espainiako itzulpen-enpresen %45,5ek daukatela postedizioa txertatuta itzulpen-fluxuan. Ikertzaileek harriduraz hartu zuten datua, zifra altuagoa izango zela aurreikusten baitzuten, eta ondorio hauek atera zituzten Espainiako postedizioaren ehuneko baxua azaltzeko: Alde batetik, litekeena da enpresa horientzat lan egiten duten azpikontrataturako itzultzaileek IA erabiltzea enpresari jakinarazi gabe. Beste alde batetik, baliteke enpresek, bezeroei ezer esan gabe, IA erabiltzea. Eta azkenik, gerta liteke itzultzaileek ez jakitea IA erabiltzen ari diren edo ez, gero eta ohikoagoa baita IA OLI tresnetan integratzea eta itzulpen-memoriak eta IAK proposaturako itzulpenak elkarren artean nahastea. Beraz, itxura guztien arabera, errealitatean Espainia mailan IAri ematen zaion erabilera ehunekoak adierazten duena baino handiagoa izango da segur aski. Gainera, denbora kontua dela ere ondorioztatu dute adituek: duela urte batzuk itzulpen memoriak ere zalantzan jartzen ziren, eta ondoren haien erabilera orokortu egin zen. Pentsa liteke gauza bera gertatzea IArekin ere (Torres-Hostench et al.: 2016).

Euskararekin, ordea, ez ditugu hain emaitza baikorrak IA eta postedizioaren inguruan.

Jakina da euskara diglosia egoeran dagoen hizkuntza gutxitu bat dela, eta bere hiztun kopurua ez dela inguratzen dituen hizkuntza nagusien parekoa inondik inora ere. Dena dela, ekimen ugari egin dira hizkuntzaren normalizazioa bultzatzeko, horien artean baita itzulpen automatikoko sistemak garatzeko ere. Hala ere, IA horiekin lortzen den emaitza ez da espero bezain ona, euskararen eta inguruko hizkuntzen egituraketa oso desberdinak direlako batik bat. Hori dela eta, IA itzulpen-merkatuan beste hainbat hizkuntza konbinaziotan adina erabiltzea ez da posible izan.

Hiru dira euskararekin lan egiten duten itzultzaile automatikoak: Google Translate, Matxin eta Eusko Jaurlaritzaren itzultzaile automatikoa.

Inguruan ditugun hizkuntzen kasuan, corpusetan oinarritutako estrategiak erabili izan dira azken hamarkada hauetan, eta alde batera utzi izan dira erregeletan oinarritutakoak. Euskararen kasuan, aldiz, zailtasun handiak izan dituzte corpusetan oinarritutako estrategiek, batetik, teknika estatistikoek emaitza onargarriak lortzeko behar dituzten corpus zabalak ez ditugulako eskuragarri gaur egun euskaraz, eta, bestetik, euskara

hizkuntza eranskaria delako eta “morfologia aberatsa duten hizkuntzetara itzultzean sistema estatistikoak erregeletan oinarritutako sistema komertzialen atzean gelditzen direlako” (Mayor et al.: 2009).

Hori dela eta, adituek aurreikusten zuten bezala, euskarara itzultzeko erregeletan oinarritutako teknikak izan dira nagusi orain arte.

Horren adibide dugu, alde batetik, Matxin itzultzaile automatikoa, EHUko IXA taldeak eta Elhuyarrek elkarlanean garatutakoa. “Gaztelaniatik euskarara itzulpenak automatikoki egiten dituen eta publikoki erabilgarri dagoen itzulpen automatikoko lehen sistema da” (Matxin 2.0), eta testuak, dokumentuak nahiz webguneak itzultzeko aukera eskaintzen du, gaztelaniatik euskararako norantzan.

Googleren itzultzaileak ez bezala, zeinak itzulpen estatistikoa duen oinarri eta hizkuntzen arteko baliokidetzetan oinarritzen den, erregeletan oinarritutako tekniken bidez hizkuntzaren arauak eta egiturak ulertzen ditu Matxinek eta “gaztelaniazko eta euskarazko baliabideekin sortutako itzulpen automatikoko sistema bat” eskaintzen digu, Itziar Cortes Elhuyarreko langilearen hitzetan (Cortes: 2014).

Aipatzekoa da, bestalde, Lucy Software Ibérica-k Eusko Jaurlaritzarentzat garatu zuen itzultzaile automatikoa (www.itzultzailea.euskadi.net). Hau ere erabiltzaile ororen eskura dago sarean¹.

Bestetik, garrantzitsua da azpimarratzea Elhuyar fundazioaren lana, zeinaren helburua euskara zientzia eta teknologiaren arloan garatzea eta indartzea den, eta gizartean duen erabilera hedatzea euskal komunitatea lurralde eta jakintza-alor gehiagotara zabal dadin. Itziar Cortesek elkarrizketa batean jakinarazi zigunez, Elhuyarreko itzulpen-fluxuan oraindik ez dute postedizioa txertatu, baina horren inguruan ikertzen ibili dira azken urteotan. Postedizioarekin zuzenean lotutako Consumer.eus proiektuan lanean jardun dute azken hiru urteotan (2015-2017), Eroski Fundazioarekin eta EHUko IXA ikerketa-taldearekin elkarlanean. IArekin lotutako proiektu experimental bat da Consumer.eus, Matxin ES-EU itzultzaile automatikoa elikaduraren arlora egokitzea helburu duena. Horretarako, gaztelaniazko testuak abiapuntutzat hartuta, Matxinen bidez itzulitako testuak posteditatzeko aukera eman zitzaion erabiltzaileari. Plataforma elikaduraren arlora egokitu zuten, elikagaien arloko hiztegiak eta itzulpen memoriak integratuz.

¹ <http://www.itzultzailea.euskadi.net/traductor/portalExterno/text.do>

Lankidetzan oinarritutako proiektua da, eta erabiltzaile arruntak eskura izan du plataforma, nork bere ekarpena egin ahal izateko. Hala, erabiltzaileak egindako zuzenketetatik ikasten joan da sistema, akats bera ondorengo itzulpenetan ez errepikatzeko eta ondoren postedizio automatikoko sistema bat garatzen laguntzeko.

Formakuntzari dagokionez, aipatzekoa da UEUk 2015ean postedizioaren inguruan antolatutako 47 orduko ikastaroa. Ikastaroaren helburua “teknologia honekiko mesfidantza uxatzeko eta oinarridun iritzi bat eratzeke, ikastaro honetan itzultzaileei itzulpen automatikoa baliagarria zaizuen ala ez” (Itzultzaile izatetik posteditore izatera, UEU) aztertzea izan zen, eta “itzulpen automatikoa erabiltzera, postedizioa taxuz probatzera, egin beharreko aldaketak finkatzen hastera, produktibitatea neurtzera, norberaren itzulpenen kalitatea analizatzera, eta, orokorrean, teknologia honen egoeraz eta baliagarritasunaz hausnartzera” (Itzultzaile izatetik posteditore izatera, UEU) gonbidatu zituzten ikastaroan parte hartu zuten itzultzaileak. Nora Aranberri, Karlos del Olmo eta Ane López izan ziren ikastaroko irakasleak eta ikuspegi teoriko eta praktikotik landu zuten postedizioa, betiere kalitatearen gaineko hausnarketa batekin lagunduta.

Azkenik, ezinbestekoa da itzulpen automatiko neuronala (*Neural Machine Translation*, NMT) aipatzea, azken boladan garatuz doan sistema, oso emaitza onak ematen ari da, baita euskararekin ere. Neurona-sare artifizial (*Artificial Neural Networks*) konplexuetan oinarritutako itzultzaile automatikoei garunaren funtzionamendua imitatuz ikaskuntza sakona (*deep learning*) erabiltzen dute, eta horri esker gai dira beren kabuz ikasteko eta hobetzeko (Garmendia et al.: 2017).

ISEA korporazioko Igor Ellakuriak jakinarazi zigunez “jada estatistikan oinarritutako sistemak baztertuta geratzen ari dira” eta sare neuronalak gailentzen. Eusko Jaurlaritzako ekonomia departamentuak eta FEDER funtsek finantzatuta MODELA proiektua bukatu berri dute arlo honetan eta bertan euskarazko NMT sistema bat garatu dute teknologia neuronaletan oinarrituta beste ikerketa zentro batzuekin batera: Vicomtech, EHUKo IXA Taldea, Elhuyar eta Ametzagaiña eta ISEA. Esan daiteke proiektu horretan garatutako euskarazko itzultzailea gaur egun existitzen den onena dela eta merkaturatze lanetan dabilta orain. Proiektuaren webgunean demo bat dago eskura, gaztelaniatik euskararako norantzan eta alderantziz².

² <http://modela.eus/probatu.html>



Irudia 1 Modela Proiektuaren demoa

Proiektu horiek guztiek martxan badiraute ere, itzultzaile automatikoaren erabilera baxua da oraindik ere euskararen kasuan, lortutako emaitzaren kalitatea ez baita nahi bezain ona oraindik orain. Lierni Garmendiaren, Naroa Lasarteren eta Maialen Pinarren Master Amaierako Lanean (2017) oinarrituko gara ondorengo datuak aurkezteko.

Aipatutako lanaren muina euskarazko itzultzaile automatikoen inguruan dauden iritziak eta erabilpena ezagutzeko inkesta bat izan zen. Guztira 115 pertsonak erantzun zituzten galderak, batzuk itzulpengintzarekin harremana zutenak eta beste batzuk (parte-hartzaileen %36,5a) itzulpengintza sektorearekin inolako harremanik ez zutenak.

Inkestaren arabera, parte-hartzaileen %40k erabili du IA euskararako norantzan. Horietako %71,7ak gutxitan erabili du, %15,2k noizean behin eta %13k sarritan. Euskarako IA inoiz erabili ez dutenetatik (parte-hartzaileen %60a), aipagarria da %66,7ak erabili izan duela IA beste hizkuntza-bikote batzuetan: ingelesera (%44,9), espainolera (%40,6), frantsesera (%18,8) eta abar.

Euskarara IA ez erabiltzeko arrazoen artean, ondorengoak aipatu zituzten inkestan: ez dute euskarara itzultzeko beharrik (%34,8), ez zaie fidagarria iruditzen (%40,6), ez zaie produktiboa iruditzen (%17,4), testuak ulertzeko beste hizkuntza batzuk erabiltzen dituzte (%11,6) edo ez dute itzultzaile automatikorik ezagutzen euskararako norantzan (%7,2).

Hori hala izanik, ondorengo galderak datozkigu burura: Prest ote daude euskararako itzulpen automatikoko sistemak itzultzaileen jardunean laguntzeko? Ba al dute

euskararako edo euskaratiko IA sistemen emaitzek gutxieneko kalitaterik? Baliagarri al zaie itzultzaileei IAK eskainitako laguntza? Ba al dago aurreiritzirik euskarazko IAKo sistemen inguruan? Gaur egun, eraginkorra da postedizioa?

ZERGATIK GRADU AMAIERAKO LANA POSTEDIZIOARI BURUZ?

Postedizioari buruzko azterketa bat egitea izan dugu helburu lan honen bitartez, euskarazko postedizioan sakontzeko. Postedizioari buruz egin izan diren ikerketetan jarduera horri buruzko hainbat alor landu izan da. Orain arte postedizioa, produktibitatearen inguruan ez ezik, beste ikuspegi batzuetatik ere ikertu izan da. Horren adibide dugu Lierni Garmendiak, Naroa Lasartek eta Maialen Pinarrek Master Amaierako Lanean (2017) egindako euskarazko postedizioaren erabileraren gaineko inkesta. IAre eta postedizioaren erabilera aztertu duten ikerketak ere badaude, lehenago aipatu dugun Espainia mailako ikerketa, adibidez, Madrilgo Universidad Europeak egina (Torres-Hostench et al.: 2016).

Postedizioaren produktibitatea aztertu duten lanak ere badira, eta haien artean azpimarratzekoa da Plitt eta Masseloten ikerketa, software lokalizazioaren arloan egindako esperimentua postedizioaren eraginkortasuna neurtzeko eta IARI esker produktibitatea igoko zen edo ez jakiteko (Plitt & Masselot: 2010). Horrez gain, beste ikerketa askotan ere ondorioztatu izan da IARI esker produktibitateak gora egiten duela (Aranberri: 2014).

Lan honetarako guk erabiltzaileen pertzepzioan jarri dugu fokua, eta graduako ikasleek postedizioari buruz duten pertzepzioa ezagutu nahi izan dugu. Orain arte ez da inoiz mota honetako ikerketarik egin graduako ikasleekin, eta alde horretatik berritzailea izan liteke proba hau, orain arte aztertu gabeko kolektibo bat barne hartzen baitu. Kontuan izan behar dugu, aurretik aipatu bezala, proban parte hartu duten itzultzaileek ez dutela aurrez postedizioaren inguruko formakuntzarik jaso, eta gehienek ez dutela inoiz postedizioan jardun.

Graduko ikasleek aurreiritzirik ba ote duten jakin nahi dugu itzultzaile automatikoaren erabilgarritasunaren inguruan, eta aurreiritzi horiek, praktikan, postedizio proba egin ostean mantentzen ote diren. Horrez gain, proba baten bitartez etorkizuneko itzultzaileei beren jardunerako baliagarri izan lekieken tresna batekin lan egiteko aukera ematea eta tresna horren erabilera errealari buruzko gogoeta bultzatzea ere bada gure xedea.

Horretarako, postedizio proba bat diseinatu dugu Itzulpengintza eta Interpretazioko hiru eta laugarren urteko ikasleekin gauzatzeko. Txandaka, posteditatu edo hutsetik itzuli dituzte segmentuak eta bi galdetegi bete dituzte (bat proba aurretik eta bestea ondoren)

IAri ematen dioten erabilera, tresnarekiko duten iritzia eta proba ondorengo gogoetak gurekin partekatzeko. Itzulpen enkargu zehatz honetarako metrika automatikoetan emaitza onena eman duen itzultzaile automatikoaren laguntza baliagarria izan ote den aztertu dugu eta ondorioak atera ditugu, betiere itzultzaileen pertzepzioari erreparatuz.

METODOLOGIA ETA PROBAREN DISEINUA

Atal honetan Gradu Amaierako Laneko ardatz izan den probaren metodologia deskribatuko dugu. Bi izan dira lan honetarako datuak lortzeko erabili ditugun informazio iturriak: parte-hartzaileek erantzun beharreko bi galdetegi itzultzaileen IArekiko pertzepzioa aztertzeko, eta TAUSen DQF Tools plataformarekin diseinatutako proba, produktibitate-tasa erreala neurtuko duena.

Esperimentua Itzulpengintza eta Interpretazioko graduako hiru eta laugarren mailetako ikasleekin egin dugu, haiek izango baitira etorkizunean itzulpengintzan arituko direnak. IA eta postedizioa merkatuan oraindik oso gutxi erabiltzen da euskaraz eta etorkizunean merkatu horretan arituko direnek horiekiko duten pertzepzioa ezagutzea interesgarria da. Gainera, izan duten unibertsitate-mailako formakuntzari esker, kalitate oneko edo onargarriko itzulpena egingo dutela espero da.

Guztira, graduan euskara A hizkuntza duten eta graduan zehar euskarara itzultzen eskarmentua lortu duten 22 ikasle bildu ditugu eta gaztelania-euskara hizkuntza-konbinazioko proba bat diseinatu dugu, euskararekin ohikoena den hizkuntza-bikotea baita.

Probarako hizkuntza-bikoteari dagokionez, euskararekin ohikoena den hizkuntza-konbinazioa gaztelania-euskara dela pentsatu genuen, eta, beraz, gure probarako interesgarriena izango zela. Itzultzaile automatikoaren aukeraketan ere eragin zuen erabaki horrek, itzultzaile automatiko guztiek ez baitute aukera ematen hizkuntza-konbinazio jakin batzuetan itzultzeko. Gaztelaniatik euskararako norantzan itzultzeko, hiru itzultzaile automatikoren artean egin behar izan genuen hautua: Google itzultzailea, Matxin, edo Eusko Jaurlaritzaren itzultzaile automatikoa.

Postedizio proba egiteko TAUSen DQF Tools (*Dynamic Quality Framework*) plataforma erabili genuen.

Enpresek produktibitate probak egin ohi dituzte eskaintzen dituzten zerbitzuen eta produktuen efizientzia neurtzeko, eta gauza bera gertatzen da itzulpen enpresetan ere. 2005az geroztik TAUS plataformak mota horretako probak egiteko aukera eskaintzen du. Itzulpenaren eta hizkuntzen industrian aritu da lanean TAUS, esparru horretan lan egiten duten profesionalei haien lanerako laguntza eskainiz. Horretarako, itzulpen merkatuari eta teknologiei buruzko informazioa eskura jartzen dute, eta baita metrikak

eta askotariko ebaluazio-probak ere (kalitatea edo produktibitatea neurtzen dutenak, esaterako).

Horien artean, DQF Tools erreminta dago sarean eskura TAUSen izena emanda dagoen erabiltzaile ororentzat. IAren eta postedizioaren erabilerari buruzko ebaluazio-sistema eskaintzen du plataforma horrek, bai enpresei eta bai hezkuntza-erakundeei. Itzulpen-enpresek plataforma horri esker egiten dituzte, besteak beste, eskaintzen dituzten zerbitzuen efizientzia neurtzeko produktibitate probak.

Hauek izan dira esperimentuaren faseak:

Prestakuntza

Lehenik, beste esperimentu batzuen erreferentziak jaso genituen, izan ere, hutsetik postedizio proba bat sortzeko garrantzitsua da aurretik egin direnak aztertzea, haien helburuak eta emaitzak ikusteko. Horien artean Plitt eta Masselotenarena interesgarria iruditu zitzaigun. Bertan, IAren produktibitatea aztertu zuten, eta horretarako hamabi itzultzaile bildu. Lehenik, IAren laguntzarik gabe itzuli zuten testu bat ingelesetik frantsesera, italierara, alemanera eta espainolera (hizkuntza bakoitzerako hiru itzultzaile). Bigarrenik, IAk itzulitako testu bat posteditatu zuten hizkuntza-konbinazio horietan (Plitt & Masselot: 2010).

Guk euskara eta postedizioa uztartuko zituen azterketa bat egin nahi genuen, eta horretarako, itzulpengintzan doktorea eta postedizioan aditua den Nora Aranberriekin hitz egitera joan ginen Leioako Irakasle Eskolara. Haren laguntzarekin, esperimentuaren nondik norakoak zehaztu genituen bilera horretan, eta gai hauek aztertzen dituzten ikerketetan eman ohi diren urratsetarako hurbilpen bat eskaini zigun. Urrats nagusien artean, probarako ikuspegi bat finkatu beharra aipatu zigun (produktibitatea, kalitatea, erabiltzaileen gaitasuna lantzea...), horren arabera diseinatzeko proba. Behin probaren nolakotasunak finkatuta, probarako erabiliko genuen testua bilatzeari ekin genion.

Probari enkargu erreal baten itxura eman nahi genion, eta, horretarako, testu erreal bat aukeratu genuen. Hala ere, terminologia espezializaturik gabeko testu bat behar genuen. Izan ere, testu espezializatu bat itzuli nahi izango bagenu, alor horretako terminologiara moldatu beharko genuke itzultzaile automatikoa ataza hori ondo egiteko. Bestalde, testuari inolako egokitzapenik ez egitea erabaki genuen, eta segmentu solteak barik testu

oso bat erabiltzea. Modu horretan, itzultzaileak egoera erreal batean kokatu genituen, itzulpena testu jakin horretako testuingurua kontuan izanda egin behar izan baitzuten.

Beraz, ondorengo faktoreak izan genituen kontuan parte-hartzaileek proban itzuli beharreko testua aukeratzean. Lehenik, testu espezializatua ez izatea. Bigarrenik, testua oso luzea ez izatea, proba asko luza ez zedin. Azkenik, esaldi oso luzedun testuak baztertu genituen, itzultzaile automatikoaren proposamenak okertzen baitituzte halakoek.

Ezaugarri horiek guztiak kontuan izanik, haurrei zuzenduta dagoen *¡La leche!* izeneko aldizkaritik ateratako gaztelaniazko testu labur bat aukeratu genuen (264 hitzekoa): “La carrera de los descubridores”³.

Lehenago aipatu dugun bezala, ezinbestekoa da kalitate oneko itzultzaile automatikoa aukeratzea, horrek zuzenean eragingo baitu postedizio lanaren eraginkortasunean. Horregatik, gure probarako ere, behin testua aukeratuta hiru itzultzaile automatikoren itzulpenak bildu genituen (Googleren itzultzaile automatikoa, Eusko Jaurlaritzarena, eta Matxin) emaitza onenak hiruretatik zeinek ematen zituen ikusteko, eta, horren arabera, probarako egokiena aukeratzeko.

Metrika automatikoak erabili genituen, automatikoki itzulpen baten kalitatea ebaluatzen duten plataformak, alegia, hiruren artean itzulpen automatiko onena zein zen erabakitzeko. Antzeko ikerketetan ere plataforma horiek erabili ohi dira itzulpen baten kalitatea ebaluatzeko algoritmotzat. Guztion eskura ez daudenez, datu horiek lortzeko gaian aditua den eta plataforma horietara sarbidea duen Nora Aranberrirekin kontaktuan jarri ginen. Hark pasa zuen testuen kalitate kontrola eta emaitzak jakinarazi zizkigun.

IAk emandako itzulpena giza-itzulpen batekin erkatzen dute metrika automatiko horiek. Beraz, testua guk geuk itzuli genuen itzulpenak metrika automatikoetatik pasa baino lehen, gure itzulpena erreferentziazko itzulpen gisa erabiltzeko.

Aipatutako metrika automatikoen helburua itzulpenak modu objektibo, azkar eta merkean ebaluatzea da. Beste batzuen artean, BLEU⁴ eta TER dira ezagunenetariko bi.

³ Hirugarren eranskinean aurkituko duzue testua.

⁴ Plitt eta Masseloten ikerketan erabili zuten, adibidez, BLEU metrika automatikoa

BLEUk (*Bilingual Evaluation Understudy*) itzulpenen egokitasuna neurtzen du, eta, horrenbestez, zenbat eta altuagoa izan BLEU metrika automatikoaren emaitza, orduan eta altuagoa izango da itzulpenaren kalitatea. TERek (*Translation Error Rate*), aldiz, itzulpen automatikoaren errore maila neurtzen du. Kasu honetan, beraz, zenbat eta baxuagoa izan TER metrika automatikoaren emaitza, orduan eta altuagoa izango da itzulpenaren kalitatea (Aranberri, 2014).

Metrika automatiko gehiago ere badauden arren, guk aipatutako biei erreparatu diegu batez ere, horiek baitira gurea bezalako ikerketetan gehien erabiltzen eta aipatzen direnak. Ondorengo taulan metriken emaitzak daude adierazita, Nora Aranberrik eman zizkigun formatuan. Garbi ikus daiteke Eusko Jaurlaritzaren itzultzaile automatikoa (Lucy) dela egokiena, hari dagozkio eta BLEU emaitzarik altuena eta TER emaitzarik baxuena.

	MBLEU	BLEU	NIST	METEOR	TER	WER	PER
GOOGLE	5.50 (BP=93)	11.36	2.99	29.38	72.34	73.41	60.75
MATXIN	5.70 (BP=100)	11.62	2.95	33.19	74.04	73.83	60.75
LUCY	14.27 (BP=100)	20.54	3.77	48.55	60.00	62.02	48.94

Irudia 2 Metrika automatikoen emaitzak

Behin jatorrizko testua eta itzultzaile automatikoaren emaitza esku artean izanda, proba nagusia diseinatu genuen. Horretarako, bi testuak esaldiz esaldi parekatu behar genituen, eta itzultzaileentzat proba egiteko euskarria erabaki behar genuen. Aukera ugari zeuden proba gauzatzeko: alde batetik, testu-prozesatzaile bat edo kalkulu orri bat erabiltzea, itzultzaileek aldaketak programa horietakoren batean egin zitzaten. Bestetik, OLI tresnaren bat ere erabil zitekeen, itzulpen automatikoa bertan txertatuta. Azkenik, TAUS plataformaren alde egin genuen, Elizabete Manterola lan honetako tutoreak eta Nora Aranberrik aurretik esperientzia baitzuten plataforma hori erabiltzen.

Esan bezala, TAUS plataformaren bitartez proba bera eta proba aurreko demoa sortu genituen. Hori egiteko, lehenik eta behin izena eman genuen TAUSen webgunean. Kontua sortu eta gero, proba sortzeari ekin genion.

Welcome

The TAUS Dynamic Quality Framework tools enable members to measure [MT productivity](#), [rank MT engines](#), evaluate [adequacy](#), [fluency](#) and/or undertake [error typology](#) review. You can start an evaluation project using your own evaluators or linguists, or you can choose to outsource the evaluation tasks to professional linguistic consultants.

[Getting Started](#)

[Create Project](#)

[Outsource Evaluation](#)

Irudia 3 TAUSen proiektua sortzeko

Hiru proba mota eskaintzen dizkigu TAUSen webguneak, 79 hizkuntza desberdinetan, horien artean euskara. Hautatutako probaren arabera, erabiltzaileek itzultzaile automatikoak itzulitako segmentuak posteditatu ditzakete produktibitatea neurtzeko (produktibitate proba), hizkuntza-bikote jakin batean lan egiten duten hiru itzultzaile automatiko desberdinen jarduna ebaluatu segmentuz segmentu (konparaketa proba), edo itzulpen baten egokitasuna eta jariorasuna ebaluatu (kalitate ebaluazioa).

Guk produktibitate proba aukeratu genuen gure ikerketarako, txandakatuz segmentuak posteditatzeko eta hutsetik itzultzeko aukera ematen duelako. Bi jardueren artean saltoka ibiltzeak itzultzaileei biak alderatu eta horien inguruan hausnartzeko aukera ematen die, eta produktibitatea kalkulatzeko ere laguntzen digu, jarduera bakoitza egiten tardatzen duten denbora neurtzen baitu plataformak.

Project Creation

[Start Over](#)

Project Name

Project Type

Productivity [\(Info\)](#)

Post-editing Quality

Good Enough [\(Info\)](#)

Process

- ☐ Human [\(Info\)](#)
- ☐ MT+PE [\(Info\)](#)
- ☐ TM+Human [\(Info\)](#)
- ☒ MT+PE+Human [\(Info\)](#)
- ☐ MT+PE+TM+Human [\(Info\)](#)

Irudia 4 TAUSen proiektua sortzeko

Proba amaitu eta gero, programak txosten batean biltzen ditu parte-hartzaile bakoitzari dagozkion datuak: posteditatzen eta hutsetik itzultzen emandako denbora, produktibitate-ratioa, egindako itzulpenak...

Behin jatorrizko testua eta itzultzaile automatikoak eginiko itzulpena programan parekatuta, parte-hartzaileei aurretiaz mezu elektroniko bat bidali genien probarako estekarekin. Proba online egin behar da, eta esteka hori proba egunean bertan ireki zuten guztiek batera.

Proba aurreko prestaketarekin amaitzeko, bi galdetegi prestatu genituen. Lehen galdetegiak zazpi galdera ditu eta parte-hartzaileek IAren inguruan daukaten iritzia eta ematen dioten erabilera ezagutzea du helburu. Bigarren galdetegiak bost galdera ditu, eta proban edo ataza jakin honetan IA erabilgarria izan zaien jakitea du helburu, hau da, probaren ondorengo gogoetak eta parte-hartzaileen pertzepzioa ezagutzea.

Galdetegiko galderak diseinatzeko Nora Aranberrirekin egindako bileran aipatutako puntu nagusiak hartu genituen kontuan. Gai hau sarritan landu izan duenez, zenbait gidalerro eta aholku eman zizkigun gure lana ondo bideratzeko, eta bileran aipatutakoa oinarri hartuta diseinatu genituen bai galdetegiak eta baita proba bera ere.

Parte-hartzaileak biltzea

Esperimentua egiteko bi saio presentzial antolatu genituen parte-hartzaileekin. Eguna eta ordua zehazteko Doodle galdetegi bat prestatu genuen, eta A hizkuntza euskara daukaten Itzulpengintza eta Interpretazioko graduako hiru eta laugarren mailetakako ikasleen artean zabaldu. Bertan, ikaskide bakoitzak bere eskuragarritasuna jakinarazi zigun. Hori kontuan hartuta, lehen saioa antolatu genuen eta laugarren mailako 10 ikaskidek hartu zuten parte proban.

Ondoren, graduako hirugarren mailako ikasleekin ere proba egiteko, hirugarren mailako irakasle den Naroa Zubillagarekin kontaktuan jarri ginen eta eskola-orduan bertan proba egiteko aukera eman zigun. Bigarren saio horretan beste 12 parte-hartzaileekin errepikatu genuen proba.

Guztira, beraz, Itzulpengintza eta Interpretazioko graduako hiru eta laugarren mailetakoko 22 ikasleko lagina hartu dugu proba egiteko. Aipatzekoa da, bestalde, proba anonimoa izan zela eta anonimotasuna bermatzeko parte-hartzaile bakoitzari zenbaki bat esleitu geniola ausaz.

Esperimentua edo ataza

Proba, guztira, lau fasetan banatu genuen.

Lehenik, datuak lortzeko, proba aurreko galdetegi bat erantzun zuten parte-hartzaileek. Proba aurreko galdetegiaren⁵ helburua parte-hartzaileek IAren zeukaten esperientzia edo ematen zioten erabilera ezagutzea zen, jakitea IAren aurreiritzirik bazuten edo ez, eta zer aurreikusten zuten parte-hartzaileek postedizio atazarekiko. Bestela esanda, haien ustez IAren laguntza baliagarri izango zitzaien edo ez ataza honetan edo testu bat itzultzean.

Bigarrenik, probako postedizio ariketa txiki bat egin genuen proba nagusia baino lehen, kontuan izanda parte-hartzaileak ez zeudela probarako erabili dugun tresnara edo plataformara ohituta. Programara ohitzeko ez ezik, probarako baliagarri izango zitzaizkigun gidalerro guztiak bisualki azaltzeko ere balio izan zigun demoak. Modu horretan, fidagarriagoak izango dira probako emaitzak, ezjakintasunak edo zalantzek

⁵ Lehenengo eranskinean dago ikusgai proba aurreko galdetegia.

eraginda gorabeherak izateko eta gorabehera horiek emaitzetan eragiteko arriskua txikiagoa izan baita.

Probari ekin aurretik antolatutako demo txiki horretan, guk aurrez prestatutako lau esaldi edo segmentu posteditatu edo hutsetik itzuli zituzten. Proban bertan ahalik eta zalantzarik gutxien izateko, plataformaren inguruko argibide edo gidalerro batzuk eman genizkien itzultzaileei, modu horretan, arreta berezia eskaini zien ondorengo proban kontuan izan beharreko alderdi teknikoei zegozkien berezitasunei. Hala eta guztiz ere, baliteke oraindik proban gorabehera batzuk egon izana, ohitura faltak eraginda.

Hirugarrenik, proba nagusia gauzatu genuen, TAUS plataformaz baliatuz. Parte-hartzaileen egitekoa segmentuak posteditatu edo hutsetik itzultzea izan zen, bata eta bestea txandakatuz. Modu horretan, bi jardueren artean saltoka aritu ziren parte-hartzaileak eta bataren eta bestearen inguruko gogoeta bultzatu zen⁶. Guztira hamalau segmentu zituen landu beharreko testuak.

Esan bezala, segmentu batzuetan itzulpen automatikoaren laguntza izan zuten itzultzaileek eta beste batzuetan hutsetik hasi behar izan zuten itzultzen. Bi kasuetan, momentuan itzuli beharreko segmentuaz gain, aurretik landutakoa eta ondoren landu beharrekoa ere ikusgai uzten du programak, itzultzaileak testuingurua ezagutzeko. Bestalde, kontuan izan behar da behin hurrengo segmentura pasata plataformak ez duela aurreko segmentuetara atzera egiten uzten. Hurrengo orrialdean ikusgai daude probako bi irudi: lehenengoa posteditatu beharreko segmentu baten adibide bat da (posteditatzeko interfaza), eta bigarrena hutsetik itzuli beharreko batena (itzultzeko interfaza).

⁶ TAUS plataformak berak eskaintzen du aukera halako probak egiteko, dagoeneko definitutako probak baitira.

Home

postedizio proba irati (Productivity)

Information

Required Level of Quality: [Good Enough](#)
Content Type: Knowledge Base
Filename: postedizio_proba_2.tmx
Segment: 3 of 14

Source: Spanish (Spain)

Previous EE.UU., Francia & Ingalaterra, 1837-1843

Current Cuando por fin se demostró que la Antártida era un continente, comenzó la ávida carrera por conquistarlo.

Next El rey francés Luis Felipe I ordenó encontrar y reclamar el polo magnético a Dumont d'Urville.

Target: Basque

Previous AEB, Frantzia eta Ingalaterra, 1837-1843

Current Azkenik Antartika kontinente bat zela frogatu zenean, karrera egarria hasi zuen hura konkistatzeagatik.

PAUSE

NEXT

Or Press Enter

Please write to us with any questions at dqf@taus.net.
Copyright TAUS 2014

Irudia 5 Posteditatzeko interfaza

Home

postedizio proba irati (Productivity)

Information

Required Level of Quality: [Good Enough](#)
Content Type: Knowledge Base
Filename: postedizio_proba_2.tmx
Segment: 4 of 14

Source: Spanish (Spain)

Previous Cuando por fin se demostró que la Antártida era un continente, comenzó la ávida carrera por conquistarlo.

Current El rey francés Luis Felipe I ordenó encontrar y reclamar el polo magnético a Dumont d'Urville.

Next A pesar de que el capitán soñaba con aguas más cálidas, se vio obligado a comandar una expedición en 1837.

Target: Basque

Previous Azkenik Antartika kontinente bat zela frogatu zenean, karrera egarria hasi zuen hura konkistatzeagatik.

Current

PAUSE

NEXT

Or Press Enter

Please write to us with any questions at dqf@taus.net.
Copyright TAUS 2014

Irudia 6 Itzultzeko interfaza

TAUS programari esker, segmentu bakoitza itzultzeko behar izan zuten denbora ere neurtu genuen, parte-hartzaileek izan zuten pertzepzioa errealitatearekin bat zetorren ala ez ikusteko.

Azkenik, proba ondorengo galdetegia⁷ bete zuten parte-hartzaileek. Izan berri zuten esperientziaren inguruko galderak erantzun zituzten proba ondorengo galdetegian. Besteak beste, itzulpen-lan jakin honetan erabilgarri izan zitzaien IAren laguntza edo ez, edo ataza honetan itzultzaile automatikoari esker denbora aurreztu zuten edo ez.

Guztira, ordu eta erdi inguruko iraupena izan zuen probak saio bakoitzean.

Proba egin ostean jasotako emaitzak aztertu ditugu eta ondorengo atalean aurkeztuko ditugu, xehe-xehe.

⁷ Bigarren eranskinean dago ikusgai proba ondorengo galdetegia.

IKERKETAREN MUGAK

Adituen arabera oso zaila da zehatz adieraztea gurea bezalako probetan itzultzaileen eguneroko errealitatea zein den, oso zaila baita itzulpen-prozesuan eragiten duten faktore guztiak kontuan izatea (Garmendia et al.: 2017).

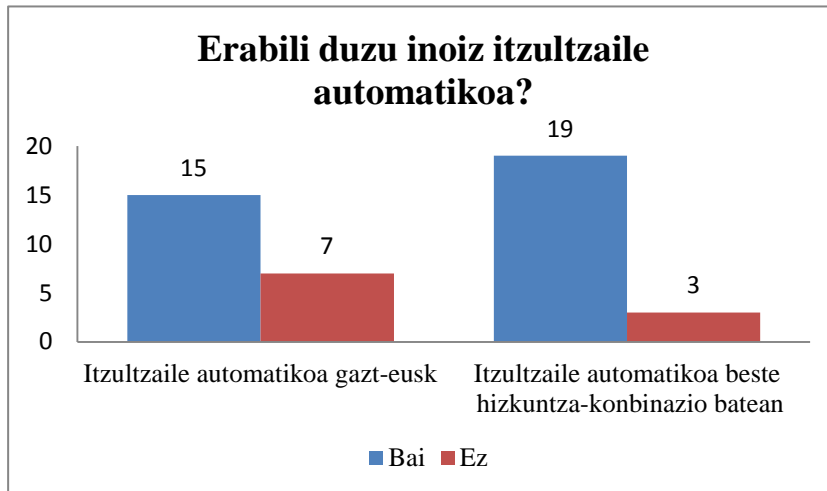
Alde batetik, proban parte hartu zuten ikasleek ez dute graduan zehar postedizioaren inguruan inolako formakuntzarik jaso eta horrek eragina izan lezake emaitzetan. Proba baino lehen postedizioari buruzko inolako gidalerro teorikorik ez eskaintzea erabaki genuen, itzultzaileen jarduna eta probaren emaitzak ez mugatzeko eta itzultzaile bakoitzak libreki itzuli eta itzulpen automatikoa nahierara moldatzeko. Hala ere, jakitun gara itzultzaileek enkargupean lan egiten dutela eta postedizioaren inguruko inolako gidalerrorik gabe lan egiteak errealitatetik urrun dezakeela proba.

Bestalde, probarako erabili genuen TAUS plataformaren mugei dagokienean, itzuli beharreko segmentua, aurretik itzultitako segmentua eta hurrengo itzuli beharrekoa bakarrik jartzen ditu ikusgai plataformak, eta behin segmentua onartuta ez du atzera egiteko aukerarik ematen. Hiru parte-hartzailek jakinarazi ziguten faltan bota zutela testu osoa irakurtzeko aukera.

Azkenik, galdetegian aukera anitzeko galderak jarrita erantzunak baldintzatzen genituela iruditzen zitzaigun eta hori saihestu nahi izan genuen. Parte-hartzaileen iritzia hobeto ezagutze aldera, galdera ireki asko egin genituen galdetegietan, parte-hartzaileek garatu beharrekoak. Ateratako emaitza orokorrak eskaintzeko ikasleek emandako erantzunak multzokatu ditugu, eta kasu batzuetan zailtasunak edo zalantzak izan ditugu antzeko erantzunak multzokatzeari, askotariko erantzunak jaso baititugu.

EMAITZEN AZTERKETA: LEHEN GALDETEGIA, ITZULTZAILE AUTOMATIKOAREKIKO (AURRE)IRITZIAK ETA PARTE-HARTZAILEEN OHITURAK EZAGUTZEA

1. Inoiz IA erabili duen parte-hartzaile kopurua

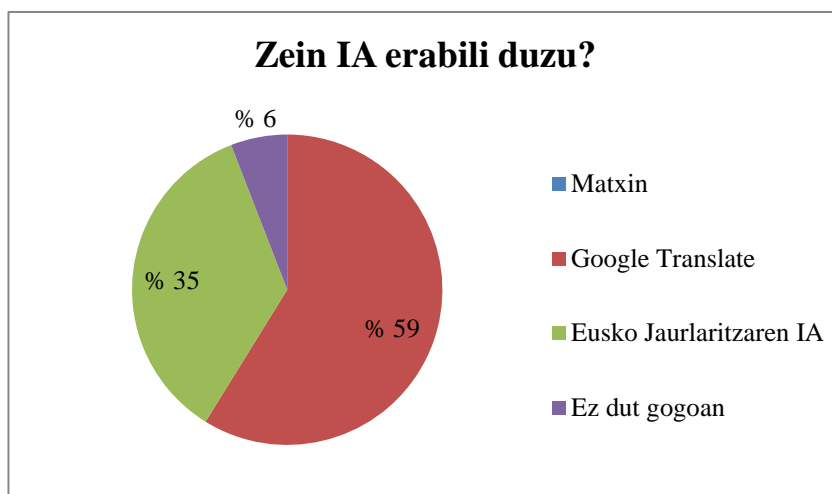


Grafikoa 1 Erabili duzu inoiz itzultzaile automatikoa?

Grafikoan ikus dezakegun bezala, baxuagoa da IA gaztelania-euskara hizkuntza-bikotean erabili duten parte-hartzaile kopurua beste hizkuntza-konbinazioekin alderatuz. Galdetegia erantzun duten 22 ikasletatik 15 dira noizbait gaztelania-euskara hizkuntza-bikotean IA erabili dutenak; beste hizkuntza-bikote batzuetan, aldiz, 19 ikaslek erabili dute noizbait IA.

Aipatzekoa da, bestalde, gaztelania-euskara ez diren hizkuntza-bikoteetan IA erabili ez duten hiru ikasleek aitortu dutela noizbait IA erabili izana gaztelania-euskara hizkuntza-bikotean.

1.1. Zein IA erabili izan dute gehien parte-hartzaileek ES-EU konbinazioan?

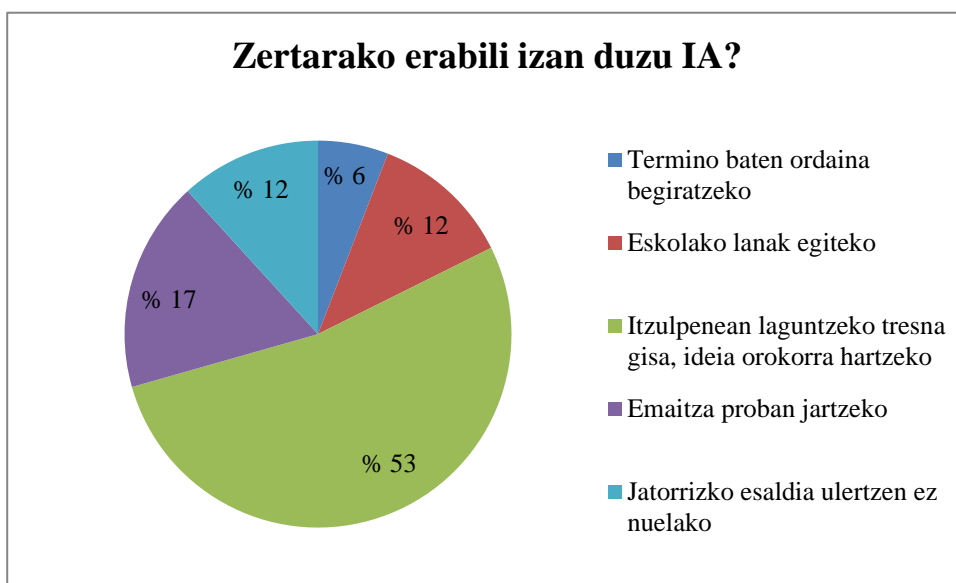


Grafikoa 2 Zein IA erabili duzu?

Gaztelania-euskara bikotean IA erabili izan dutenen artean gehien erabili izan den itzultzaile automatikoa Google Translate da (10 pertsona). Bigarren postuan, Eusko Jaurlaritzaren IA izan da parte-hartzaileek gehien erabili dutena (6 pertsona). Azkenik, aipatzekoa da parte-hartzaile batek adierazi duela ez duela gogoan zein itzultzaile automatiko erabili duen. Segur aski, guztiak multzo berean sartuko ditu eta ez du IA bataren eta bestearen artean inolako alderik ikusiko.

Bestalde, harrigarria dirudien arren, inork ez du Matxin itzultzaile automatikoa erabili. Azken urteotan garapen handia izan duen proiektua izan den arren, badirudi oraindik ez dela ezaguna Itzulpengintza eta Interpretazioko graduko ikasleen artean.

1.2. Zer erabilera eman diote parte-hartzaileek IAri?

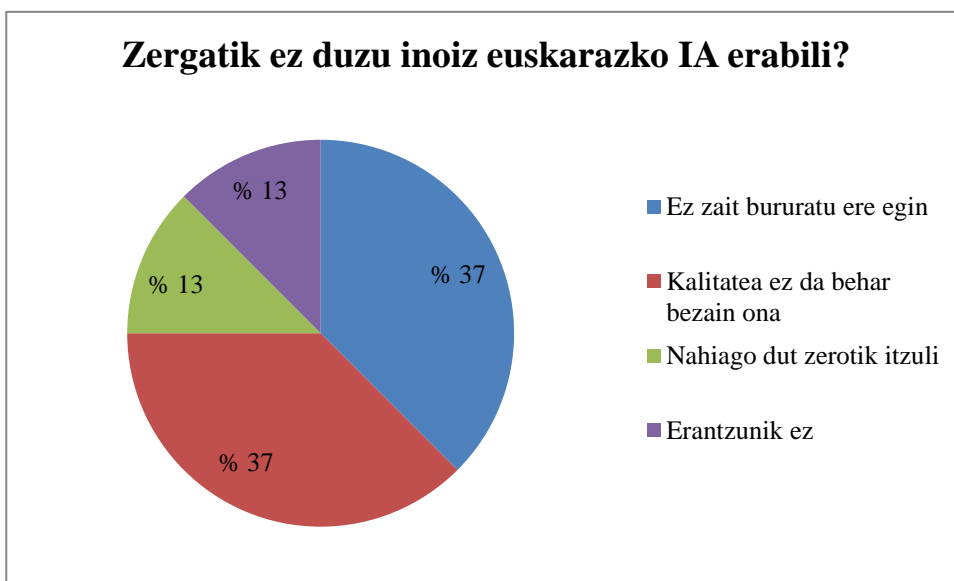


Grafikoa 3 Zertarako erabili izan duzu IA?

IA noizbait erabili izan dutenetatik, erdiak baino gehiagok erabili izan du itzulpen-prozesuan laguntzeko tresna gisa, hau da, jatorrizko testuak diotenaren ideia orokorra ulertzeko. Bestalde, hiru lagunek aipatu dute emitza proban jartzeko erabili dutela IA, jakin-min hutsak bultzatuta, proposatzen duen itzulpena onargarria edo kaskarra den ikusteko. Jarraian, bi pertsonak aipatu dute inoiz eskolako lanekin laguntzeko erabili izan dutela IA. Azkenik, pertsona batek adierazi du termino baten ordaina begiratzeko erabili izan duela IA.

Ondoriozta dezakegu, beraz, aurreko galderari baietz erantzun dioten parte-hartzaile guztiek ez dutela IA helburu profesionalerako (itzulpen-jarduerarako) erabili. Hain zuzen ere, 15 lagunetik 10 dira IA itzulpen-prozesuan lagungarri izango zaien tresna gisa erabili dutenak eta beste 5ek jarduera akademikoari edo profesionalari loturiko itzulpengintzaz kanpo dauden funtzioak betetzeko erabili izan dute (itzulpengintza-jardueratik kanpoko helburu akademiko nahiz pertsonaletarako).

1.3. ES-EU bikotean IA ez erabiltzeko arrazoiak

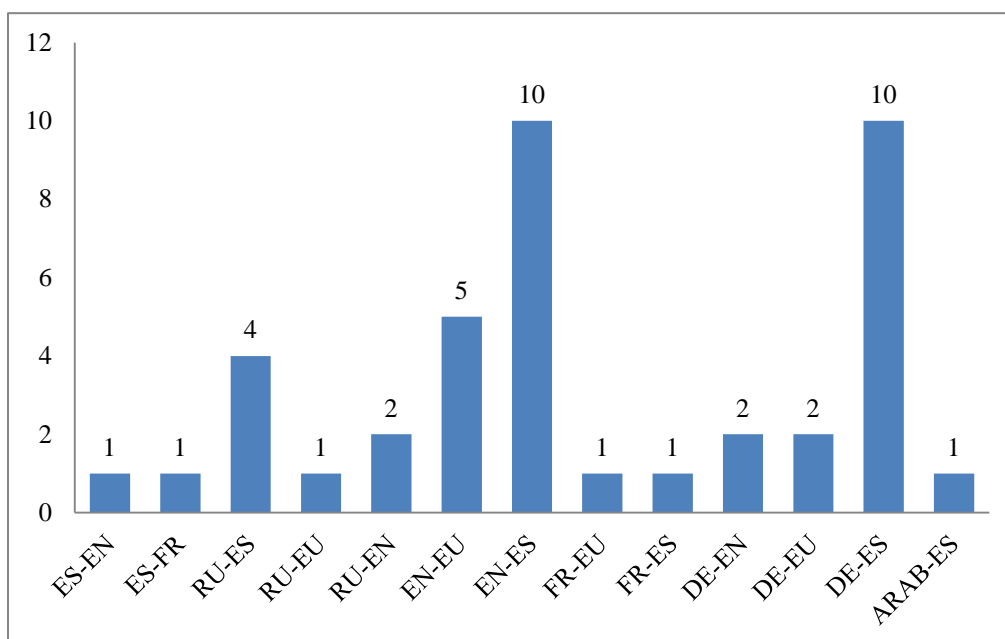


Grafikoa 4 Zergatik ez duzu inoiz euskarazko IA erabili?

ES-EU bikotean IA inoiz erabili ez dutenen artean, honako hauek dira gehien errepikatu diren arrazoiak: 3 pertsonak aipatu dute bururatu ere ez zaiela egin baliabide hori erabiltzea. Beste 3 pertsonak aipatu dute IAk eskaintzen duen kalitatea txarra dela, eta, azkenik, pertsona batek aipatu du nahiago duela zerotik itzuli, kasu honetan ere, segur aski, emaitzak txarrak izango direlakoan.

Emaitza horietatik ondoriozta dezakegu Itzulpengintza eta Interpretazioko graduan oraindik IAren erabilera normalizatu gabe dagoela, eta, ES-EU hizkuntza-bikotean oraindik ikasle askok ez dutela IA baliabide edo tresna lagungarritzat edo erabilgarritzat jotzen.

1.4. IA beste hizkuntza-bikote batzuetan



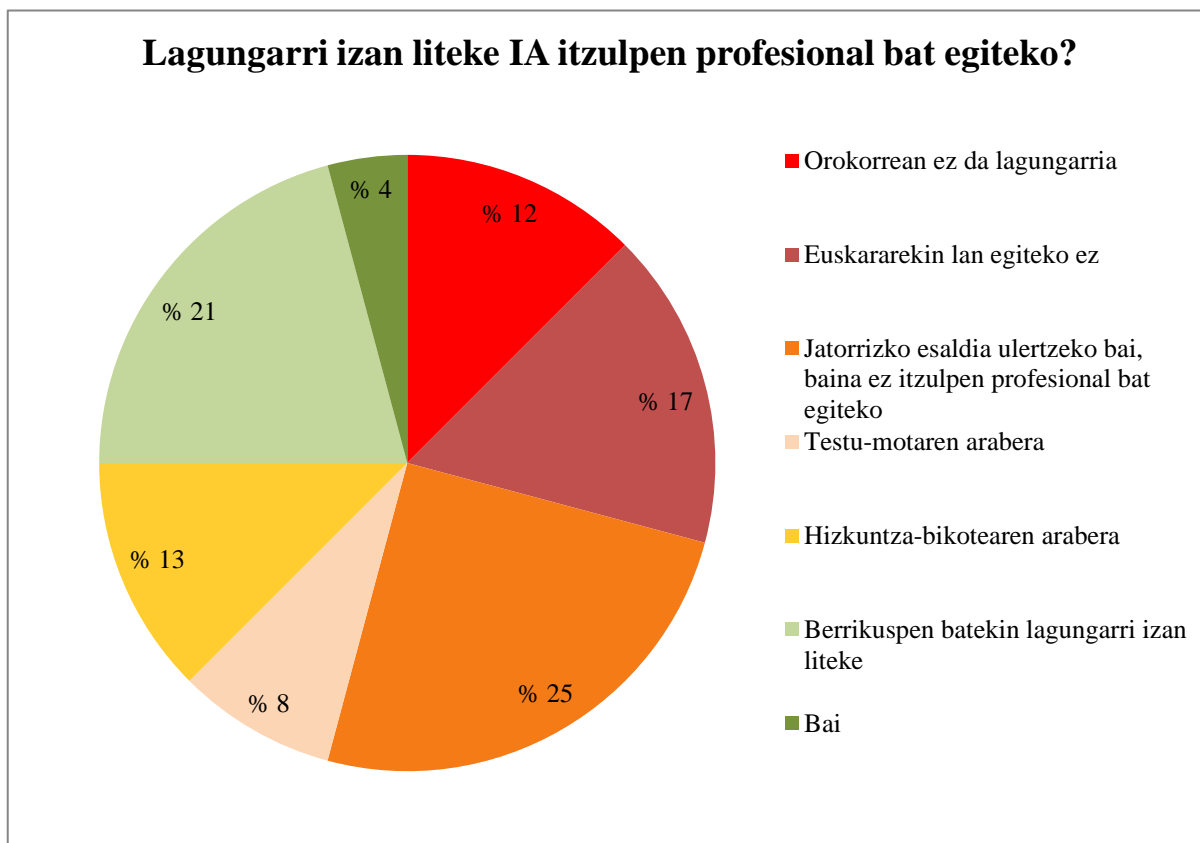
Grafikoa 5 IA beste hizkuntza-bikote batzuetan

Grafiko honek garbi adierazten du zein hizkuntza-bikotetan erabiltzen duten gehien IA proba honetan parte hartu duten ikasleek. 10 ikaslek erabili izan dute IA EN-ES eta DE-ES hizkuntza-bikoteetan. Aipatu behar da, bestalde, ez ditugula kontuan izan proban parte hartu duten ikasleen B eta C hizkuntzak. Hori dela eta, ezin ditugu objektiboki erkatu alemanaren, frantsesaren edo errusieraren datuak, ez baitakigu parte-hartzaileetako zenbatek duten alemana C hizkuntza, zenbatek frantsesa eta zenbatek errusiera.

Alde batetik, ondorio garbi bat atera dezakegu: hizkuntza-bikote hegemonikoetan altuagoa da IArenerabilerak. Horren adierazgarri garbia dugu EN-ES edo DE-ES konbinazioei dagozkien zifra altuak. Izan ere, proban parte hartu duten ikasleek euskaraz ikasten dute gradua, eta ikasketa-planaren barruan ez dago EN-ES hizkuntza-konbinaketa lantzen duen irakasgairik, EN-EU baizik. Hala eta guztiz ere, altuagoa da IA ingelesetik gaztelanarako norabidean erabili duten ikasle kopurua, segur aski emaitza hobea izango zela aurreikusita joko baitzuten zuzenean hizkuntza-bikote horretara bilaketa egiteko.

Euskarari dagokionean, ES-EU hizkuntza konbinazioan ez ezik, oro har euskara barne hartzen duten hizkuntza-konbinazioak gutxienekoak dira ikasleek IArenerabiltzen dituzten erabileran.

2. Lagungarri izan liteke IA itzulpen profesional bat egiteko?



Grafikoa 6 Lagungarri izan liteke IA itzulpen profesional bat egiteko?

Galdetegian neurtu nahi genuen beste faktore bat IArekiko aurreiritzirik ba ote zegoen zen. Horretarako, proba aurreko galdetegian galdera hau egin genien parte-hartzaileei: Lagungarri izan liteke IA, zure ustez, itzulpen profesional bat egiteko?

Grafikoan ikus dezakegun bezala, iritzi kontrajarriak daude, baina begi-bistakoa da baiezko borobila eman duten ikasle kopurua oso txikia dela (ikasle bakarra). Eraitza baikorrak eman dituztenak, baiezko borobila (lagun bat) nahiz berrikuspen batekin erabilgarria izan litekeela (5 lagun) adierazi dutenak, parte-hartzaileen %25a besterik ez dira guztira.

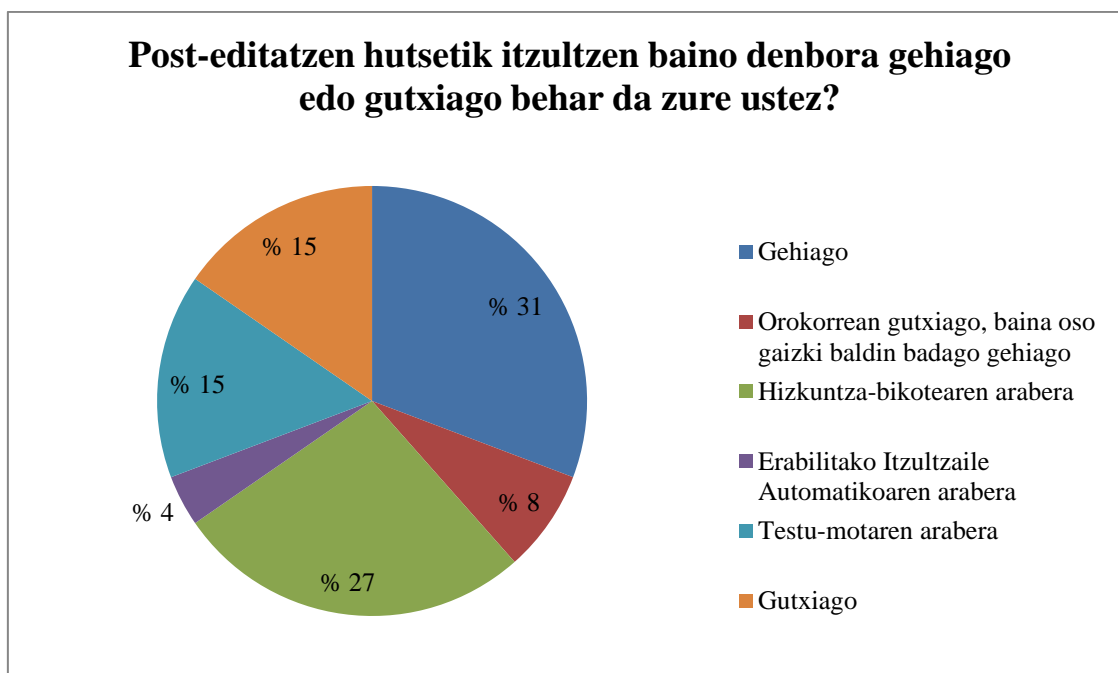
Gehiengoak mesfidantza adierazi du, eta baldintzak jarri ditu: euskararekin lan egiteko balio ez duela diote 4 lagunek, ulermen mailan lagun dezakeela baina itzulpen profesional bat egiteko ez duela balio diote 6 lagunek, IArek erabilgarritasuna testu-motak eta hizkuntza-konbinaketak zehaztuko dutela diote beste 5 lagunek, eta beste 5 lagunek diote berrikuspen batekin erabilgarria izan litekeela, zuzenketa eta

egokitzapenekin, alegia. Ezezeko borobila eman duenik ere bada, 3 lagun, hain zuzen ere.

Testu-motari dagokionez, itzultzaile askok aipatu dute literatur-testuak itzultzeko ez dela lagungarria, baina administrazio-testuak itzultzeko, aldiz, erabilgarria izan daitekeela, formulak askotan errepikatzen baitira eta askotan formula horiek IAen corpusetan gordeta baitaude. Testu-mota batzuetan IAko tresnek ez dutela fidagarritasunik ere aipatu dute batzuek.

Itxura guztien arabera, nahiko orokorra da IArekiko mesfidantza, IA noizbait erabili dutenen kasuan eta baita tresna inoiz erabili ez dutenen kasuan ere.

3. Postedizioa: produktibitatearen gainera iritzia



Grafikoa 7 Posteditatzen hutsetik itzultzen baino denbora gehiago edo gutxiago behar da zure ustez?

Proban parte hartu dutenen artean, gehiengoak (8 pertsonak) aurreikusi du orokorrean posteditatzen denbora gehiago beharko dela testu bat hutsetik itzultzen hasita baino. Esaldiz esaldi jatorrizko testua eta itzulpen automatikoa alderatu beharrari hutsetik hasteari baino denbora gehiago eskaini behar zaiola diote, eta baita zuzenketak eta egokitzapenak egiteari ere. Aipagarria iruditzen zaigu, gainera, postedizioak esfortzu kognitiboa arindu beharrean astunago bihurtzen duela iruditzen zaiola itzultzaile bati.

Beste batzuen ustez beste hainbat faktoreren baitan egongo da denbora gehiago edo gutxiago behar izatea (hizkuntza-bikotea, testu-mota eta testuaren konplexutasun-maila edo itzulpen automatikoaren kalitatea).

Hizkuntza-bikoteari dagokionez, bi hizkuntza hegemonikoren artean ari bagara posteditatzen denbora gutxiago emango dugula uste dute askok, baina euskararen kasuan denbora gehiago beharko dugulakoan daude, urruneko hizkuntzak direlako eta, batez ere, tresnak oraindik guztiz garatzeke daudelako. Askok halaxe esplizituki adierazi ez badute ere, euskararekin lan egiteko orduan itzultzaile automatikoekiko mesfidantzaren adierazgarri izan liteke hori ere. Aurreko galderan ere ikusi dugu asko errepikatzen den ideia dela euskararen kasuan IA erabilgarria ez izatearena.

Hizkuntza-bikotea aipatzen duten beste askok sorburu-hizkuntzaren ezagutza-mailari egiten diote erreferentzia. Izan ere, jatorrizko hizkuntza ondo menperatzen ez badu itzultzaileak, testua ulertzeko eman beharreko urratsek asko luzatzen dute itzulpen-prozesua. IAK, aldiz, dezente azkartzen du lana lexiko aldetik.

Testu-motari dagokionez, hainbatek aipatu dute administrazio-testuak posteditatzeko denbora gutxiago beharko luketela hutsetik itzultzeko baino.

Harrigarria iruditu zaigu parte-hartzaile bakarrak aipatzea erabilitako itzultzaile automatikoaren arabera ere aldatuko dela denbora. Izan ere, probarako IA aukeratzean metrika automatikoek garbi erakutsi ziguten bataren eta bestearen arteko aldea handia zela eta batzuek beste batzuek baino itzulpen txukunagoa eskaintzen zutela.

Badira, azkenik, posteditatzen denbora gutxiago emango dutela dioten ikasleak ere (4 ikasle).

POSTEDIZIO PROBAN ATERATAKO DATU OROKORRAK

Proba bukatutakoan, TAUS plataformak txosten baten bitartez jakinarazi zizkigun probaren emaitzak. Atal honetan azalduko ditugu, xehe-xehe, proba honetan ateratako ondorioak produktibitateari dagokionez.

Aipatu behar dugu produktibitate datuei erreparatu diegun arren, ez dugula itzulpenen kalitatea ebaluatu itzultzaileen pertzepzioa aztertzea delako gure xede nagusia ikerketa honetan. Beraz, postedizio ariketak ez ditugu bere horretan aztertu, baina aparteko fitxategi gisa kontsultagarri jarri ditugu.

Emaitza orokorreari erreparatzen badiegu, orduko hitz kopurua askoz altuagoa izan da posteditatutako segmentuetan (868) hutsetik itzulitakoetan baino (707). Ondorioz, produktibitatea handiagoa izan da posteditatutako segmentuetan denbora laburragoa behar izan baitute lana gauzatzeko. Itzultzen eta posteditatzen emandako denborari dagokionez ere, aldea nabaria da: segmentuak hutsetik itzultzen denbora gehiago eman zuten (16.535 segundo) posteditatzen baino (11.916 segundo). Hala ere, datu horiek interpretatzean kontuan izan behar dugu, aurretik aipatu bezala, testuari ez diogula inolako egokitzapenik egin, ezta jarduera bakoitzean itzuli beharreko hitz-kopurua berdintzeko ere. Guztira 14 segmentu landu dituzte itzultzaileek proban: 7 posteditatu dituzte eta beste 7 hutsetik itzuli. Gerta liteke segmentu batzuk beste batzuk baino luzeagoak edo laburragoak izatea, horregatik, desoreka txiki bat dago hutsetik itzulitako hitzen artean (141 hitz) eta posteditatutakoen artean (123 hitz).

Probako produktibitate-ratioa (posteditatzen orduko egindako hitz kopurua / itzultzen orduko egindako hitz kopurua) 1.22koa da. Formula horretatik abiatuz, emaitzak 1 baino altuagoa izan behar du baikorra izateko, hau da, postedizioari esker produktibitate altuagoa lortu dugula ondorioztatzeko.

Average Productivity (Info)

Language Pair	Number of Segments	Number of Words	Translation (WPH)	Post-edit (WPH)	Time Spent for Translation (seconds)	Time Spent for Post-edit (seconds)	Post-Editing Productivity Ratio (Info)	MT Engine
Spanish (Spain) > Basque	14	264	707	868	16,535	11,916	1.22	Lucy Software

[Irudia 7 Probako emaitza orokorrak](#)

Parte-hartzaileen datu zehatzei erreparatzen badiegu, parte-hartzaile gehienen kasuan (18) orduko hitz kopurua altuagoa izan da posteditatutako segmentuetan hutsetik itzulitakoetan baino. Beraz, gehiengoak azkarrago lan egin du segmentuak posteditatuta hutsetik itzultzaile automatikoaren laguntzarik gabe itzulita baino. Denborari dagokionez ere, batek izan ezik beste guztiek behar izan dute denbora gehiago segmentuak hutsetik itzultzeko posteditatzeko baino.

Ondorengo atalean ikusiko dugun bezala, proba ondorengo galdetegian parte-hartzaileen % 18ak (4 lagunek) erantzun zuen segmentuak posteditatzeko denbora gehiago behar izan zutela hutsetik itzultzeko baino. TAUS plataformaren datuei erreparatuz, postedizioaren produktibitate-ratioa positiboa izan da 18 kasutan (kasu guztietan 4 parte-hartzaileen kasuan izan ezik). Hortik ondoriozta dezakegu itzultzaileen pertzepzioa ez dela errealitatetik oso urrun ibili.

Productivity per Editor

Evaluator Name	Translation (WPH)	Post-Edit (WPH)	Time Spent for Translation (seconds)	Time Spent for Post-edit (seconds)	Post-Editing Productivity Ratio (Info)
Bat	665	872	762	507	1.31
Bi	813	831	623	532	1.02
Hiru	566	738	895	600	1.3
Lau	894	1,069	567	413	1.19
Bost	588	515	862	858	0.87
Sei	683	1,263	742	350	1.84
Zazpi	631	705	803	627	1.11
Zortzi	761	801	666	552	1.05
Bederatzi	723	762	701	580	1.05
Hamar	1,204	1,009	421	438	0.83
Bat	643	908	788	487	1.41
Hamalau	745	822	680	538	1.1
Bost	584	651	867	679	1.11
Hemezortzi	559	718	907	615	1.28
Hogeiabat	519	690	976	640	1.32
Hogeiabi	732	845	692	523	1.15
Bederatzi	464	714	1091	619	1.53

Hogeitalau	562	1,282	902	345	2.28
Hogeita bi	1,045	1,463	485	302	1.4
Hamaika	670	599	756	738	0.89
Hamabi	823	817	616	541	0.99
Hogeita sei	692	1,024	733	432	1.47

Irudia 8 Probako parte-hartzaile bakoitzaren emaitzak

EMAITZEN AZTERKETA: BIGARREN GALDETEGIA, PROBA ONDORENGO ONDORIOAK ETA IRITZIAK EZAGUTZEA

Proba amaitu eta gero, parte-hartzaileen iritzia eta pertzepzioa ezagutu nahi izan ditugu. Horretarako, bigarren galdetegiko galderak erantzun dituzte parte-hartzaileek eta jarraian horien azterketa egingo dugu.

1. Itzultzaileen iritzia itzulpen kalitatearen inguruan



Grafikoa 8 Noiz izan da hobe, zure ustez, itzulpenaren kalitatea?

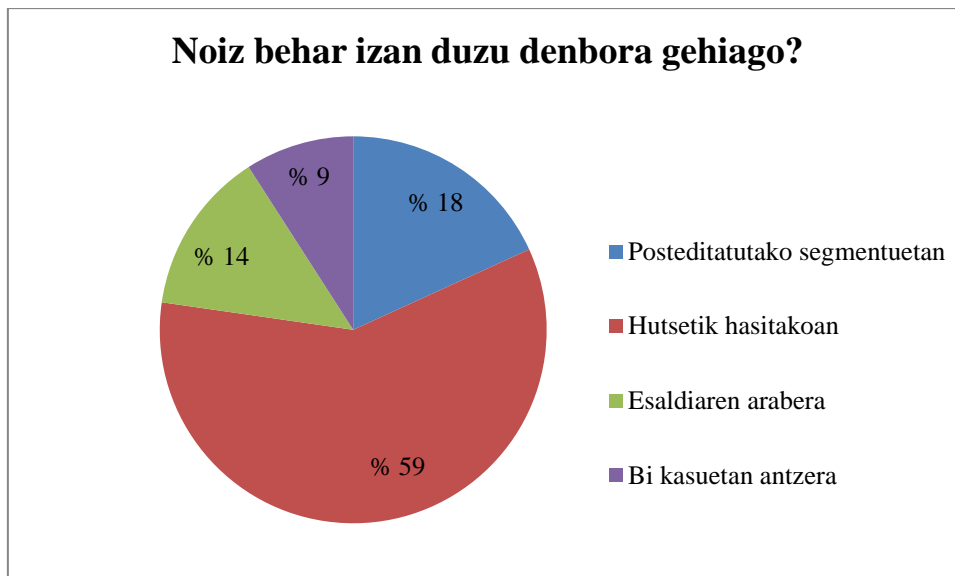
Itzulpenaren kalitateari dagokionean, parte-hartzaileen erdiak, 11 lagunek, hutsetik hasitako itzulpenean kalitate hobe lortu dutela diote. Orokorrean ez da egon bere horretan, itzulpen automatikoari aldaketa bakar bat ere egin gabe, utzi duten segmenturik. Gainera, itzulpen posible bat irakurri dutenez, baldintzatuta sentitu dira asko, eta ez dira asko urrundu proposamenetik. Hutsetik hasitako segmentuetan, aldiz, testutik nahi beste urruntzeko eta libreago jokatzeko aukera izan dutela aipatu dute, naturaltasunaren eta jariotasunaren mesedetan.

Beste batzuek, aldiz, itzulpenaren kalitatea antzekoa izan dela diote. Izan ere, itzulpen automatikoak zituen akatsak zuzendu dituzte eta onargarria eta egokia den itzulpen bat egitea lortu dute. Haien aburuz, hutsetik hasitako testuaren kalitate bera lortu dute, nork buruan zituen kalitate-baldintzak edo irizpideak bete behar izan baititu segmentua onartzeko eta hurrengora pasatzeko. Gainera, askatasun osoz aldatu edo egokitu dute

itzulpen automatikoa. Horrez gain, askotan errepikatu den ideia bat itzulpen automatikoaren kalitatea izan da, nahiko txukuna iruditu baitzaie.

Azkenik, badira posteditatutako segmentuetan kalitate altuagoa lortu dutela uste dutenak ere. Azken horien ustez, lexiko aldetik lagungarri izan zaie itzulpen automatikoa eskuratzatea, eta akatsak zuzentzea ez zaie lan nekeza iruditu. Bestalde, itzulpena hutsetik hasi beharrean oinarri batetik hastea ere lagungarria izan dela aipatu dute, itzultzaile automatikoak haiei bururatuko ez zitzaizkien proposamenak eman dizkielako.

2. Itzultzaileen aurreikuspena itzultzeko edo posteditatzeko behar izan duten denboraren inguruan



Grafikoa 9 Noiz behar izan duzu denbora gehiago?

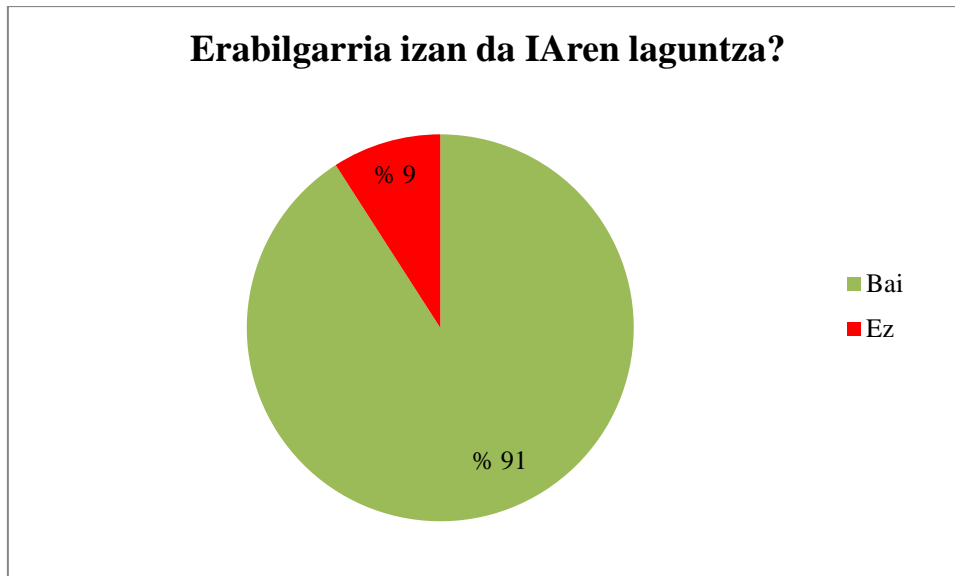
Proba egin ondoren, parte-hartzaileek denborarekiko izan duten aurreikuspena edo pertzepzioa ezagutu nahi izan dugu. Hau da, parte-hartzaileen iritziz zein kasutan behar izan duten denbora gehiago eginkizuna betetzeko, posteditatu beharreko segmentuetan ala hutsetik itzuli beharrekoetan.

Grafikoak adierazten duen bezala, gehiengoak (13 lagunek) hutsetik hasitako segmentuak itzultzeko denbora gehiago behar izan duela dio, posteditatu beharrekoetan baino bilaketa gehiago egin behar izan dituztelako. Posteditatutakoetan, bestalde, esaldiak zuzendu, egokitu eta birformulatzearekin nahikoa izan dela aipatu dute.

4 lagunek adierazi dute posteditatu beharreko segmentuetan denbora gehiago eman dutela. Gehien errepikatu den arrazoiak berrikusketan eta zuzenketetan denbora asko eman dutela izan da, zentzugabeko esaldiak zuzentzen edo terminoak ondo itzuli direla egiaztatzen.

Hiru lagunek diote esaldiaren arabera denbora gehiago edo gutxiago behar izan dutela segmentu bat posteditatzeko edo hutsetik itzultzeko. Eta, azkenik, bi kasuetan gutxi gorabehera denbora bera eman dutela uste duten bi itzultzaile daude.

3. Itzultzaile automatikoaren erabilgarritasuna ataza honetako itzulpen-prozesuan



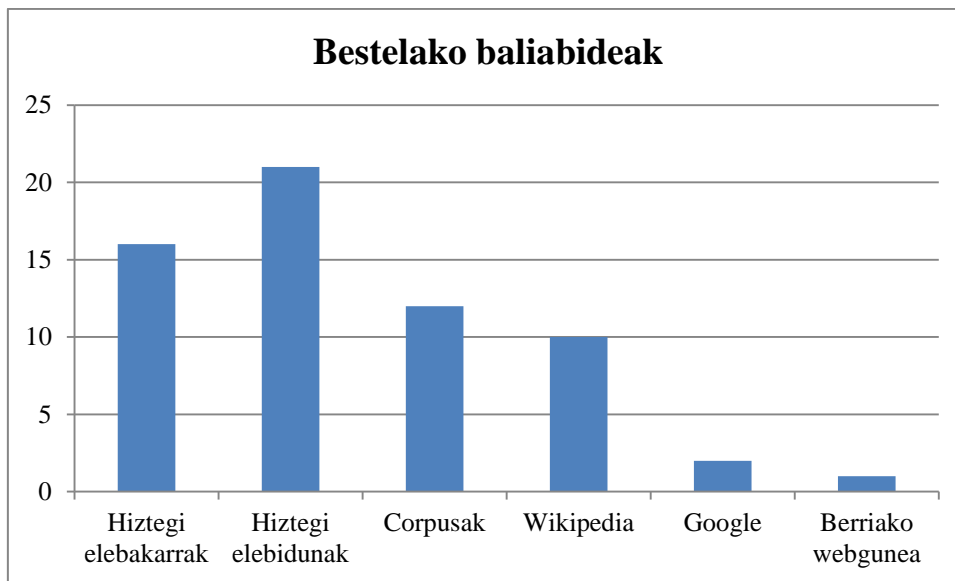
Grafikoa 10 Erabilgarria izan da IAren laguntza?

Grafikoak adierazten duen bezala, gehiengoak dio (20 lagunek⁸) itzultzaile automatikoaren laguntza erabilgarria izan zaiola ataza honetan. Maila lexikoan izan da erabilgarria, batez ere, itzultzaile automatikoaren laguntza, askok aipatu baitute hari esker hiztegi-tako bilaketa-lana aurreztu dutela zuzenean termino eta hitz askoren ordainak eman dizkietelako. Beste kasu batzuetan, aldiz, hiztegi-tara eta bestelako baliabide-tara jo dute itzulpen automatikoko proposamena egiaztatze-ko.

Aldiz, hitzen ordenan, deklinabidean eta syntaxian ez da lagungarria izan hainbaten iritziz.

⁸ 20 lagun horietatik 17k baie-zko borobila eman dute, eta 3 lagunek esan dute kasu batzuetan lagungarria izan dela itzultzaile automatikoaren laguntza.

4. Zer beste baliabide erabili dituzte parte-hartzaileek itzulpen-prozesuan?



Grafikoa 11 Bestelako baliabideak

Itzulpen automatikoz aparte ondorengo baliabideak erabili dituzte parte-hartzaileek posteditatu beharreko segmentuetan nahiz hutsetik itzuli beharrekoetan: batek izan ezik gainerako guztiek erabili dituzte hiztegi elebidunak, 16 parte-hartzailek hiztegi elebakarrak, 12k corpusak, 10ek Wikipedia, 2k Google eta Berriako webgunea batek.

Itzulpen automatikoa erabiltzeak ez ditu salbuetsi ikasleak ohiko itzulpen-prozesuetan erabiltzen dituzten baliabideak erabiltzetik. Kontuan izan behar dugu, dena dela, galdera honetan ez dugula neurtu posteditatutako esaldietan edo itzulitako esaldian erabili dituzten aipatutako baliabideak. Izan ere, oso nahasgarria izango litzateke behin itzulpena amaitu eta gero baliabide bakoitza zein jardueratan erabili duten gogoraraztea.

Bestalde, galdetegietan esan digutenetik ondoriozta dezakegu parte-hartzaile gehienek nahiago izan dutela itzulpen automatikoak proposatutako segmentuetan lexikoa egiaztatu.

5. Parte-hartzaileen iruzkinak eta gogoetak (hautazko galdera)

Bigarren galdetegiaren amaieran gogoetak edo iruzkinak partekatzeko tarte bat utzi genien itzultzaileei, hautazko galdera gisa. Guztira parte hartu zuten 22 itzultzaileetatik 10ek gehitu zuten oharra galdetegiaren bukaeran esanaz esperotakoa baino esaldi txukunagoak proposatu dituela itzultzaile automatikoak eta espero baino lagungarriagoa izan dela IA ataza honetan. Bestalde, hiruk jakinarazi digute hemendik aurrera agian itzulpen automatikoa itzulpen prozesuan txertatuko dutela, beste tresna baten gisa.

ONDORIOAK

Proba hau 22 ikaslek osatutako lagin txiki batekin egin dugu, ezaugarri jakinak dituen testu mota batekin eta IA tresna jakin batekin, beraz, azterketa honetako emaitzak lagin txiki horren testuinguruaren barruan eta probaren ezaugarri jakin horien barruan ulertu behar ditugu.

Lan honetan postedizioari buruzko aurkezpen bat eskaini dugu, eta postedizioak gure testuinguruan duen zabalkundeari erreparatu diogu. Ikusi dugu IA sistemak garatu direla eta postedizioari buruzko azterketa zenbait gauzatu. Hala ere, euskarazko itzultzaile automatikoaren erabilerari dagozkion datuak egun oso baikorrak ez direla ikusi dugu: parte-hartzaileen %40k erabili du IA euskararako norantzan, eta horietatik %13k sarritan.

Jarraian, gure ikerketa azaldu dugu urratsez urrats, eta probatik ateratako emaitzak edo ondorioak aurkeztu ditugu.

Orokorrean, esan dezakegu IA jada badela kontsultarako baliabide bat graduko ikasleen artean, nahiz eta ez izan graduan lantzen dituzten hizkuntza konbinaketa nagusietan, hizkuntza-bikote hegemonikoetan altuagoa baita graduko ikasleek IAri ematen dioten erabilera. IAri ematen zaion erabilerari dagokionez, IA gaztelania-euskara konbinazioan inoiz erabili izan dutela aitortu diguten 15 ikasleetatik 10ek baino ez dute erabili helburu profesionalerako, itzulpen-prozesuan laguntzen duen tresna gisa, alegia.

Proba aurretik parte-hartzaileen %25ak baino ez du erantzun IA lagungarri izan litekeela itzulpen profesional bat egitean eta gehienek mesfidantza adierazi dute. Hala ere, behin proba amaituta, TAUSEk eskainitako emaitzak oso baikorrak izan dira, eta baita parte-hartzaileek eurek atazaren inguruan egindako hausnarketak ere. Itzultzaileen %91k esan du IAren laguntza baliagarria izan zaiola ataza honetan, eta parte-hartzaileen erdiak baino gehiagok uste dute denbora gehiago behar izan dutela segmentuak hutsetik itzultzeko posteditatzeko baino. TAUSEn txostenek berretsi digute hala izan dela, lau parte-hartzailearen kasuan izan ezik beste guztietan izan baita baikorra postedizioaren produktibitate-ratioa.

Azterketa xume honen bitartez postedizioari buruzko hausnarketa egiteko modua izan dugu, ez bakarrik proba antolatzen eta diseinatzen jardun dugunok, baizik eta baita parte-hartzaile guztiek ere. Izan ere, aurreiritzien gaineko hausnarketa egin dute

galdetegian, praktikan IA proban jartzeko aukera izan dute postedizio proban, eta ondoren hausnarketak edo gogoetak partekatu dituzte gurekin proba ondorengo galdetegian.

Orain arte euskarazko postedizioaren ikerketan urrats bat gehiago eman da azterketa honen bitartez eta orain arte aztertu gabeko kolektibo baten, graduiko ikasleen, inguruko informazioa jaso dugu.

Postedizioaren erabilgarritasuna frogatuta geratu da hainbat ikerketetan, eta ikasle gehienentzat ere erabilgarria izan dela ikusi dugu oraingo honetan. Hala ere, lan honetan behin eta berriz errepikatu dugun bezala, faktore askok eragiten dute produktibitatean. Oro har, esan dezakegu “postedizioaren produktibitatea itzultzaile automatikoaren kalitatearen eta posteditoreen espezializazio-mailaren arabera” (Aranberri, 2014: 474) izango dela, eta itzultzailearen hautua izango dela IA itzulpen-prozesuan txertatzea edo ez, lanerako dauzkan ohituren, lehentasunen eta gustuen arabera.

BIBLIOGRAFIA

- Aranberri, N. (2014). Ebaluatoia: crowd evaluation of English-Basque machine Translation. [Kontsulta: 2018-05-22] Erabilgarria: https://addi.ehu.es/bitstream/handle/10810/17841/hapAranberri_ADDI.pdf?sequence=1 helbidean
- Aranberri, N. (2014). Posedición, productividad y calidad. *Tradumàtica: Tecnologies de la Traducció*, 12, 471-477 or. [Kontsulta: 2018-05-22] Erabilgarria: https://ddd.uab.cat/pub/tradumatica/tradumatica_a2014n12/tradumatica_a2014n12p471.pdf helbidean
- Aranberri, N. (2016). Ba al dago lekurik euskararako postedizioarentzat?. *Senez*, 47, 195-203 or. [Kontsulta: 2018-05-22] Erabilgarria: http://www.eizie.eus/Argitalpenak/Senez/20161103/13aranberri/senez47_aranberri.pdf helbidean
- Cortes, I. (2014). Matxin itzultzaile automatikoa. Elhuyar Fundazioa. [Kontsulta: 2018-05-22] Erabilgarria: <http://zientzia.eus/zientzia-irratia/matxin-itzultzaile-automatikoa/> helbidean
- Elhuyar. (D.g.) [Kontsulta: 2018-05-10] Berreskuratua: <https://www.elhuyar.eus/>
- Eusko Jaurlaritzaren itzultzaile automatikoa. (D.g.) [Kontsulta: 2018-03-15] Berreskuratua: <http://www.itzultzailea.euskadi.net/traductor/portalExterno/text.do>
- Garmendia, L., Lasarte, N. & Pinar, M. (2017). Situación actual y viabilidad de la TA al euskera: posedición y análisis de los resultados de un motor de TABR español-euskera. Universitat Autònoma de Barcelona.
- Googleren itzultzailea. (D.g.) [Kontsulta: 2018-03-15] Berreskuratua: <https://translate.google.com/?hl=es>
- Itzultzaile izatetik, posteditore izatera. (D.g.) Berreskuratua: <http://www.ueu.eus/ikasi/ikastaroa-ikusi/855/Itzultzaile%2Bizatetik%252C%2Bposteditore%2Bizatera>
- Ixa taldea. (D.g.) [Kontsulta: 2018-04-28] Berreskuratua: <http://ixa.si.ehu.es/?language=eu>

Koehn, P. & Germann, U. (2014): The Impact of Machine Translation Quality on Human Post-editing. *Workshop on Humans and Computer-assisted Translation*, 38-46 or. Gothenburg: Sweden. [Kontsulta: 2018-05-22] Erabilgarria: <http://www.aclweb.org/anthology/W14-0307.pdf> helbidean

¡La leche! Aldizkaria. (D.g.) [Kontsulta: 2018-03-12] Berreskuratua: <http://revistalaleche.com/somos-la-leche/>

Matxin 2.0 - Euskarazko itzultzaile automatikoaren bertsio berria. (D.g.) [Kontsulta: 2018-03-15] Berreskuratua: <http://matxin.elhuyar.eus/>

Mayor, A.; Alegria, I.; Díaz de Ilarraza, A.; Labaka, G.; Lersundi, M. & Sarasola, K. (2009). Matxin, euskararako lehenengo itzultzaile automatikoa. *Senez*, 37, 197-220 or. [Kontsulta: 2018-05-22] Erabilgarria: <http://www.eizie.eus/Argitalpenak/Senez/20091026/matxin> helbidean

MODELA proiektua. (D.g.) [Kontsulta: 2018-05-21] Berreskuratua: <http://modela.eus/probatu.html>

Plitt, M. & Masselot, F. (2010). A productivity test of statistical machine translation postediting in a typical localisation context. *The Prague Bulletin of Mathematical Linguistics*, 93, 7-16 or.

TAUS. (D.g.) [Kontsulta: 2018-05-01] Berreskuratua: <https://www.taus.net/>

TAUS (2016). Machine Translation Post-editing Guidelines. [Kontsulta: 2018-05-22] Erabilgarria: <https://www.taus.net/think-tank/articles/postedit-articles/taus-post-editing-guidelines>

TAUS (2016). Translation Technology Landscape Report. [Kontsulta: 2018-05-22] Erabilgarria: <https://www.taus.net/think-tank/reports/translate-reports/taus-translation-technology-landscape-report-2016>

Torres-Hostench, O., Presas, M. & Cid-Leal, P. (2016). El uso de traducción automática y posesición en las empresas de servicios lingüísticos españolas: Informe de investigación ProjeCTA 2015. *Bellaterra*. [Kontsulta: 2018-05-22] Erabilgarria: https://ddd.uab.cat/pub/estudis/2016/148361/usotraaut_2016.pdf helbidean

UEU. (D.g.) [Kontsulta: 2018-05-18] Berreskuratua: <http://www.ueu.eus>

ERANSKINAK

I. ERANSKINA: LEHENENGO GALDETEGIA

PROBA AURREKO GALDETEGIA

Erabili al duzu inoiz itzultzaile automatikorik (gaztelania-euskara konbinazioan)?
(Bai/Ez)

- Erantzuna **baiezkoa** bada, markatu erabili izan dituzun itzultzaile automatikoak:
 - Matxin
 - Google Translate
 - Eusko Jaurlaritzaren itzultzaile automatikoa
 - Beste batzuk: _____
- Erantzuna **baiezkoa** bada, zein helbururekin erabili izan duzu itzultzaile automatikoa?

- Erantzuna **ezezkoa** bada, esan zergatik ez duzun inoiz erabili:
 - Ez zait bururatu ere egin.
 - Kalitatea ez da behar bezain ona.
 - Nahiago dut zerotik itzuli.
 - Beste batzuk: _____

Erabili al duzu inoiz itzultzaile automatikorik beste hizkuntza konbinazioen batean?
(Bai/Ez) Zeinetan? _____

Zein da zure iritzia itzultzaile automatikoaren inguruan? Itzulpen profesional bat egiteko lagun diezazukeela uste duzu? Erantzuna baiezkoa bada, esan mesedez zertan gerta dakizukeen lagungarri.

Itzulpen automatikoaren laguntzaz itzulitako testu bat posteditatzen denbora gehiago edo gutxiago emango duzu, zure ustez, testua itzultzen hutsetik hasita baino? Zergatik?

MILA ESKER!

II. ERANSKINA: BIGARREN GALDETEGIA

PROBA ONDORENGO GALDETEGIA

Noiz izan da hobe zure **itzulpenen kalitatea** zure ustez, hutsetik itzulitako segmentuetan edo posteditatutakoetan? Zergatik?

Eta zein kasutan behar izan duzu **denbora** gehiago? Zergatik?

Ataza honetan **erabilgarria** izan da itzultzaile automatikoaren laguntza? Zertan lagundu dizu eta zertan ez?

- Zer **beste baliabide** erabili dituzu itzulpen-prozesuan?

- Hiztegi elebakarrak
- Hiztegi elebidunak
- Corpusak
- Beste batzuk (zehaztu): _____

Beste iruzkinik edo gogoetarik partekatu nahi baduzu, hemen duzu aukera.

MILA ESKER ZURE LAGUNTZAGATIK!

III. ERANSKINA: JATORRIZKO TESTUA

La carrera de los descubridores

EE.UU, Francia & Inglaterra, 1837-1843

Cuando por fin se demostró que la Antártida era un continente, comenzó la ávida carrera por conquistarlo.

El rey francés Luis Felipe I ordenó encontrar y reclamar el polo magnético a Dumont d'Urville. A pesar de que el capitán soñaba con aguas más cálidas, se vio obligado a comandar una expedición en 1837. Un año más tarde, el neoyorkino Charles Wilkes dirigía la expedición científica norteamericana.

Tras innumerables penurias, ambas expediciones se cruzaron tan solo un instante y, sin saber quién ofendió a quién, se produjo el primer conflicto internacional de la historia antártica.

En 1840, James Clark Ross, un experimentado oficial británico de la Royal Navy, partió desde Australia con las naves Terror y Erebus.

En apenas unos meses llegó al cabo Adare y tuvo acceso al navegable mar de Ross. Sin embargo, pronto los bloques de hielo le impidieron seguir avanzando y decidió desembarcar. Ross descubrió el monte Erebus (uno de los mayores volcanes activos de la Antártida); bautizó la costa con el nombre de la reina Victoria. Batió el récord de latitud alcanzado hasta ese momento y, ya de vuelta, cartografió toda la costa pacífica hasta el mar de Weddell.

D'urville, Wilkes y Ross lograron abrir nuevas vías de exploración para la posteridad. También advirtieron al resto del mundo de las duras condiciones del viaje: la infranqueable barrera de bloques de hielo que hacían que las naves se astillaran, la ventisca helada que cortaba las velas como si fueran de papel y las densas brumas que hacían tan difícil la orientación.

IV. ERANSKINA: ERREFERENTZIAZKO ITZULPENA

Aurkitzaileen lasterketa

AEB, Frantzia eta Ingalaterra, 1837-1843

Behingoz Antartika kontinente bat zela frogatu zenean, hura konkistatzeko grina piztu zen.

Luis Filipe I.a Frantziako erregeak Dumont d'Urville-ri polo magnetikoa aurkitzeko agindu zion. Kapitainak ur epelagoak nahi bazituen ere, 1837an espedizio bateko buru izan behar izan zuen. Urtebete beranduago, Charles Wilkes newyorktarra izan zen espedizio zientifiko estatubatuarren buru.

Gorabehera ugari izan eta gero, bi espedizioak momentu batez gurutzatu egin ziren, eta, norgehiagoka nork hasi zuen ez badakigu ere, historia antartikoko lehen gatazka internazionala gertatu zen.

1840an, James Clark Ross, Royal Navy-ko eskarmentudun ofizial britainiar bat, Australiatik abiatu zen Terror eta Erebus ontziekin.

Hilabete batzuen buruan Adare lurmuturrera iritsi zen eta Ross itsaso nabigagarria sarbidea izan zuen. Hala ere, izotz blokeek segidan eragotzi zioten aurrera egitea eta lehorreratzea erabaki zuen. Rossek Erebus mendia aurkitu zuen (Antartikako sumendi aktibo handienetako bat); Reina Victoria izena eman zion kostari. Ordura arteko latitude errekorra gainditu zuen, eta itzulitakoan kostalde Pazifiko osoa kartografiatu zuen, Weddell itsasora arte.

D'urvillek, Wilkesek eta Rossek esploraziorako bide berriak irekitzea lortu zuten etorkizunerako. Bidaia zein gogorra zen ere eman zuten jakitera: ontziak ezpaltzen zituen izotz-blokeen barrera zeharkaezina, belak paperezkoak izango balira bezala hausten zituen haizete hotza eta orientazioa horrenbeste zailtzen zuten lanbro trinkoak.

V. ERANSKINA: ITZULPEN AUTOMATIKOA (MATXIN)

Aurkitzaileen EE lasterketa.

UU, Frantzia & Ingalaterra, 1837-1843

Cuandok azkenik erakutsi zuen Antartika gorpuzkera bat zela, lasterketa egarria hasi zuen lortzeagatik.

Errege frantsesak Luis Filipe I. Aurkitzea ordenatu zuen eta polo magnetikoa eskatzea Dumont D' Urville. Nahiz eta kapitainak urekin amets egiten zuen gehiago beroak, espedizio bat 1837an flota-buru izatera behartua ikusi zen. Geroago urte batean, neoyorkino Charles Wilkesek iparramerikarra espedizio zientifikoa zuzentzen zuen.

Eskasia zenbatezinen atzean, espedizio biak bakar-bakarrik gurutzatu ziren instant bat eta, nor jakin gabe nor mindu zuen, historia antartikoko lehen gatazka nazioartekoa gertatu zen.

1840an, Clark Ross jan dezazun, ofizial britainiarra Royal Navyen esperimentatu bat, Terrorrek eta Erebusen Australiatik zatitu zuen nabeekin.

Apenas hilabete batzuetan Adare azkenean iritsi zen eta Rossen itsaso nabigagarrian sartu ahal izan zen. Hala ere, laster izotzeko blokeek aurreratu galarazitu zioten eta lehorreratzea erabaki zuen. Rossek mendi Erebus aurkitu zuen (sumendi aktibo nagusietako Antartikaren bat) ; kostaldea bataiatu zuen Viktoria erreginaren izenarekin. Latitudeko errekorra menderatu zuen une hori lortua eta, jadanik bueltan, kostalde baketsu guztia cartografiatu zuen Weddellen itsasoraino.

D' urville dezan, Wilkesek eta Rossek esplorazioaren bide berriak irekitzea geroarentzat lortu zuten. Ere bidaiako baldintza gogorretako gainerako munduari ohartarazi zioten: blokeetako izotzeko egiten zuten nabeak ezpal zitezen hesi iraganezina, belak ebakitzen zituen bailitzan bisuts izoztuak paper izan zitezen eta egiten zuten hain zaila orientazioa lainobera dentsoak.

VI. ERANSKINA: ITZULPEN AUTOMATIKOA (GOOGLE TRANSLATE)

Aurkikerien lasterketa

AEB, Frantzia eta Ingalaterra, 1837-1843

Antartikako kontinente bat zela frogatu zuenean, hasperen hasi zen lasterketa aberatsa hasi zen.

Luis Felipe I errege frantsesak Dumont d'Urville-ko Dumont d'Urne magnetikoa aurkitu eta eskuratu nahi zuen. Kapitainak uretan beroago amestutako arren, espedizio bat agindu zuen 1837an. Urtebete geroago, New Yorkeko Charles Wilkesek Ipar Amerikako zientzia-espedizioari ekin zion.

Hainbat zailtasunen ostean, bi espedizioek une batez bakarrik zeharkatu zuten eta, nazien aurka minik egin gabe, Antartikako historiako lehen nazioarteko gatazka gertatu zen.

1840an, James Clark Ross, Royal Navy britainiarraren ofizial esperientziaduna, Australia eta Terror eta Erebus ontziak utzi zituen.

Hilabete gutxiren buruan, Gorputz Adare iritsi eta Ross Itsas nabigaziorako sarbidea izan zuen. Hala eta guztiz ere, laster izotz blokeak eragozten zioten aurrera egiteko eta desegiteko erabaki zuten. Rossek Erebus mendia aurkitu zuen (Antartikako sumendirik aktibo handienetakoa) Queen Victoria izenarekin duen kostaldea, ordura arte iritsi zen latitudea hautsi zuen eta, ondoren, Pacific kostaldeko itsaso osoa mapatu zuen de Weddell.

D'urville, Wilkes eta Rossek jarraitzaileentzako esplorazio modu berriak ireki zituzten. impassable izotz bloke hesia ontziak printza egin duen, izozteak Blizzard bela moztu balitz bezala paper ziren eta trinko lainoa da orientazio hain zaila egin duten: ere munduko gainerako bidaiaren baldintza gogorrak batetik ohartarazi dute .

VII. ERANSKINA: ITZULPEN AUTOMATIKOA (EUSKO JAURLARITZA)

Aurkitzaileen karrera

EE.UU, Frantzia & Ingalaterra, 1837-1843

Azkenik Antartika kontinente bat zela frogatu zenean, karrera egarria hasi zuen hura konkistatzeagatik.

Luis Felipe I errege frantziarrak aurkitzea eta polo magnetikoa eskatzea ordenatu zion Dumont d'Urville-ri. Kapitainak ur beroagoekin amestu arren, 1837an ikusi zen espedizio bat buru izatera behartuta. Urte bat geroago, neoyorkino-a Charles Wilkes-ek espedizio zientifiko iparramerikarra zuzentzen zuen.

Ezin konta ahala eskasien ondoren, bi espedizioak une baino ez bat gurutzatu ziren eta, nori nor mindu zuen jakin gabe, historia antartikoaren nazioarteko lehen gatazka gertatu zen.

1840an, James Clark Ross-ek, Royal Navy-aren eskarmentudun ofizial britainiarrak, Izua eta Erebus zatitu zituen Australiatik ontzietan.

Nekez hilabete batzuetan ez zen Adare lurmuturrari iritsi eta Ross-eko itsaso nabigagarriko sarbidea izan zuen. Hala ere, goiz izotz-blokeek aurrera egiten jarraitzea galarazi zioten eta lehorreratzea erabaki zuen. Ross-ek Erebus mendia aurkitu zuen (Antartikako sumendi aktibo handienetako bat); kostaldea izenarekin bataiatu zuen erreginarekin izenarekin Victoria. Une horretara arte heldutako latitude-marka hautsi zuen eta, itzuliko jada, Weddell-eko itsasorainoko kostalde baketsu guztia kartografiatu zuen.

D'urville-k, Wilkes-ek eta Ross-ek esplorazio-bide berriak gerorako irekitzea lortu zuten. Ohartarazi zioten ere bidaiaren baldintza gogorretako gainerako munduari: ontziak ezpal zitezela eragiten zuten izotz-blokeetako hesi igaroezina, belak papera izango balira bezala mozten zituen izoztutako bisutsa eta orientazioa hain zail egiten zuten lanbro dentsoak.