

Edukia eta indarra: intentsitate emozionala euskarazko itzulpen automatikoan

NORA ARANBERRI MONASTERIO

Hasierako urrats gisa diseinatutako azterketa enpiriko txiki honetan, gaztelaniazko sare sozialetako mezuak itzulpen automatikoaren bidez euskaratzean jatorrizko indar afektiboari nola eusten zaion arakatu du egileak artikulu honetan, txio simulatuen corpus bat osatu, hiru itzulpen-sistemarekin probatu (Itzuli, ChatGPT eta Latxa) eta *crowdsourcing* bidez giza ebaluatzaileekin aztertu ondoren. Emaitez garbi erakutsi dute makinek, mezu biziak itzultzean, intentsitatea lausotzeko eta testua tonu neutroago batean emateko joera dutela.



Edukia islatzea ez da nahikoa

Azken urteotan maiz entzun dugu itzulpen automatikoa gero eta hobe dela. Eta, neurri handi batean, egia da. Euskararen kasuan ere, azken urteetako aurrerapenak nabarmenak izan dira: duela urte batzuk akats nabarmenak egiten zituzten sistemek gaur egun testu ulergarriagoak, zuzenagoak eta erabilgarriagoak sortzen dituzte askotan. Ondorioz, tresna horiek egunerokoan erabiltzen dituzte ikasleek, erakundeek, kazetariak, enpresek eta, jakina, itzultzaileek ere bai, haien tresna espezializatuen artean beste bat.

Hala ere, «itzulpen automatikoa hobetu da» baieztapena potoloegia da, hainbat alderdi on gutxiesten dituen eta hainbat alderdi ahul ezkutatu. Teknologiaren aurrerapena behar bezala deskribatzeko, ezinbestekoak dira askotariko galdera xeheagoak: Zer hizkuntza-bikotetarako hobetu da? Zer testu motatan? Zer komunikazio-egoeratan? Zer irakurlerentzat? Zer helburutarako?

Ez dugu ahaztu behar itzulpenaren kalitatea ez dela ezaugarri bakarra eta abstraktua. Ez dago «kalitate» bakar bat testu guztietan esanahi bera duena. Administrazio testu batek, literatur pasarte batek, osasun-txosten batek, publizitate-kanpaina batek, sare sozialetako txio batek eta lagun bati bidalitako mezu informal batek ez dituzte eskakizun berak. Testu bakoitzak bere funtzioa du, bere igorlea eta bere hartzailea, bere erregistroa eta bere arrakasta-irizpideak.

Batzuetan nahikoa izan daiteke informazio nagusia zuzen ulertzea. Beste batzuetan, ordea, terminologia zehatza eta fraseologia dira funtsezkoak. Edo testuaren eragina bera izan liteke gakoa: konbentzitu egin behar du, hunkitu, haserrea piztu, konplizitatea sortu, barre eragin edo jarrera jakin bat transmititu. Horrelakoetan, «edukia» mantentzea ez da nahikoa; esanahia tonuan, intentsitatean eta testuak irakurlearengan sortzen duen efektuan ere badago.

Horregatik, itzulpen automatikoaren egoeraz hitz egitean, arriskutsua izan daiteke kalitateaz orokorrean aritzea. Baliteke sistema batek oso emaitza onak ematea testu informatibo laburretan, baina sistema bera askoz ere okerrago aritzea testu ironikoetan. Baliteke euskara zuzena darabilen itzulpen bat sortzea, baina jatorrizkoaren ahozkoatasuna, tentsioa edo indarra galtzea. Baliteke testuak «ondo» esatea zer gertatu den, baina ez transmititzea nola bizi duen hori idazleak.

Artikulu honetan, hain zuzen ere, azken alderdi horri begiratuko diogu: emozioari. Zehazki, gaztelaniaz idatzitako sare sozialetako mezuak euskarara automatikoki itzultzean zer gertatzen den aztertuko dugu. Gure interesa ez da soilik jakitea sistemek edukia mantentzen duten ala ez, baizik eta aztertzea ea gai diren jatorrizko testuaren tonu emozionala eta intentsitatea ere emateko.

Zergatik emozioa eta zergatik orain?

Emozioaren azterketa ez da kapritxo teoriko bat, itzulpen automatikoaren eremuan kalitatea ebaluatzeko modu tradizionala kolokan jartzen duen premia praktikoa baizik. Gaur egun, itzulpen automatikoak jauzi kuantitatibo nabarmena egin baitu eta gure komunikazio digitalaren erdigunean kokatu baita, inoiz baino beharrezkoagoa da aztertzea zer gertatzen den hitzez harago. Izan ere, testu baten azken efektua edukia aurkezteko moduan eta eraikitzen duen jarreran dago maiz.

Alabaina, itzulpen automatikoaren ebaluazio estandarrak batez ere bi alderdi aztertu ditu azken hamarkadetan: zehaztasuna, hots, ea jatorrizko esanahia mantentzen den, eta naturaltasuna, hau da, ea testua jatorra den helburu-hizkuntzan. BLEU neurri klasikoetatik hasi eta COMET neurri berrietaraino, helburua izan da sistemak giza itzulpenetara zenbateraino hurbiltzen diren neurtzea (Papineni *et al.*, 2002; Rei *et al.*, 2020). Neurri horiek oso erabilgarriak izan dira sistemen garapenean, baina giza ebaluazioaren inguruko ikerketek erakutsi dute itzulpen automatiko modernoaren kalitatea neurtzea lan konplexua dela, batez ere oinarritzko kalitate mailatik gorantz goazenean, eta ebaluazio-prozedura desegokiek ondorio okerrak ekar ditzaketela (Freitag *et al.*, 2021). Muga horiek direla eta, azken urteotan gero eta interes han-

diagoa piztu du emozioaren, estiloaren eta jarreraren transmisioak hizkuntza-teknologiaren esparruan, argi baitago testu baten baliokidetasuna esanahi semantiko hutsetik harago doala.

Alabaina, emozioa bera ez da erraz definitzen, ez eta neurtzen ere. Hizkuntza-teknologiaren eremuan, sentimenduen eta emozioen analisiek tradizio luzea duten arren, gehienetan lexikoan, polaritatean (positibo/negatibo) edota emozio-kategoria itxietan oinarritu izan dira. Mohammad-ek (2018), ordea, ikuspegi hori zabaldu eta baliabide lexikal sakon bat proposatu zuen, hiru ezaugarritan oinarritua: balentzia, aktibazioa, eta kontrola edo nagusitasuna. Ekarpen horrek argi erakusten du emozioaren adierazpena ezin dela «positibo» edo «negatibo» izatera soilik mugatu, jokoan baitaude intentsitatea, eragin emozionala eta kontrol-sentsazioa bera ere. Itzulpengintzaren ikuspegitik, horrek kezka bat dakar: zer gertatzen da nabardura eta ezaugarri horiek guztiak hizkuntza batetik bestera eramatean?

Ildo horretako emaitzak hasiak dira jada azaltzen itzulpen automatikoaren literaturan ere. Fukuda eta Jin-en (2022) ikerketak, esaterako, japoniar-ingeles txioak aztertuta, agerian utzi zuen itzulpen automatikoak erabat alda ditzakeela jatorrizko emozio-profilak: itzulitako mezuetan, intentsitatea apalagoa izan ohi da, bereziki aktibazio handiko emozioen kasuan. Bestalde, Saadany-k eta kideek (2023) ingelesa, arabiera eta gaztelania barne hartzen zituen azterketa eleaniztunean ondorioztatu zuten, kaleko hizkuntzak, emojiak, kode-aldaketeak edota ortografia ez-estandarrak emozioa distortsionatzea dakarte askotan, nahiz eta itzulpenak zuzenak eta semantikoki sinesgarriak izan. Antzera, Qian-ek eta kideek (2023) txinera-ingeles mikroblog emozionalean ikusi zuten itzulpen askok ez ziotela jatorrizko emozioari behar bezala eusten. Lan horiek guztiek iradokitzen dute emozioaren transmisioa ez dagoela beti bermatuta, testua linguistikoki zuzena eta irakurgarria izan arren. Hala ere, orain arteko ikerketa gehien-gehienak hizkuntza nagusietan egin dira, eta ez zegoen, guk dakigula, euskarazko sare sozialetako edukiaren emozio-intentsitateari buruzko azterketa zehatzik.

Eztabaida horrek, hain zuzen ere, garrantzi handia bereganatu du azkenaldian hizkuntza-eredu handien iraultzarekin. ChatGPTk eta era horretako sistemak aurreko belaunaldiko itzulpen automatikoek baino testu naturalagoak eta diskurtsiboki koherenteagoak sortzen dituzte maiz. Horrek erabiltzaile askoren ikuspegia goitik behera aldatu du, eta makinek testua «ulertzen» duten irudipena zabaldu da gizartean. Alabaina, naturaltasun formal horrek ez du berez bermatzen testuaren dimentsio sozial eta pragmatikoak zuzen jasotzea. Hovy eta Yang-ek (2021) ohartarazi bezala, hizkuntza-teknologiak maiz ezinbidean geratzen dira hizkuntzaren faktore sozialak interpretatu behar dituztenean: nork hitz egiten duen, nori, zer egoeratan, zer harreman sozialetatik eta, batez ere, zer helbururekin.

Muga hori argi ikusten da egun-eguneko eta, nolabait, muturreko testuinguru batean: sare sozialetan. Izan ere, plataformek klik bakar batekin masiboki eskaintzen dituzte gaur egun itzulpen automatikoak, eta euskara ere aukeren artean egon ohi da. Alabaina, hizkuntza oso modu markatuan erabiltzen da espazio digital horietan: berehalakotasuna da nagusi, ahozkotatunetik hurbil dagoen estiloa, umore partekatua, erreferentzia kultural inplizituak eta emozioaren presentzia itzela. Gainera, erabiltzaileek askotan hizkuntza-arau estandarrak hautsi egiten

dituzte, nahita, efektu komunikatibo jakin bat sortzeko. Horregatik guztiagatik, inguruneok erronka paregabea dira itzulpen automatikoarentzat, eta gaur egun argi erakusten dituzte haren mugak.

Euskararen kasuan, muga eta erronka horiek dimentsio berezia hartzen dute. Hizkuntza gutxituen teknologia-garapenari begiratzen badiogu, Joshi-k eta kideek (2020) erakutsi bezala, munduko hizkuntza gehienak eremu digitaletik baztertuta geratzeko arriskuan daude, ordezkaritza eskasarekin. Euskara, ordea, paradoxa baten erdian dago: haren egoera gutxitua kontuan hartuta, azpiegitura gero eta sendoagoa duen hizkuntza den arren, arrakasta horrek berak dakar erronkarik handiena. Tresna automatikoak gure egunerokoan gero eta txertatuago egoteak zuzeneko eragina baitu gure hizkuntza-erabileran. Horregatik, hain zuzen, ezinbestekoa zaigu geure buruari galdera hau egitea: zer-nolako euskara saritzen eta hedatzen dute gaur egungo sistema automatikoeak?

Galdera horri erantzutea ez da erraza. Sistema automatikoeak sortzen duten euskararen inguruko eztabaida intuizioetan eta eguneroko esperientzietan oinarritzen da askotan. Datu enpirikorik ia ez daukagu oraindik, bereziki erregistro informalei dagokienez. Horregatik, sare sozialetako hizkuntza aztergai bereziki interesgarria da. Bertan, hitzunek erregistroak eta baliabideak nahasten dituzte, ahozketasuna da nagusi, eta emozioaren adierazpena hertsiki lotuta dago komunitatearen erabilera-ohiturekin. Horregatik, testuinguru egokia da aztertzeke ea itzultzaile automatikoeak jatorrizko testuaren tonua, erregistroa eta intentsitate emozionala mantentzen dituzten ala ez.

Artikulu honetan deskribatzen den azterketarekin, galdera horiei datuetatik heldu nahi izan diegu, hasierako urrats bat baino ez bada ere. Horretarako, sare sozialetako mezu emozionalen corpus bat osatu genuen, zenbait sistema automatikorekin itzuli, eta emaitzak giza ebaluazioaren bidez aztertu. Hurrengo atalean azalduko dugu xehetasun gehiagorekin nola egin genuen lan hori.

Ikerketa: txio simulatuak, itzulpen automatikoa eta *crowdsourcinga*

Laburbilduz, hemen aurkezten den ikerketaren helburua galdera zehatz bati erantzutea zen: gartzelaniazko sare sozialetako mezuak euskarara automatikoki itzultzean, mantentzen al da jatorrizko intentsitate emozionala?

Makinek intentsitate emozionala nola kudeatzen duten aztertzeke, azterketa enpiriko txiki bat diseinatu genuen, bi ardatzetan oinarritua: aztergai izango den testu-multzoaren (corpusaren) sorkuntza eta giza ebaluatzaileen bidezko balorazio masiboa (*crowdsourcinga*).

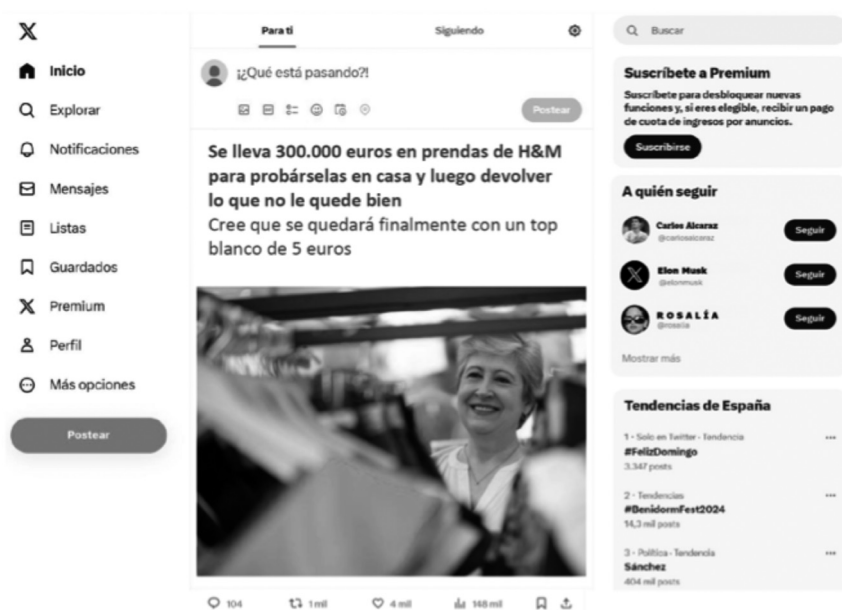
Txio simulatuen corpusa

Erronkari erantzuteko, lehenik eta behin datu-multzo egoki bat behar genuen. Guk dakigula behintzat, ez dago gartzelaniazko eta euskarazko sare sozialetako mezu emozionalen corpus pa-

ralelo egokirik gure helburua duen ikerketa baterako. Sare sozialetako corpusek gai jakin bati buruzko mezuak biltzen dituzte askotan, edo duela urte batzuetako datuetan oinarritzen dira, eta ez dute beti gaur egungo komunikazio informalararen aniztasuna islatzen. Gainera, sare sozialetako mezu publikoen erabilerak hainbat zalantza juridiko eta etiko ere sortzen ditu.

Horregatik guztiagatik, ez genuen benetako txioen bilketarik egin, eta corpus propio bat sortzea erabaki genuen. Horretarako, 19 eta 25 urte bitarteko hemeretzi gazte boluntarioen laguntza izan genuen, guztiek ere gaztelania ama-hizkuntza zutenak. Adin-tarte hori ez zen kasualitatez aukeratu: sare sozialetako plataformen erabiltzaile aktiboak izanik, espero genuen txioen ezaugarri linguistikoak modu naturalean erreproduzitzea.

Parte-hartzaileei ez genien eskatu txio arruntak idazteko; aitzitik, erreakzio-txioak sortu behar zituzten. Horretarako, X plataformaren itxura imitatzen zuen ingurune bat diseinatu genuen (ikusi 1. irudia). Pantailan, albiste baten titularra eta azpitulua agertzen ziren (batzuetan *El País* egunkaritik moldatuak, beste batzuetan *El Mundo Today* hedabide satirikotik hartuak), irudi batekin eta sare sozialetako ohiko elementuekin batera: atsegite kopuruak, birtxioak, gomendatutako kontuak edo joera diren gaiak. Parte-hartzaileek albiste horren aurrean sare sozial batean idatziko luketen erreakzioa idatzi behar zuten.



1. irudia. X plataformaren itxura imitatzen duen ingurunearen adibidea.

Diseinu horren helburua ez zen parte-hartzaileen iritziak jasotzea, baizik eta hizkuntza espontanea eta emozionalki markatua sustatzea. Izan ere, adierazi genien erreakzio-txioek ez

zutela zertan haien ikuspuntua islatu behar. Hasi aurretik, gainera, talde-eztabaida labur bat egin genuen parte-hartzaileekin, txioen ezaugarriei buruz hausnar zezaten: laburtasuna, ironia, informaltasuna edo ahozkotasuna, besteak beste. Horrela, ahalik eta hizkuntza naturalena lor-tzen saiatu ginen. Guztira 472 erreakzio-txio bildu genituen.

Itzulpen automatikoa

Bigarren urratsa txio horiek euskarara itzultzea izan zen. Horretarako, hiru sistema erabi-li genituen. Lehenengoa Itzuli izan zen, Eusko Jaurlaritzaren itzulpen automatikoko zerbi-tzua. Bigarrena ChatGPT (GPT-4o-mini bertsioa) izan zen, gaur egungo erabilera orokorreko hizkuntza-eredu handi hedatuenetakoa bat. Hirugarrena Latxa izan zen, HiTZ zentroak garatu eta euskararako bereziki egokitutako hizkuntza-eredua. Hiru sistemek filosofia desberdinak ordezkatzen dituzte, hala nola itzultzaile neuronal edo AA deritzoguna izatea, hau da, hiz-kuntza-eredu handia edo LLMa, eta pribatua edo askea izatea, eta horrek aukera ematen zuen ikusteko ea emozioaren tratamenduan alde nabarmenik ba ote zegoen. Itzulpenak 2025eko udaberrian bildu ziren.

Hasierako esperimentuan, sistema guztiak ahalik eta baldintza sinpleenetan erabili geni-tuen. Hizkuntza-ereduen kasuan, adibidez, ez genuen jarraibide berezirik eman: txioa itzultze-ko agindu hutsa (oinarrizko *prompta*) erabili genuen. Erabaki hori nahita hartu genuen, sistema guztiak baldintza ahalik eta antzekoenetan erabiltzeko.

Ebaluazio-prozesua eta *crowdsourcinga*

Behin itzulpenak sortuta, ebaluazio-faseari ekin genion. Horretarako, *crowdsourcing* bidezko ebaluazio ireki bat antolatu genuen. Zenbait bide erabili genituen egitasmoa gizarteratzeko eta boluntarioak lortzeko, tartean EIZIEk kudeatzen duen ItzuL posta-zerrenda, deialdia itzultzaile profesionalen komunitatean zabaltzeko. Baliteke, beraz, artikulua honen irakurleren batek ere parte hartu izana ekimenean. Hala bada, eskerrik asko!

Ebaluaziorako, web plataforma espezifiko bat garatu genuen: Berbak.eus. Parte-hartzaile bakoitzak hiru testu ikusten zituen pantailan: albistearen jatorrizko txioa (guk sortua), gazte-laniazko erreakzio-txioa (boluntarioek sortua) eta euskarazko itzulpen bat (hiru sistemetako batena, ausaz aukeratua). Ondoren, zenbait balorazio egiteko eskatu genuen. Lehenik, zehazta-suna neurtzen zen: ea itzulpenak jatorrizko txioaren eduki proposizionala, mezua, mantentzen ote zuen. Bigarrenik, naturaltasuna: ea itzulpena jatorra zen euskaraz. Bi alderdi horiek zazpi puntuko Likert eskala baten bidez ebaluatu ziren, ohiko kalitate-neurketetan ohi denez.

Hirugarren alderdia izan zen gure ikerketaren muina: intentsitate emozionala. Kasu horre-tan, parte-hartzaileek bereiz baloratu zituzten gaztelaniazko txioaren eta euskarazko itzulpe-naren intentsitateak, bost puntuko eskala batean. Mohammad-en (2018) ereduari jarraituz, eskala horren bidez ez genuen bildu nahi hizkuntza bakoitzeko txioek zer emozio maila zehatz zuten, baizik eta bi testuen arteko intentsitate-aldaketaren inguruko balorazioa zein zen.

Jakina, diseinu horrek badu muga bat ere. Parte-hartzaileek jatorrizko txioa eta itzulpena batera ikusten zituztenez, haien balorazioak ezin dira ulertu gaztelaniazko eta euskarazko testuen intentsitatearen neurketa «absolutu» gisa, hizkuntzekiko dugun harremana desberdina delako eta, batera erakustean, bata bestean eragina duelako. Gure helburua, ordea, ez zen hori. Esan bezala, interesatzen zitzaiguna itzulpen-prozesuan hautemandako aldaketa neurtzea zen: irakurle elebidun batek bi bertsioak alderatzen dituenean, emozioaren indarra mantentzen den ala ez esatea; kasu konkretu horretan, haren iritzia. Zentzu horretan, diseinua ez zegoen pentsatuta euskarazko testua bere testuinguruan ebaluatzeko, baizik eta jatorrizkoarekiko hautemandako distantzia neurtzeko.

Balorazio horiek egiteaz gain, parte-hartzaileei aukera eman genien beren euskarazko itzulpen-proposamen propioa idazteko. Helburua etorkizuneko ikerketetarako erreferentzia-materiala biltzea zen, baina, oraindik azterketa hori egiteko bidean gauden arren, jada baliagarriak izan zaizkigu euskarazko erregistro informalaren inguruko hausnarketak egiteko, aurrerago ikusiko dugun bezala.

Guztira 1.098 ebaluazio jaso genituen. Datuak garbitu ondoren, hau da, itzultzaile guztientzako ebaluazioa jaso ez zuten txioak baztertu eta ebaluazio bikoiztuak bateratu ondoren, azken analisian, 334 txiori dagozkien 1.002 ebaluazio erabili genituen. Bestela esanda, gaztelaniazko 334 txio oinarri hartuta, hiru itzultzaile automatikorentzat jaso genituen ebaluazioak.

Merezi du aipatzea parte-hartzaileen gehiengo zabalak aitortu zuela euskal hitzun gaitua zela: ebaluazioen % 75 ama-hizkuntza euskara zuten pertsonen egin zuten, eta % 91k adierazi zuten C1 maila edo altuagoa zutela.

Horrela, ohiko zehaztasunaz eta naturaltasunaz gain, itzulpen automatikoaren beste ezaugarri bati buruzko iritzia ere bildu ahal izan genuen: emozioaren intentsitatearen pertzepzioari buruzkoa.

Emaitza nagusiak

Hasteko, hiru sistemen emaitzak alderatu genituen (Itzuli, ChatGPT eta Latxa), zehaztasuna eta naturaltasuna 1etik 7rako eskalaren bidez neurtuta. 1. taulan ikus daitezke lortutako batezbestekoak.

Sistema	Zehaztasuna	Naturaltasuna
Itzuli	5,28	4,70
ChatGPT	5,21	4,78
Latxa	5,10	4,58

1. taula. Zehaztasunaren eta jariakortasunaren balorazioa (1-7 eskala).

Ikus daitekeenez, hiru sistemek erraz gainditu zuten eskalaren erdiko muga (4 puntukoa). Kasu guztietan zehaztasunari naturaltasunari baino balorazio pixka bat hobea eman bazioten

ere, aldeak ez dira nabarmenak. Askok esango lukete emaitzak ez direla batere txarrak, erdi-muga gaitzitu dutela sistemek. Nik bestela interpretatzen ditut: lortutako zenbakietatik 7ra, hau da, goiko muturrera, bi puntuko tarte dagoela oraindik; nire ustean, hobekuntzarako tarte handia. Hala ere, ez diet sistemei meriturik kendu nahi, eta esango dugu baliabide urriko hizkuntza izanda emaitzak dotoreak direla.

Azpitarratu nahi nukeen beste kontu bat da sistemen artean ez dagoela alderik, hau da, gure txio multzorako, Itzuli erabili zein ChatGPT edo Latxa erabili *prompting* oinarritzkoenarekin, gauza bera da. Egia esan, itzulpenak ez dira berak, baina kalitate orokorra berdina da. Bestela esanda, erabilitako teknologiaren izaera edozein izanda ere, emaitza orokorrak berdinak izan dira.

Gure interes nagusia, ordea, ez zegoen ohiko neurri horietan, itzulpenek emozioaren intentsitateari nola eusten zioten argitzean baizik. Horretarako, ebaluatzaileek preseski baloratu zituzten gaztelaniazko txioen eta euskarazko bertsioen karga emozionalak (1etik 5erako eskalan). Behatutakoa interesgarria da oso (ikusi 2. taula).

Intentsitate emozionala (1-5)	
Gaztelaniazko txioak	3,61
Euskarazko itzulpenak	2,80

2. taula. Intentsitate emozionalaren batezbesteko indizea (1-5 eskala).

Datu horiek ia puntu bateko galera uzten dute agerian. Itzulpenak gramatikalak eta semantikoki zuzenak izan arren, jatorrizko testuaren indar bera transmititzeko ezintasuna erakutsi dute. Horrela, haserrea, harridura, ironia edo ilusioa adierazten duten txioek tonu neutroagoa eta apalagoa hartu dute euskarazko bertsioetan.

Joera hori, gainera, ez zaio sistema bati soilik egotzi; gaur egungo teknologiaren ezaugarri orokorra dela dirudi. Euskarazko intentsitatearen batez besteko puntuazioak oso parekoak izan dira hiruretan: 2,83 Itzuliren kasuan, 2,76 ChatGPTrean eta 2,82 Latxaren azterketan. Diferentzia horiek ez dira estatistikoki esanguratsuak, eta horrek argi iradokitzen du emozioaren moteltzea ez dela tresna jakin baten muga, gaur egungo euskarazko itzulpen automatikoak oro har duen gaitza baizik.

Nolanahi ere, emaitzarik mamitsuena intentsitate mailen arabera azterketan azaleratu da (ikusi 3. taula). Karga emozionalik gabeko txioetan, neutroetan, sistemek txukun mantentzen dute jatorrizko tonua. Alabaina, jatorrizko testuaren emozioa handitzen den neurrian, galerak gora egiten du nabarmen, puntu batetik gora galtzeraino. Txioa zenbat eta biziagoa izan jatorrian, orduan eta handiagoa da itzulpen automatikoak nabardura afektiboak lausotzeko eta testua berdintzeko duen joera.

Gaztelaniazko intentsitatea	Euskarazko intentsitatearen aldaketa		
	Itzuli	ChatGPT	Latxa
1	0,00	0,09	0,05
2	-0,23	-0,25	-0,28
3	-0,67	-0,54	-0,39
4	-1,00	-1,05	-1,00
5	-1,33	-1,34	-1,19

3. taula. Emozio-galera intentsitate-mailaren arabera (batez besteko diferentzia).

Lehen emaitza horiekin kezkatuta, bigarren esperimentu bati ekin genion. Helburua ez zen tresna berri bat probatzea, AA bati testuinguru eta instrukzio (*prompt*) aberatsagoak emateak laguntzen ote zuen neurtzea baizik. Azken finean, horixe da AAren abantailetakoa bat, jarraibideak emanda emaitza hobetzeko edo doitzeko aukera ematen digutela, sistema neuronal tradizionalen aldean.

Horretarako, lehen fasean gaztelaniaz karga emozional handia izan arren euskaraz epel eman ziren 86 txio hautatu genituen. Zehatzago, gaztelaniaz 4 edo 5 mailan kokatutakoak eta euskaraz 1-3 maila esleitu zitzaienak.

Kasu honetan, Latxa erabili genuen AA sistemen ordezkari gisa. Zenbait instrukzio (*prompting*) estrategia baliatu genituen: lehenak testuinguru orokorra eta itzultzaile profesional baten rola zehazten zituen (testuinguru aberastua); bigarrenak, gainera, euskarazko benetako txioen adibideak eskaintzen zizkion sistemari (*few-shot prompting*); eta hirugarren aukerak, aurreko estrategia baliatzeaz gain, Latxaren parametroetan tenperatura altuagoa ezarri zuen, eta horrek sorkuntza-askatasun handiagoa ematen dio AAri.

Emaitzek argi utzi dute *prompting*ak —hau da, jarraibide zehatzek— bide ona zabaltzen dutela (ikusi 4. taula). Hasteko, kalitate orokorrak (zehaztasuna eta naturaltasuna) gorakada nabarmena izan zuen. Baina, horrez gain, emozio-galera nabarmen gutxitu zen: batez beste -1,85 puntutik -1,11 puntura jaitsi zen. *Prompting* estrategien artean ez dago alde estatistikoki esanguratsurik. Hala ere, hobekuntza hori ez zen nahikoa izan arazoa erabat gainditzeko: emaitzarik onenetan ere, emozioaren intentsitatea puntu batetik gora jaisten zen batez beste.

Prompt mota	Zehaztasuna	Naturaltasuna	Emozio-galera
Oinarrizkoa	3,94	3,47	-1,85
Testuinguru aberastua	4,56	4,68	-1,20
<i>Few-shot prompting</i>	4,61	4,84	-1,11
Tenperatura altuagoa	4,35	4,55	-1,20

4. taula. Instrukzioen (*prompting*) araberako kalitatearen eta emozioaren bilakaera.

Ikerketa txikia den arren, eta muturreko testuinguru baten kokatutakoa, zenbait joera agertzen dira. Esan genezake gaur egungo sistemek nahiko ondo itzultzen dutela *zer esaten den*, baina zailtasun gehiago dituztela *nola esaten den* islatzeko, formarekin. Eta zailtasun hori, ikusi dugunez, bereziki larria da karga emozional handiko testuetan. Zehaztasuna eta naturaltasuna soilik neurtuz gero, errealitate hori erraz pasa liteke oharkabeaz; intentsitate emozionalari begiratzean, ordea, gaur egungo sistemen muga agerian geratzen da.

Emozioa, teknologia eta euskararen adierazkortasuna

Aurreko ataleko emaitzek itzultzaile askori ezagun egingo zaien sentsazio bat azaleratu dutelakoan nago. Euskarazko bertsioek ez dute zertan gaizki egon linguistikoki; askotan, erraz ulertzen dira eta jatorrizkoaren oinarritzko esanahia fidelitasunez mantentzen dute. Hala ere, badute hutsune nabarmen bat: ez dute jatorrizko testuaren indar bera. Neurri batean, horixe da gure ikerketan azaleratu den joera nagusia: itzulpenek esanahia transmititzen duten arren, emozioaren intentsitatea ahuldu egiten da maiz, bereziki jatorrizko testuak karga afektibo handia dakarrenean. Horrek, nahitaez, hurrengo galderara garamatza: zergatik dira hain zailak emozionalki markatutako testuak gaur egungo itzulpen automatikoko sistementzat?

Sistema estandarren eta datuen arkitektura-mugak

Gure ikerketaren diseinu enpirikoak galdera horri erantzun borobilik ematen ez dion arren, gure datuetatik abiatuta zenbait azalpen-bide proposa daitezke. Lehenik eta behin, gaur egungo itzulpen-sistema gehienak zehaztasuna eta naturaltasuna optimizatzeko entrenatzen dira. Beste modu batean esanda, sistemek esanahia ez galtzen eta testu sintaktikoki naturalak sortzen ikasten dute, baina ez dute zertan kode berean barneratu emozioaren intentsitatea, ironia, gehiegi-keria edota ahozkotasan markatua mantentzeko algoritmoa.

Bigarrenik, muga entrenamendu-datuetan bertan egon liteke, non erregistro neutroek eta estandarrek pisu erabakigarria duten. Itzulpen automatikoaren corpus klasikoek zati handi bat administrazio-testuek, prentsa-albisteek edota dokumentazio teknikoak osatzen dute. Sare sozialetako hizkuntza, aldiz, erabat urruntzen da formaltasun horretatik: puntuazio ez-estandarra darabil, esamolde efimeroak, umorea, hiperboleak, kode-aldaketak edota komunitate itxi batek baino ulertzen ez dituen erreferentzia kulturalak. Datuen desoreka horren ondorioz, sistemek hobeto ikasten dituzte erabilera neutro eta konbentzionalen ereduak adierazpide markatuagoak baino. Litekeena da horregatik izatea makinek, jatorrizko testuak emozio handia duenean, intentsitatea berdindu eta lausotzeko joera.

Gure bigarren esperimentuak, bestalde, pista interesgarri bat utzi du agerian. Sistemari testuinguru gehiago emateak, itzulpenaren helburua argitzeak edota euskarazko adibide errealek erakusteak emaitzak nabarmen hobetu zituen. Horrek iradokitzen du emozioaren transmisioa ez dagoela erabat sistemaren gaitasunetik kanpo. Hala ere, hobekuntza horiek ez ziren aski

izan emozio-galera erabat desagerrarazteko, eta horrek argi uzten du, *prompting* estrategiek laguntzen duten arren, muga garrantzitsuak dituztela oraindik sistemek. Muga horien jatorria, entrenamendu-datuen izaeraz gain, hizkuntzaren alderdi afektibo eta pragmatikoak modelatzeko sistemek dituzten berezko zailtasun teknikoei egotz dakieke.

Euskarazko ingurune digitalaren erronka soziolinguistikoa

Ebaluazio-prozesuan parte-hartzaileei alternatibak proposatzeko aukera eman zitzaienean, azterketarako guztiz baliagarria den fenomeno bat geratu zen agerian. Zenbait ebaluatzailek adierazi ziguten ez zekitela nola eman euskaraz jatorrizko txioetako esamolde edo ñabardura informal batzuk. Ulertzen zuten gaztelaniazko testuaren indarra, eta hautematen zuten proposatutako itzulpen automatikoa kamuts geratzen zela; baina, alternatiba bat bilatzerakoan, ez zuten beti irtenbide argirik aurkitzen.

Datuetatik ondorio absolutuak ateratzerik ez dagoen arren, galdera bera pizgarria da. Izan ere, sare sozialetako hizkuntza erregistro guztiz konplexua da: hiztunek beren identitatea erakitzen dute bertan, talde jakin bateko kide direla erakusten dute, eta baliabide oso pertsonalak erabiltzen dituzte emozioak adierazteko (interjekzioak, euskalkiak, maileguak edota kodealdaketak). Euskararen kasuan, fenomeno horrek garrantzi berezia hartzen du, euskarazko komunikazio digital informala oraindik ere erabat finkatu gabe dagoen eta etengabe eratzen ari den erabilera-eremua delako. Horrek ez du esan nahi hizkuntzak baliabiderik ez duenik, inondik ere; baina bai adierazkortasun-baliabide informal horiek ez daudela beti modu berean hedatuak edota partekatuak hiztun guztien artean. Horrenbestez, baliteke sistema automatikoen zailtasun batzuk ez egotea soilik teknologian bertan, baizik eta baita modelatu behar duten erregistro digitalaren beraren izaeran eta heldutasunean ere.

Ondorioak eta etorkizuneko norabideak

Errealitate honek galdera sakon bat irekitzen du hizkuntza-teknologiaren etorkizunari eta hizkuntza-politikari begira. Zer egin behar dute sistema automatikoek horrelako egoeretan? Erabilera erreala ahalik eta fidelkien erreproduzitu, ala euskarazko adierazpide eta erregistro informal jakin batzuk indartzen eta zabaltzen lagundu? Bi aukera horien arteko muga ez da beti dirudien bezain garbia. Hizkuntza-teknologiak ikuspegi deskriptibo batekin garatzen badira ere, haien erabilera masiboa denean, deskribapen hutsak ere ondorio normatiboak ditu gizar-tean. Itzulpen automatikoek behin eta berriz formulazio jakin batzuk proposatzen badituzte, erregistro horiek zabaltu, finkatu eta legitimatu egiten dira, adierazkortasun askotarikoago edo berezituago baten kaltetan. Hori horrela, agian merezi du hasieratik hausnartzea zer-nolako euskara sortu eta bultzatu nahi dugun, bai orokorrean, bai ingurune digitaletan.

Ikerketa honek bidea ematen du itzulpen automatikoaren kalitateaz hitz egitean kontuz aritzeko: jada ez da nahikoa galdetzea ea testu batek esanahia mantentzen duen. Batzuetan,

garrantzitsuagoa da aztertzea ea tonu afektiboari eusten dion, ea sinesgarria den, edota ea gai den irakurlearengan jatorrizko efektu eta intentsitate emozional bera sortzeko. Hizkuntza-teknologiek gure eguneroko komunikazioan gero eta presentzia handiagoa izango duten tuinguruetan, ebaluatzen eta neurtzen dugun horrek bideratuko du etorkizuna: neurtzen ditugun alderdiak izango baitira, azken finean, sistemek hobekien ikasiko dituztenak.

Oharra: ikerketaren xehetasunak 15th Workshop on Computational Approaches to Subjectivity, Sentiment Social Media Analysis ekitaldiko aktetan daude irakurgarri (Aranberri, 2026).

BIBLIOGRAFIA

- ARANBERRI, N. (2026ko martxoa). Crowd-based evaluation of emotion intensity preservation in Spanish–Basque tweet machine translation. In *The Proceedings for the 15th Workshop on Computational Approaches to Subjectivity, Sentiment Social Media Analysis (WASSA 2026)* (123.-133. or.).
- FREITAG, M., FOSTER, G., GRANGIER, D., RATNAKAR, V., TAN, Q., & MACHEREY, W. (2021). Experts, errors, and context: A large-scale study of human evaluation for machine translation. *Transactions of the Association for Computational Linguistics*, 9, 1460–1474.
- FUKUDA, K., & JIN, Q. (2022). Analyzing change on emotion scores of tweets before and after machine translation. In *Social Computing and Social Media: Design, User Experience and Impact* (pp. 294–308). Springer International Publishing.
- HOVY, D., & YANG, D. (2021). The importance of modeling social factors of language: Theory and practice. *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, 588–602.
- JOSHI, P., SANTY, S., BUDHIRAJA, A., BALI, K., & CHOUDHURY, M. (2020). The state and fate of linguistic diversity and inclusion in the NLP world. *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, 6282–6293.
- PAPINENI, K., ROUKOS, S., WARD, T., & ZHU, W. J. (2002ko uztaila). Bleu: a method for automatic evaluation of machine translation. In *Proceedings of the 40th annual meeting of the Association for Computational Linguistics* (311.-318. or.).
- QIAN, S., ORASAN, C., DO CARMO, F., LI, Q., & KANOJIA, D. (2023). Evaluation of Chinese-English machine translation of emotion-loaded microblog texts: A human annotated dataset for the quality assessment of emotion translation. In *Proceedings of the 24th Annual Conference of the European Association for Machine Translation* (125.-135. or.).
- REI, R., STEWART, C., FARINHA, A. C., & LAVIE, A. (2020). COMET: A neural framework for MT evaluation. *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing*, 2685-2702.
- SAADANY, H., ORASAN, C., CARO QUINTANA, R., DO CARMO, F., & ZILIO, L. (2023). Analysing mistranslation of emotions in multilingual tweets by online MT tools. In *Proceedings of the 24th Annual Conference of the European Association for Machine Translation* (275.-284. or.).
- SAIF MOHAMMAD. (2018). Obtaining Reliable Human Ratings of Valence, Arousal, and Dominance for 20,000 English Words. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 174.-184. or., Melbourne, Australia. Association for Computational Linguistics.