

GRADU AMAIERAKO LANA

EUSKAL WIKIPEDIA AZTERTZEN: ITZULPEN PROZESUAREN AUTOMATIZAZIOA ETA POSTEDIZIOA

eman ta zabal zazu



Universidad
del País Vasco

Euskal Herriko
Unibertsitatea

Esther Oar-Arteta Estebanez

Itzulpengintza eta Interpretazioko Gradua

2022/2023 Ikasturtea

Tutorea: Zuriñe Sanz Villar

Saila: Ingeles eta Aleman Filologia eta Itzulpengintza eta Interpretazioa

1. Laburpena

Lan hau Wikipedian gauzatzen den itzulpenean eta postedizioan zentratu da, helburuak honako bi hauek izanik: alde batetik, itzultzaile profesionalak ez diren wikilarien lana arautzen duen politikarik dagoen ikustea; eta, bestetik, Elia *Content Translation* plataforman txertatu denetik, Wikipedian burutzen den postedizioa (PE) zer nolakoa den aztertzea.

Horretarako, lehenik eta behin, Wikipediak eskaintzen dituen eztabaida-orriak aztertu ditugu, bai eta Euskal Wikilarien Kultur Elkarte (EWKE) deritzon webgunea ere. Wikipedia *crowdsourcing* bidez elikatzen den entziklopedia aske handienetarikoa bat da, hau da, edukiak milaka erabiltzaile boluntarioren elkarlanari esker osatzen dira. Halere, edukiez gain, politika formal eta informalak ere lankideen artean negoziatzen dira, horretarako ezarritako eztabaida-orrietan. EWKE, ostera, euskal wikilari osatutako elkarte bat da, zeinak online plataforma bat baituen entziklopediari buruzko berriak eta artikulak argitaratzeko.

Bigarren helburua lortzeko, bestalde, postedizio-azterketa bat gauzatu dugu. Horretarako, lehenik eta behin, bi lagin sortu ditugu: bata, sorburu-testuez eta *Content Translation* gailuan instalatuta dagoen itzultzaile automatikoaren (Eliak) bitartez sortutako itzulpenez osatua; eta bestea, berriz, Eliak sortutako itzulpenak abiapuntutzat hartuta, wikilariak posteditatutako artikuluez osatua. Behin hori eginda, postedizio-azterketa bi zatitan banatu dugu. Lehen partean, Eliak egindako akatsak identifikatu eta zenbatu ditugu. Bigarren partean, ostera, erabiltzaileen postedizioak aztertu ditugu eta gauzatu dituzten aldaketak anotatu.

Lana hiru bloketan banatuta dago. Lehenengo blokea eduki teorikoari eskaini diogu. Hala, zati honetan zehar Wikipediari buruzko informazioa helarazteaz gain (entziklopedia askea zer da, euskal Wikipedia komunitatea etab.), Wikipedia eta itzulpenari buruz gaur arte argitaratu diren ikerketa ezberdinak jaso ditugu. Horrez gain, postedizioa definitzen ere saiatu gara. Bigarren blokea, berriz, alderdi praktikoari eskaini diogu, hots, ikerketan zehar erabilitako metodologia zein izan den azaldu dugu. Hirugarren eta azken blokean, azterketen emaitzak plazaratu ditugu, hau da, Wikipediako itzulpen politikez eta postedizio-azterketen emaitzez jardun dugu. Amaitzeko, ikerketa honetatik ateratako ondorioak partekatu ditugu.

AURKIBIDEA

1.Laburpena	2
2.Sarrera.....	5
3.Wikipedia, entziklopedia askea	6
4.Wikipedia ikergai gisa	7
5.Postedizioa	11
6.Helburua eta metodologia.....	12
6.2.Euskal wikilarien lagina osatzeko irizpideak	13
6.3.Itzulgaien aukeraketa	15
6.4.Postedizio-azterketaren prozedura	17
6.4.1.Akatsen identifikazioa: IAk emandako itzulpenetako akatsen anotazioa.	17
6.4.2.Posteditatutako testuen analisisia: zuzenketa fasea.....	19
7.Emaizak	20
7.2.Euskal Wikipediako itzulpen-politikak.....	20
7.3.Postedizio-azterketaren emaitzak.....	22
7.3.1.Akatsen identifikazioa: IAk emandako itzulpenetako akatsen anotazioa.	22
7.3.2.Posteditatutako testuen analisisia: zuzenketa fasea.....	25
7.3.2.1.Azterketa kuantitatiboa	25
7.3.2.2.Azterketa kualitatiboa	28
8.Ondorioak	33
9.Erreferentzia bibliografikoak	35
10.Eranskinak	39
10.1.Erdal Wikipedietako artikuluak	39
10.2.Euskal Wikiledian argitaratutako bertsiok	40

IRUDIEN AURKIBIDEA

Irudia 1: Lankideen babel informazioa.....	15
Irudia 2: MQM taxotomia.....	18
Irudia 3: Itzulpenerako argibideak <i>Conten Translationen</i>	21
Irudia 4: Eliaren akatsak guztira	24
Irudia 5: Eliaren akatsak: doitasuna.....	24
Irudia 6: Eliaren akatsak: jariotasuna.....	25
Irudia 7: Wikilarien postedizioak guztira	27
Irudia 8: Wikilarien postedizioak: gehitutako akatsak	27
Irudia 9: Wikilarien postedizioak: mantendutako akatsak.....	28

TAULEN AURKIBIDEA

Taula 1: Eliak egindako akatsak (balore absolutuak)	22
Taula 2: Wikilarien postedizioak (balore absolutuak).....	25
Taula 3: Gehitutako akatsa (jariotasuna)	29
Taula 4: Gehitutako akatsa (doitasuna)	29
Taula 5: Gehitutako akatsa (tokiko konbentzioak).....	30
Taula 6: Gehitutako akatsa (estiloa)	30
Taula 7: Mantendutako akatsak (jariotasuna).....	31
Taula 8: Mantendutako akatsak (doitasuna)	31
Taula 9: Mantendutako akatsak (tokiko konbentzioak).....	31
Taula 10: Mantendutako akatsak (estiloa).....	32

2. Sarrera

“Wikipedia entziklopedia askea” gaur egun gehien kontsultatzen den informazio-iturrietako bat da; nahiz eta sarritan ez den fidagarritzat jotzen. Bere izenak dioen bezala, *askea* da, hau da, edonork editatu edo kontsultatu dezake. Are gehiago, ikerlariek Wikipedia *crowdsourcing* terminoarekin deskribatu dute, hizkuntza-komunitate ezberdinetako boluntario talde batek egindako elkarlanari esker osatzen baita. Gainera, informazioa hutsetik sortu daiteke, edo bestela, itzulpenaren bitartez osatu. Azken horren presentzia handiagoa da komunitate txikietan handietan baino; lehenengoen eduki-falta nabarmenagoa eta wikilari gutxiago baitituzte. Hori horrela, Wikipedia ingurune eleanitz bat da, eta itzulpena, ostera, entziklopedia askea osatzen duten komunitate guztietara informazioa modu azkar eta eraginkor batean zabaltzeko tresna. Dena den, plataforman gauzatzen den itzulpena ez da profesionala, eta, hortaz, erabiltzaileek ez dute izaten itzulpengintzari lotutako formakuntzarik.

Azken urteotan itzulpenak bizitako automatizazio-prozesuaz geroztik, fokua itzulpena izatetik PEa izatera pasatu da. Baina aldaketa hori ez da soilik itzulpen-ingurune profesioaletan gauzatu, baita Wikipedian ere. Alegia, 2014. urtetik martxan da *Content Translation* gailua, zeinak sorburu-testua eta xede-testua parez pare ikusteko aukera ematen baituen. Horrez gain, hasieran, erregeletan oinarritutako itzultzaile automatikoa erabil zitekeen plataforma horretan, baina, gaur egun, itzultzaile automatiko neuronalak dira erabilgarri daudenak. Euskararen kasuan, 2021. urtetik, Elhuyarrek garatutako IA neuronalak, Elia, da wikilariak baliatzen. Hori horrela, gaur gaurkoz erabiltzaileek ez dute testuak hutsetik itzultzeko beharrik, horiek zuzentzekoa baizik, hots, posteditatzekoa.

Horrenbestez, jakinda itzulpena baliabide lagungarria dela entziklopedia askean, lan honen bitartez Wikipedian itzulpen-politikarik dagoen aztertuko dugu. Bestetik, ikusirik Wikipediako itzulpenaren automatizazioak gero eta posteditazio behar handiagoa dakarrela, azken hori zer nolakoa den ikertuko dugu, lagin batean oinarrituta.

Lan hau hiru bloketan banatu dugu. Lehenengo blokean alderdi teorikoari erreparatuko diogu: hasteko, Wikipedia zer den azaldu eta euskal Wikipediako komunitatearen berri emango dugu (3. atala). Ondoren, orain arte Wikipedia eta itzulpenaren inguruan egin diren ikerketak plazaratuko ditugu (4. atala). Jarraian,

postedizioa definitzen saiatuko gara (5. atala). Bigarren blokean (6. atala), ikerketa aurrera eramateko erabilitako metodologia aurkeztuko dugu (euskal wikilarien lagina osatzeko irizpideak, itzulgaien aukeraketa eta postedizio-azterketaren prozedura). Azkenik, hirugarren blokean, ikerketaren emaitzak eta azken ondorioak jasoko ditugu (7. eta 8. atala).

3. Wikipedia, entziklopedia askea

Wikipedia Jimmy Wales eta Larry Sangerrek, 2001. urtean, elkarlanean sortutako eduki askeko entziklopedia bat da. Entziklopedia irabazi asmorik gabeko Wikimedia Fundazioak babesten du (Wikipedia, 2023). Wikipediaren helburua ahalik eta hizkuntza gehienetan idatzita dagoen doako entziklopedia fidagarri bat sortzea da. Gaur egun, 323 hizkuntza komunitate daude entziklopedia askean, zeintzuen artean ingeles komunitatea baiten handiena 6,5 milioi artikulurekin.

Wikipedia lankidetzatza proiektu erraldoi bat den neurrian, milioika erabiltzailearen lanari esker sostengatzen da. Erabiltzaile horiek ez dute informazio pertsonalik partekatzeko beharrik, hau da, anonimoak dira. Gainera, Wikipediak bi multzotan banatzen ditu erabiltzaileak: aktiboak, azken 30 egunetan ekarpenak egin dituztenak, eta pasiboak.

Bestalde, Wikipediako jarduera “Bost euskarriak” deituriko oinarritzko bost legek arautzen dute, hala nola: 1) Egiaztagarritasuna. Wikipedia entziklopedia bat denez gero, ezinbestekoa da edukiak egiaztagarriak izatea eta, hortaz, iturri fidagarrietara garamatzaten erreferentziak txertatu behar dira. 2) Neutraltasuna. Wikipediako artikulua ezin daitezke subjektibotasunez josi. Hau da, gai baten inguruan existitzen diren ikuspegi guztiak bildu behar dira, bakar bat ere goraipatu gabe. 3) Eduki askea. Wikipedian jasotako informazioa askea da, edonork erabil eta edita dezake. Horrenbestez, ezin daiteke egile eskubide librea ez duen materialik erabili. 4) Errespetua. “Wikipedian dihardutenek errespetuz eta gizalegez tratatu behar dute elkar” (Wikipedia: Bost euskarriak, 2022). 5) Arau finkorik eza. Bost euskarriez gain, Wikipediak ez du arau finkorik eta, hortaz, ekarpenak egiteko gune aproposa da (Wikipedia, 2022).

Euskal Wikipedia komunitateari dagokionez, 404.994 artikulua sortu dira jada. Gaur gaurkoz, Euskal Wikipedia komunitateak 458 wikilari aktibo ditu. Dena den, zifra horrek ez du errealitatea islatzen, lehen esan bezala, erabiltzaile aktiboak azken 30 egunetan ekarpenak egin dituztenak baino ez direlako. Aipatzekoa da, gainera, erabiltzaile aktiboak zazpi multzo ezberdinetan sailkatzen dituela euskal Wikipediak, zeintzuen arteak administratzaileak eta burokratak baitaude. Administratzaileei Wikipediaren garbitzaile ere esaten zaie, hain zuzen, entziklopedia askea txukun mantentzeko ardura baitute. Bigarrenik, burokratek beste edozein erabiltzailek egin ditzakeen jarduerak gauzatzeaz gain, wikilari bat administratzaile edo burokrata bihur dezakete, bai eta lankide bati *Bot* estatusa eman edo kendu ere (Wikimedia, 2023).

Euskal Wikipedia, edukiari erreparatuta, handia izan ez arren, komunitate aktibo eta ekintzaile batek osatzen du. Horren erakusle da, adibidez, Euskal Wikilarien Kultur Elkarte (EWKE). Elkarte hau 2016. urtean sortu zen bi helburu nagusi betetzeko asmoarekin: alde batetik, euskal Wikipedia osatzea; bestetik, Wikimedia Fundazioak babestutako proiektuak euskaraz garatzea.

4. Wikipedia ikergai gisa

Wikipedia proiektu eleaniztun erraldoi bat da eta, horrenbestez, ezin daiteke ukatu itzulpena azken horren euskarrietako bat dela. Góngora-Goloubintseff-ek (2017) dioenez, entziklopedia askea fenomeno berria ez izanik ere, ikerketa gutxi egin dira haren inguruan. Baina Wikipediaren izaera aldakor eta kolektiboak adituen arreta erakarri du azken hamarkadan. Izan ere, 2021. urteko abenduaren 15 eta 17 artean, Hong Kong Baptist University-k *Understanding Wikipedia's Dark Matter: Translation and Multilingual Practice in the World's Largest Online Encyclopedia* izeneko kongresua antolatu zuen. Bertan hainbat ikertzaile elkartu ziren, ikerketen emaitzak eta ideiak partekatzeko: besteak beste, Henry Jones-ek Wikipediako itzulpena ikertzen dutenen erronka teoriko eta metodologikoei buruz hitz egin zuen; Mark Shuttleworth, berriz, Wikipedia aztertzeke garatutako tresna digitalei buruz jardun zen. Hori horrela, jarraian, Wikipedia eta itzulpena ikergaitzat dituzten artikuluetako batzuk aipatuko ditut, gaur arte, alor horretan egin denaren ikuspegi orokor bat emate aldera.

Wikipedia eta itzulpena biltzen dituzten proiektuei dagokienez, McDonough Dolomaya (2014) ikerlariak *Wikipedia: Pages needing translation to English* Wikiproiektuan zerrendatuta zeuden 94 itzulpen hartu, eta corpus bat sortu zuen haiekin. Hain zuzen, itzulpenon “aldaketen historialak” aztertuz, editoreek egindako edizio, errebisio eta berridazketak ikusi ahal izateko. Hala, McDonough ohartu zen entziklopediaren izaera askea lagungarria izan daitekeela itzulpenetan akatsak identifikatu eta zuzentzeko, nahiz eta denbora-tarte luze bat pasatu erroreak guztiz konpondu arte. Bestetik, ikusi zuen batez ere transferentzia eta gramatika akatsak direla erabiltzaileek sarrien konpontzen dituztenak.

Wikipedia entziklopedia aske bat izan arren, ez da harrizkoa askatasun hori mugatzen duten neurriak existitzea, baita itzulpen praktikei dagokienez ere. Hori erakusten du, behintzat, Góngora-Goloubintseff-ek (2021) bere tesian. Ikerlariak Wikipediako lau komunitate ezberdin aztertu ditu (espainiarra, frantsesa, alemana eta suediarra) entziklopedia askeko arauak nola erregulatzen eta negoziatzen diren jakiteko, bai eta lau komunitate horietako erabiltzaileek arauok zein neurritan barneratu dituzten ezagutzeko ere. Hori lortzeko, bi metodo erabili ditu: batetik, Wikipediak eskaintzen dituen “Eztabaida”¹ izeneko orriak kontsultatuz eta, bestetik, komunitate horietako 16 kide elkarrizketatuz.

Góngora-Goloubintseffen aurkikuntzen arabera, Wikipediako jarduera bost legek (“Bost euskarriak”) arautzen dute, baina badaude baita *policy* (politika formalak) eta *guidelines* (politika informalak) deiturikoak ere. Politika formalak komunitate-bozketa bidez erabakitzen dira, eta gehiengoaren babesa jaso behar dute. Politika informalak, berriz, “wikilari batek edo wikilari talde batek proposatutako politikak dira”. Gainera, politika informalak formal bihurtu daitezke bozketa bidez, nahiz eta horretarako ez den administratzaile baten onespenerik behar (Wikipedia, 2018). Politikak negoziatzeko, erabiltzaileek alde aurre aipatutako eztabaida-orriak baliatzen dituzte. Nabarmentzekoa da, printzipioz, negoziaketak wikilari guztien artean egin beharko liratekeela; hain zuzen, Wales-ek (2022), Wikipediako sortzaileetariko batek, dio wikilari berriak ongietorriak direla eta ezin daitekeela egon erabiltzaile berriei sarrera oztopatuko dien elite edo hierarkiarik. Dena den, errealitatea oso bestelakoa da. Hain zuzen ere,

¹ “Eztabaida” orriak (*Talk pages*): Wikipediaren paratestuetako bat dira. Bertan sartzeko, artikuluko bakoitzaren goi-ekerraldean dagoen “Eztabaida” etiketaren gainean klikatu behar da. Eztabaida orri horiek foro bat bailitzan funtzionatzen dute, hots, erabiltzaileek elkarren artean hitz egiteko balio dute (Jones, 2013, 84. or.).

Góngora-Goloubintseffek adierazi duenez, politikak ezartzeko bozketetan *full members*² gisa izendatzen diren erabiltzaileek baino ez dute parte hartzen, hots, elite batek. Bestalde, itzulpengintzari begira, ikerlariaren arabera komunitate guztiek politika eta gidalerro berak partekatu ez arren, “the core set of values, expectations and materials that underpin and configure the practice are similar across the Wikipedia communities under investigation” (Góngora-Goloubintseff, 2021, 231. or.).

Bigarren helburuari dagokionez, elkarrizketatuen erantzunek erakusten dute Wikipediako lankideek oinarrizko legeak baino ez dituztela barneratu. Hau da, lehenetsua ematen diete goian aipatutako derrigorrezko edizio-arauei eta baztertu egiten dituzte tokian tokiko itzulpenerako gidalerroak. Hori horrela, Góngora-Goloubintseffek Shuttleworthek (2017) dioena berresten du, zeinak defendatzen baituen Wikipedian itzulpenak eta edizioak continuum bat osatzen dutela, hots, itzulpena eta edizioa ez direla bi fenomeno bereiz, elkarrengarriak baino (Góngora-Goloubintseffek, 2021).

Wikipediako politiken hariari tira eginez, McDonough Dolmayak (2017) entziklopedia askeko hizkuntza-politikak, itzulpen-politikak eta itzulpen-norabideak aztertu ditu, bai eta itzulpen-fluxu horiek modu justu eta eraginkor batean kudeatzeko era ere. Wikipedian bertan hizkuntza-politikarik aurkitu ezin daitekeen arren, Wikimediak *Language Proposal Policy* deituriko arau-multzoa du. McDonoughen arabera, hizkuntza-politika hori Alan Patten-ek (2001) proposatutako eruedetako batekin lotu daiteke: eleaniztasun ofiziala. Horrek esan nahi du entziklopedia askeak errekonozimendu bera ematen diela hizkuntza komunitate guztiei. Itzulpen politikari dagokionez, ez Wikimediak ezta Wikipediak ere, ez dute itzulpen-politikarik (McDonough 2017; Góngora-Goloubintseff 2021); nahiz eta *Language Proposal Policy* deritzon dokumentuan eleaniztasun ofizialaren eruedari jarraitzen dioten itzulpenari buruzko iruzkinak egon. Horrenbestez, itzulpenak hizkuntza handietatik handietara eta hizkuntza handietatik txikietara egiten dira, hots, ez da aparteko ahaleginik egiten artikulu kopuru txikiagoa duten komunitateen produkzioa sustatzeko; *crowdsourcing*ik gabeko testuinguruetan bezala. Hori horrela, McDonough Dolomayak justizia linguistikoaren kontzeptua gogora ekartzen du, erakusteko etorkizunean sor daitezkeen itzulpen-politika posibleek nola lor dezaketen justiziaren eta eraginkortasunaren arteko

² *Full members*: administratzaileak, burokratak eta beste funtzionario batzuk, bai eta esperientziadun erabiltzaileak ere. 100 ekarpen baino gehiago dituzten lankideak (Góngora-Goloubintseff, 2021).

oreka handiagoa; indarrean dagoen eleaniztasun ofizialaren modeloa errespetatuz, alegia (McDonough Dolomaya, 2017).

Itzulpen prozesua automatizatzeko aurrerapauso handiak eman diren arren, garrantzitsua iruditzen zait euskal Wikipedia eta erregeletan oinarritutako itzultzaile automatikoa (IA) uztartzen dituen artikulu bat mahaigaineratzea. Iñaki Alegria et al.-ek (2013), erregeletan oinarritutako itzultzaile automatikoaren bitartez, Wikipediako 100 artikulu itzuli zituzten gaztelaniatik euskarara. Ondoren, euskal Wikipediako boluntario talde batek itzulpenok zuzendu zituen, emaitzak *Statistical Post-Editing* (SPE) sisteman txertatuz. Hala, bi helburu bete zituzten: euskal Wikipediako artikuluen kopuruak gora egitea eta IA bidezko itzulpenen kalitatea hobetzea.

Esan bezala, itzultzaile automatikoei abantailak dakarzkie erabiltzaileei. Halaber, emaitzak oraindik ez dira guztiz zuzenak eta gizakion esku hartzea ezinbestekoa da. Izan ere, makinek sortutako itzulpenak berrikusi egin behar dira, posteditatu. Wikipediara bueltatuz, Solaunek (2021) ikusi zuen entziklopedia askeak ez zuela boluntarioei postedizioa aurrera eramaten lagunduko zien gidalerrorik eta, hortaz, jarraibide batzuk sortu zituen. Horretarako, postedizio proba bat diseinatu zuen. Proba hartan zazpi wikilarik hartu zuten parte, eta helburua gaztelaniaz idatzitako 200-300 hitz inguru zituzten testuak gainbegiratzea zen, postediziorako oinarritzko gidalerro batzuk erabiliz. Horrela, hasierako jarraibide minimo horiek osatzea lortu zuen.

Berrikuntzen ideiarekin jarraituz, gaur egun, erregela bidezko itzultzaile automatikoak baino garatuago dauden tresnak baliatzen dira itzulpenetan. Wikipediari dagokionez, *Content Translation* (CX) deritzon tresna martxan dago 2014ko urtarrilaren 1etik. Gailu teknologiko gehientsuekin gertatzen den bezala, hobekuntzak egitea ezinbestekoa denez gero, hala egin zuten Niklas Laxström, Pau Giner eta Santhosh Thottigalek (2015). Izan ere, Wikipediako editoreen feedbacka baliatu zuten egun entziklopedia askean hain erabilia den tresna hobetu ahal izateko. Zenbait komunitatek automatizazioari muzin egiten badiote ere (kalitate txarreko artikuluen ugaritzea dela medio), erabiltzaile askok lan-karga murrizteko baliabidetzat jotzen dute (Góngora-Goloubintseff, 2021). Alegia, tresna horrek hainbat abantaila dakarzkio artikulua itzultzen duenari, hala nola: sorburu-testua (ST) eta xede-testua (XT) parez pare ikus daitezke; aldaketak automatikoki gordetzen dira; XTaren formatua aldatzeko beharrik ez, tresnak berak egiten baitu lan hori; paragrafo baten % 85a IAren itzulpen gordina

baldin bada, hots, aldaketarik gabea, erabiltzaileak abisu bat jasoko du “arreta kantitatean baino kalitatean jar dezan animatuz”, eta abar (Laxtröm, Giner, Thottigal, 2015, 2. or.).

Aurrerapen teknologikoez jada ezarrita dauden sistemak hobetzea dute helburu, baina berrikuntzak onuragarriak eta kaltegarriak izan daitezke. Lehenago aipatutako Góngora-Goloubintseffen tesian politikak eta gidalerroak aztertzeaz gain, azken bost urteotan automatizatutako gailuek (CX eta Bot-ak) Wikipediako itzulpen-jarduerari nola eta zein neurritan aldatu duten ere ikertzen da. Elkarrizketetako erantzunen arabera, azken bost urteotan itzulpen-jarduerari azkartu eta zorrotz egin da, eta komunitate batzuek automatizazioarekiko konfiantza handiagoa lortu dute. Gainera, parte hartzaile askok gailuak perfektuak ez direla adierazi arren, gehiengo batek aitortu du onurak ekarri dizkiela beren itzulpen-jarduerari (Góngora-Goloubintseffen, 2021).

5. Postedizioa

Aurreko atalean aipatutako ikerketek esker ondorioztatu daiteke itzulpenak baduela tokirik entziklopedia askean. Baina ez soilik itzulpenak, baita postedizioak ere. Alegia, bai wikipediako kideek bai eta EWKEkoek ere, postedizioaren beharra nabarmentzen dute. Esate baterako, honako hau irakur daiteke EWKEko webgunean:

Itzulpen automatikoak jauzi ikaragarria egin du kalitatean azken bi urtetan, euskararako itzultzaile neuronalek ere bai (...), baina kontuz ibili behar da. Kalitate handikoa baldin bada ere, itzulpen automatikoaren emaitza gainbegiratu egin behar du pertsona batek, gero zuzendu egin behar da, “posteditatu” egin behar dela esaten dugu (Kepa Sarasola, d.g.).

IA itzulpen-prozesuaren parte denetik, itzultzaileen fokua sorburu-testua izatetik xede-testua izatera pasatu da. Izan ere, zenbait testuingurutan, itzultzaileek jada ez dute testurik itzultzen, azken horiek berrikusi edo egokitu baizik, hots, posteditatu. Hortaz, postedizioa giza itzultzaile batek IA sortutako itzulpen gordina zuzentzea da, gidalerro eta kalitate-irizpide batzuk oinarritzat hartuta (O’Brien, 2001).

Aipatzekoa da postedizioa itzulpena eta zuzenketarekin nahastu ohi dela. Hala ere, Iñurrieta eta Aranberrik (2021) diotenez, postedizioak baditu bere ezaugarri propioak. Alde batetik, postedizioaren kasuan, sorburu-testua IA bidez itzultzen da, eta zuzenketari dagokionez, berriz, giza itzulzaile batek egiten du itzulpen-lana. Horrenbestez, konpondu beharreko akatsak ere ezberdinak izango dira bi kasuetan. Bestalde, postedizioaren eta itzulpenaren arteko aldeari begira, lehenengoaren itzulpen-prozesuan hiru testu erabiltzen dira (itzulgaia, itzulpen automatikoa eta posteditatutako testua) eta bigarrenarenean, oster, bi (itzulgaia eta itzulpena); nahiz eta gehiago ere erabil daitezkeen (zubi-testuak, adibidez).

6. Helburua eta metodologia

GRAL honekin bi helburu lortu nahi dira: alde batetik, euskal Wikipedian itzulpen-politika formal edo informalik dagoen ikustea. Bigarrenik, Elia *Content Translation* gailuan integratu denetik, postedizioa zer nolakoa den aztertzea. Zehatzagoak izateko, Eliak egiten dituen akatsak zeintzuk diren identifikatu eta wikilariak akats horiek zenbateraino konpontzen dituzten aztertuko dugu.

Gongora-Goloubintseffen (2021) proposamena interesgarria eta berritzailea iruditu zitzaigunez gero, guk ere gauza bera egitea erabaki genuen, baina kasu honetan ikergaia euskal Wikipedia izanik. Lehenengo helburua betetzeko, beraz, euskal entziklopedia askeak eskaintzen dituen baliabideak kontsultatu genituen, hala nola: Txokoa eta eztabaida-orriak. Izan ere, gune horietan negoziatzen dira politika formal eta informalak (Wikipedia, 2018) eta, hortaz, informazio-iturri egokitzat jo genituen. Hala ere, wikilari guztien eztabaida-orriak aztertzeak denbora asko hartuko zigukeenez, erabiltzaile aktiboenak baino ez genituen ikertu. Horrez gain, Euskal Wikilarien Kultur Elkartearen webgunean jasotako informazioaz ere baliatu ginen, euskal Wikipediako erabiltzaileek osatutako elkarte Wikipedia sustatzeaz arduratzen delako eta, hortaz, pentsatu genuen posible zela bertan itzulpen jardueraren inguruko informazioa aurkitzea.

Horrez gain, bigarren helburuari begira, bi pauso eman genituen: alde batetik, akatsen identifikazioa eta, bestetik, postedizioen berrikusketa. Beraz, lehenik eta behin, Eliarekin itzuli diren eta dagoeneko argitaratuta dauden STak itzuli genituen *Content*

Translation erabiliz³. Ondoren, gauza bera egin genuen wikilariet posteditatutako dokumentuekin, hau da, Eliaren bitartez itzulitako eta posteditatutako artikuluak bildu eta Eliak emandako bertsioekin alderatu genituen.

Atal honetan, beraz, ikerketa aurrera eramateko jarraitutako metodologia azaltzen joango gara, hots, testuak hautatzerako orduan erabili genuen prozesuaren berri emango dugu, bai eta testuen azterketa egiteko prozedura argitu ere.

6.2. Euskal wikilarien lagina osatzeko irizpideak

Kontuan hartu behar da itzulpena Wikipediaren “materia beltza” (*dark matter*) dela, hau da, ezin daiteke jakin zein den Wikipedian gauzatzen den itzulpenaren portzentaje zehatza (Shuttleworth, 2017). Horrenbestez, aztertutako testuak benetan itzulpenak direla ziurtatzeko, wikilari aktiboen zerrendara jo genuen; zerrenda horretan azaltzen baitira, etiketa bidez markatuta, erabiltzaileek azken 30 egunetan burututako jarduera guztiak, baita itzulpenak ere: itzulpenari zuzendutako etiketak “Content Translation2” eta “EdukiItzulpena” dira. Lehenengoak adierazten du delako artikulua egiteko wikilariet *Content Translation* gailua erabili duela, eta bigarrenak, ostera, edukia itzulpenaren bitartez sortu dela. Hori egiteari esker, testu posibleen kopurua murriztea lortu bagenuen ere, hautagaiak asko izaten jarraitzen zuten oraindik. Horrenbestez, prozesua pixka bat arintzeko, kide aktibo batzuk hautatu eta horiet egindako itzulpenak erabiltzea erabaki genuen.

Azterketarako pertsona-kopurua xedatzeko, bi azterlan hartu genituen oinarritzat: Góngora-Goloubintseffen tesia (2021) eta Solaunen Gradu Amaierako Lana (2021). Lehenengoak hamasei erabiltzaile elkarrizketatu zituen, eta bigarrenak, berriz, zazpi. Aukeraketa horren atzean arrazoi ezberdinak egon daitezke, horietako bat ikerketaren luzera izanik. Hau da, kontuan hartu behar da Góngora-Goloubintseffena tesi bat dela eta Solaunena GRALA. Gure azterketari begira, beraz, zortzi pertsona hautestea erabaki genuen. Zenbaki hori aproposa iruditu zitzaigun, Wikipedian diharduten profil ezberdinen jarduera aztertzeko aukera izango genukeelako eta, aldi

³ Bigarren helburua lortzeko, hasieran, posteditzio-proba bat egitea pentsatu genuen euskal Wikipediako erabiltzaileen laguntzarekin. Baina, azkenean, ideia hori baztertu egin behar izan genuen. Hori horrela, bururatu zitzaigun soluzioa jada argitaratuta zeuden testuen STak Eliaren bitartez berriz itzultzea izan zen; nahiz eta, agian, erabiltzaileek eskuratutako XTa eta gurea guztiz berdinak ez izan.

berean, lana ez zelako gehiegi luzatuko. Wikipediako erabiltzaile guztien artean zortzi hautesteko, honako hiru irizpide hauek erabili genituen:

- A) Euskal Wikipediako erabiltzaile aktiboa izatea. Asko dira euskal entziklopedia askeko komunitatea osatzen duten kideak. Halere, wikilari aktiboak dira plataforman inpaktu esanguratsuenaren dutenak; maiztasun handiagoz parte hartzen dutelako. Hori horrela, euskal Wikipediako edukien zati handi bat erabiltzaile horien artean sortutakoa dela aintzat hartuta, ezin uka daiteke beraien itzulpenak direla entziklopediaren errealitatea hobean islatzen dutenak.
- B) *Content Translation* tresna gutxienez 6 artikulu sortzeko erabili izana. Horrela, gailua ezaguna egiten zaiela eta aspektu teknikoek postedizioan eraginik ez dutela ziurtatu genuen. Erabiltzaileek irizpide hori betetzen dutela egiaztatzeko, lehen aipatu bezala, wikilari aktiboen zerrendara jo genuen eta, gero, lankide bakoitzaren izenaren ondoan dagoen “Ekarpenak” deituriko esteka sakatu. Ondoren, “Content Translation2” eta “EdukiItzulpena” etiketak zituzten sarrerak identifikatu genituen, gutxienez sei artikulu sortzeko erabili zutela baieztatu arte.
- C) Sorburu-hizkuntza eta xede-hizkuntza ondo menderatzea ere garrantzitsua dela generitzon, baina ez ezinbestekoa. Izan ere, hizkuntza-mailak postedizioa baldintzatu dezake, baina gogorarazi beharra dago erabiltzaileak anonimoak direla eta ez dutela beraien informazio pertsonala partekatzeko beharrik. Beraz, kasu batzuetan, ezinezkoa izango da wikilarien hizkuntza-maila ezagutzea. Beste batzuen lankide-orrietan, berriz, behean txertatutako irudian azaltzen den taula ikusi ahal izango da. Taularen ezkerreko zutabean hizkuntzen laburdurak azaltzen dira zenbaki edo letra bati marratxo baten bitartez lotuta. Eskuineko zutabean, berriz, ikur, zenbaki eta letra horien azalpenak daude, hots, lankideak hizkuntza bakoitzean duen maila deskribatzen da.

Irudia 1: Lankideen babel informazioa.

Lankideen babel informazioa	
eu-N	Erabiltzaile hau euskaldun zaharra da (euskara du ama hizkuntza).
es-N	Esta persona tiene una comprensión nativa del español.
fr-5	Cet utilisateur dispose de connaissances professionnelles en français.
en-3	This user has advanced knowledge of English.
it-3	Questo utente può contribuire con un livello avanzato in italiano.
de-2	Dieser Benutzer beherrscht Deutsch auf fortgeschrittenem Niveau.

Behin baldintzak finkatuta, hautaketari ekin genion. Horretarako, Wikipediak guztion eskura jartzen duen wikilari aktiboen zerrenda kontsultatu eta Excel bat sortu genuen une hartan zerrendatuta zeuden erabiltzaileekin. Kontuan hartu behar da Wikipedia ez dela entziklopedia finko bat, hau da, aldatuz doa. Hortaz, posible da gaur egun wikilari aktiboen zerrendan dauden erabiltzaileak ez izatea dokumentua sortu genuen garaian zeuden berdinak.

Excel horretan kideek beraien lankide-orrietan idatzitakoa biltzeaz gain (informazio pertsonala: jaioterria, adina, lanbidea, etab), euskal Wikipediari egindako ekarpenak eta eztabaida-orrietan itzulpenari buruz esandakoak ere jaso genituen. Hori egiteari esker, wikilariak bahetu eta zortzi hautagaien zerrenda osatu genuen.

6.3.Itzulgaien aukeraketa

Itzulgaiak aukeratzeko, lehenik eta behin, irizpide batzuk finkatu genituen. Hasteko, testu kopurua zein izango zen erabaki genuen. Solaunek zazpi testu erabili zituen, hain zuzen, proban parte hartu zuen pertsona bakoitzak dokumentu bana posteditatu behar izan baitzuen. Gure kasuan, helburuetako bat Wikipediako posteditazioa

nolako den jakitea zenez gero, zortzi testu gutxiegi iruditzen zitzaigun eta, beraz, kopurua bikoiztea erabaki genuen.

Bigarrenik, Wikipedian neurri askotako artikulua daudela ikusirik, guk erdibideko tamainakoak aukeratzea erabaki genuen, hala nola, 600-1000 hitz inguruko sarrerak. Alegia, ez genituen testu laburrak aukeratu, Wikipediak ondo garatutako edukiak lehenesten dituelako, edo beste modu batera esanda, sarrera laburrak bazterten dituelako. Testu luzeegi batek, oster, denbora larregi hartuko liguke. Dena den, zenbait kasutan ez genuen lortu hitz-tarte horren barruan sartzen diren testuak aurkitzerik eta, beraz, bi itzulpenen orde, hiru hautetsi genituen. Gogoratu dezagun testuen hautaketa-prozesuaren lehen urratsa erabiltzaileak aukeratzea izan zela eta, horrenbestez, ezin genuen sarreren luzera aurreikusi. Horregatik, batzuetan hiru testu aukeratu genituen, kasu guztietan aztertu beharreko itzulpenen bolumena antzekoa izan zedin.

Bestalde, espezializazio-maila ere kontuan hartu behar genuen. Izan ere, Wikipedian lan egiten duten erabiltzaileak esparru ezberdinetan espezializatutakoak izaten dira eta arlo horretako testuak itzultzen edo posteditatzen dituzte batik bat (Góngora-Goloubintseff, 2021). Horregatik, zaila zenez jakintza-arlo bereko testuak aurkitzea, esparru ezberdinetakoak erabiltzea xedatu genuen. Era berean, baztertu egin genituen orokorregiak ziren dokumentuak, Wikipedia entziklopedia bat den neurrian, ezinezkoa baitzen teknizismoak saihestea (Solaun, 2021).

Aintzat hartu beharreko azken irizpidea sorburu-hizkuntza izan zen. Hautatutako zortzi wikilariak sorburu-hizkuntza ezberdinetatik abiatzen direnez gero, hiru hizkuntza hauetatik itzultitako testuak hautetsi genituen: ingelesa, frantsesa eta gaztelania. Erabaki hori Eliaren eta nire gaitasun-mailara egokitu genuen, tankera honetako azterketa bat aurrera eramateko ezinbestekoa baita jatorrizko testua eta xede-testua alderatzea eta, hortaz, bi hizkuntzak menderatu behar dira. Bestalde, Eliak sei hizkuntzetatik, eta hizkuntzetara, itzultzen duenez gero, sorburu-hizkuntza horietako bat izatea ezinbestekoa zen.

Laburbilduz, azterketa honetarako 600-1000 hitz inguruko 18 testu hautatu genituen⁴, guztiak gai ezberdin bati buruzkoak, hala nola politika, ekonomia, energia, biografiak, eta itzulpen guztien STak ingeles, frantsez edo espainolez idatzitakoak dira.

6.4. Postedizio-azterketaren prozedura

Bigarren helburua lortzeko, hau da, Wikipediako kideek gauzatzen duten PEA zenbaterainokoa den jakiteko, Sycz-Opón eta Galuskinaren (2017) eta Comparin eta Mendesen (2017) ikerketak hartu genituen oinarritzat, eta gure azterketaren ezaugarrietara egokitu. Hala, prozesua bi zatitan banatu genuen: IAK emandako akatsen anotazioa eta posteditutako testuen analisia.

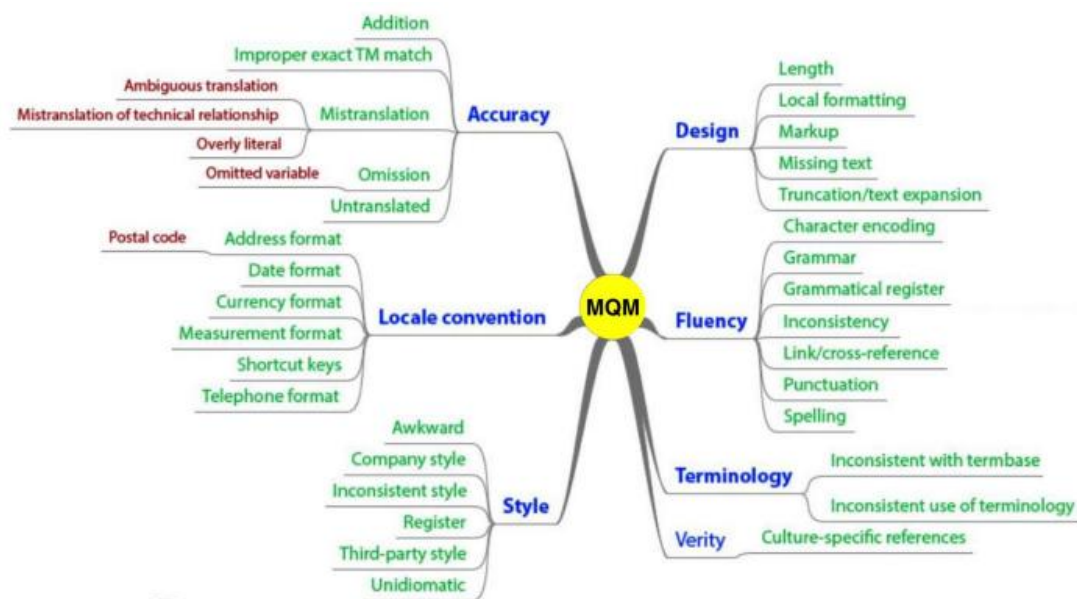
6.4.1. Akatsen identifikazioa: IAK emandako itzulpenetako akatsen anotazioa.

Akats-tipologia zehazteko Multimedia Quality Metric (MQM) deritzon sailkapena hartu genuen eredutzat. Azken hori IA bidez egindako itzulpenen kalitatea neurtzeko erabiltzen den eredu bat da. Quality Translation 21 (QT21) proiektuak garatu zuen, Europar Batasuneko *Horizon 2020* ikerkuntza eta berrikuntza proiektuaren finantzaketarekin (Comparin eta Mendes, 2017).

Beheko irudian ikus daitekeenez, MQM ereduak honako zazpi kategoria nagusi hauek bereizten ditu: doitasuna, diseinua, tokiko konbentzioak, jarioasuna, estiloa, terminologia eta egia. Kategoria nagusi bakoitzaren barnean, gainera, beste hainbat azpikategoria daude, gramatika, erregistroa eta omisioa, besteak beste.

⁴ Itzulgaien zerrenda eranskinen atalean aurki daiteke, Wikipediako artikulua kontsultatzeko estekekin batera.

Irudia 2: MQM taxotomia.



Akatsak identifikatzeko sailkapen horretan (2. irudia) oinarritu ginen arren, ez genion zehatz-mehatz jarraitu, bertan kategoria gisa azaltzen diren aspektuak guk azpikategoria bihurtu genituelako eta beste batzuk, ostera, ez genituelako kontuan hartu. Hona hemen erroreak sailkatzeko guk erabilitako kategoriak eta azpikategoriak:

- Doitasuna⁵
 - Gehikuntza: beharrezkoak ez diren elementuak gehitu ditu.
 - Gehiegizko itzulpena: itzuli behar ez dena itzuli du.
 - Itzulpen falta: itzuli gabeko hitzak/segmentuak daude.
 - Omisioa: testuan elementuak falta dira.
 - Segmentua gaizki ulertu du, hots, itzulpen txarra egin du⁶.
- Jariotasuna
 - Gramatika: deklinabide okerra, ergatiboaren erabilera txarra, aditz-denbora okerra, pertsona gramatikal okerra, subjektu aldaketa, artikulua determinatu bat indeterminatu batekin nahastea (eta alderantziz).

⁵ Gogora dezagun Eliak zerbait itzuli gabe uztea akats gisa hartu beharko litzatekeela, printzipioz, guztia itzuli beharko lukeelako. Baina, wikilarien kasuan, zerbait gehitu edo kentzea ez da errore kontsideratzen. Izan ere, Wikipediaren arabera, wikilarietako testuak osorik edo zatika itzul ditzakete. Beraz, omisioak eta aplikazioak ohikoak izango dira.

⁶ Gaizki-ulertuen ondorioz egindako akatsak ez dira zenbatuko. Adibidez, gaizki itzulitako segmentu bat zuzentzeko deklinabidea edo hitzen bat aldatu behar badira, aldaketa horiek ez dira akats gramatikal gisa ulertuko; sorburu-testua behar bezala ulertuko balu, ez litzatekeelako horrelako erroreak egingo.

- Ortografia.
- Puntuazioa.
- Sintaxia: perpaus okerra, juntagailu okerra, perpausa osatzen duten elementuen hurrenkera okerra.
- Terminologia.
- Estiloa⁷
 - Jatorrizko egituraren kalkoa: euskarazko segmentuari karaktere arrotza ematen dion egituraketa.
 - Errepikakorra.
- Tokiko konbentzioak
 - Ortotipografia.

Horrez gain, aipatzekoa da IAren kalitatea neurtzeko ereduak errore bakoitzari larritasun-maila bat esleitzen diotela, hala nola arina, larria eta kritikoa. Halaber, guk ez genuen azken hori kontuan hartu, gure helburua Eliak zein errore egiten dituen ikustea zelako, gero wikilariak horiek konpontzen edo mantentzen dituzten jakiteko. Laburbilduz, egindako akats-mota bakoitza identifikatu genuen eta akats bera behin baino gehiagotan errepikatu arren, guztiak zenbatu genituen; azken finean, wikilariak identifikatutako errore guztiak konpondu behar dituztelako.

6.4.2. Posteditatutako testuen analisisia: zuzenketa fasea.

Posteditatutako testuen analisirako, ostera, Sycz-Opón eta Gałuskina (2017) erabilitako sailkapena hartu genuen oinarritzat. Honako bi kategoria hauek berezi genituen:

- IAren akatsak zuzendu.
- IAren akatsak mantendu.
- Akatsak gehitu.

⁷ Sarritan, estiloa PE sakonetik kanpo geratzen da. Halaber, guk aztertutako testuen sorburu-hizkuntzak hain dira ezberdinak, euskararekin alderatuta, ezen beharrezkotzat jotzen baitugu kategoria hau aintzat hartzea. Hain zuzen, Eliak jatorrizko hizkuntzari gertutik jarraitzeko tendentzia baitauka, euskararen ohiko joskera bazter batera utziz.

Sycz-Opón eta Gáuskinaren lanean kategoria bat gehiago azaltzen da, hala nola “beharrezkoak ez diren aldaketak”. Halere, guk erabaki genuen horiek baztertu egingo genituela, wikilariak askatasun handia dutelako, fideltasunari dagokionez, eta denbora ez delako faktore garrantzitsu bat.

7. Emaitzak

Atal honetan, aldez aurre ezarritako bi helburuen emaitzak plazaratuko ditugu. Lehenik eta behin, euskal Wikipedian itzulpenarako politikarik dagoen azalduko dugu. Ondoren, egindako azterketaren arabera, Wikipedian zer nolako postedizioa gauzatzen den helaraziko dugu.

7.2. Euskal Wikipediako itzulpen-politikak

Wikipedia entziklopedia aske bat izanik, edonork itxuraldatu dezake: wikilariak artikulua berriak sor ditzakete, zaharrak berrikusi, akatsak konpondu eta abar. Dena den, erabiltzaileen esku-hartzeek plataforma kaltetzen ez dutela ziurtatzeko, Wikipediako komunitate guztiek oinarritzko bost arau partekatzen dituzte (ikus 3. atala).

Bost arau horiez gain, komunitate bakoitzak bere politika propioak garatzen ditu unean uneko eragozpenak konpontzeko. Lehen esan bezala, politika horiek erabiltzaileen arteko negoziazioen bitartez adosten dira horretarako egokitutako guneetan (Góngora-Goloubintseffek, 2021). Aipatzekoa da euskal Wikipedian eztabaida-orriek gain Txokoa deituriko gunea ere badagoela. Azken hori Wikipediari buruzko kontu orokorrean hitz egiteko erabiltzen da; artikulua jakin bati buruzko iruzkinak egiteko artikulua beraren edo erabiltzaileen eztabaida-orrietara jo baitaiteke. Eztabaida-orriak eta Txokoa aztertzen aritu ostean, ez dugu itzulpenari buruzko gidalerrorik aurkitu. Alegia, gaur gaurkoz, wikilariak itzultzaile automatikoari lotutako kontuez baino ez dira mintzatu: besteak beste, gailuaren erabilera bultzatzeko eta azken horrekin izandako arazo berri emateko.

Dena den, bi orri horietan euskaraz idatzitako itzulpen-politikarik aurkitu ez izanak ez du esan nahi euskal Wikipedian itzulpenaz baliatzen direnek ez dutela inolako

laguntzarik jasotzen. Izan ere, Wikipediako estilo-gidan itzulpen arazo posible batzuk argitzen dira, hala nola: liburuen, artelanen, filmen eta abarren izenburuak euskaraturik ez badaude, jatorrizko bertsioa mantenduko da eta itzulita egonez gero, itzulpen ofiziala erabili beharko da. Euskal baliokiderik ez duten erdal hitzekin gauza bera gertatzen da, hots, erabiltzaileak jatorrizko bertsioa erabili behar dute, baina letra etzanez idatzita. Eta atzerriko termino baten itzulpen edo transliterazioa egin bada, jatorrizko hizkuntzan idatzitako terminoa ere gehituko da jarraian, adibidez, “Abu Ali al-Hasan Ibn al-Haitham (arabieraz: *الهيثم بن الحسن عليّ أبو و*)” (Wikipedia, 2021). Horrez gain, orokorrean, konpainia, elkarte eta erakundeen izenak soilik beharrezkoa bada itzuliko dira eta izenak forma deskriptiboa badauka (National Health Service), norberak aukeratu beharko du zer egin. Aitzitik, horien akronimoak euskaraz zabalkunde nahikorik ez badute, jatorrizko hizkuntzan sortutakoak erabiliko dira. Gainera, bi instituzio ezberdinen izenek antzekoa edo berdina den itzulpen bat badute, ez dira euskaratuko. Bestalde, artikuluetan euskara ez den beste hizkuntza batean idatzitako aipuak sartu nahi baldin badira, itzuli egin beharko dira. Azkenik, atzerriko toponimoak ez dira itzuliko, Euskaltzaindiak kontrakorik gomendatzen ez badu.

Beheko irudian ikus daitekeenez, *Content Translation* gailuak itzulpen egoki bat gauzatzeko argibide minimo batzuk ematen ditu. Are gehiago, irudiaren azken zutabeko “Ikus itzulpen gidalerroak” izeneko estekak ingelesez idatzitako politika informaletara⁸ garamatza. Halaber, azken sarrera horrek ez du itzulpenari buruzko argibiderik ematen eta, gainera, ingeles Wikipediako edukiak sustatzera zuzenduta dago.

Irudia 3: Itzulpenerako argibideak *Content Translation*en



⁸ Goian txertatutako irudian ikus daitekeenez, *Content Translation* gailuak “gidalerroak” hitza darabil *guidelines* terminoaren itzulpen gisa. Euskal Wikipediak, ostera, politika informal terminoa baliatzen du politikei zuzendutako sarreran (Wikipedia: politikak). Beraz, nahasmenduak saihesteko, guk politika hitza erabili dugu kasu guztietan.

Euskal Wikilarien Kultur Elkartearen, ostera, *Content Translation* gailua eta Elia behar bezala erabiltzeko jarraibideak aurki daitezke, baina ez da itzulpen-politikei buruz ezer aipatzen. Elia Elhuyarrek garatutako itzultzaile automatiko neuronala da, eta itzulpenak sei hizkuntzaren (euskara, gaztelania, frantsesa, ingelesa, katalana eta galegoa) artean egin ditzake. Gainera, 2021. urtetik aurrera Elia Wikipediako edukiak itzultzeko erabil daiteke Wikimedia Foundation eta Elhuyarren arteko hitzarmenari esker (Elhuyar, 2021).

7.3. Postedizio-azterketaren emaitzak

7.3.1. Akatsen identifikazioa: IAk emandako itzulpenetako akatsen anotazioa.

Elia eta *Content Translation* plataformaren bitartez sortutako itzulpenak aztertu ostean, zenbatutako emaitzak honako taula honetan jaso ditugu:

Taula 1: Eliak egindako akatsak (balore absolutuak)

TESTUAK ⁹	DOITASUNA	JARIOTASUNA	ESTILOA	TOKIKO KONBENTZIOAK	AKATSAK GUZTIRA
1. TESTUA	33	63	15	1	112
2. TESTUA	20	37	9	0	66
3. TESTUA	12	36	7	2	57
4. TESTUA	21	38	8	7	74
5. TESTUA	23	61	21	2	107
6. TESTUA	9	36	12	2	59
7. TESTUA	7	17	6	1	31
8. TESTUA	38	52	5	1	96
9. TESTUA	30	30	6	3	69
10. TESTUA	7	37	9	2	55
11. TESTUA	32	88	13	1	134
12. TESTUA	21	41	9	6	77
13. TESTUA	15	36	5	6	62
14. TESTUA	13	28	5	3	49
15. TESTUA	24	32	6	2	64
16. TESTUA	17	20	3	5	45
17. TESTUA	17	37	4	1	59
18. TESTUA	6	16	6	0	28

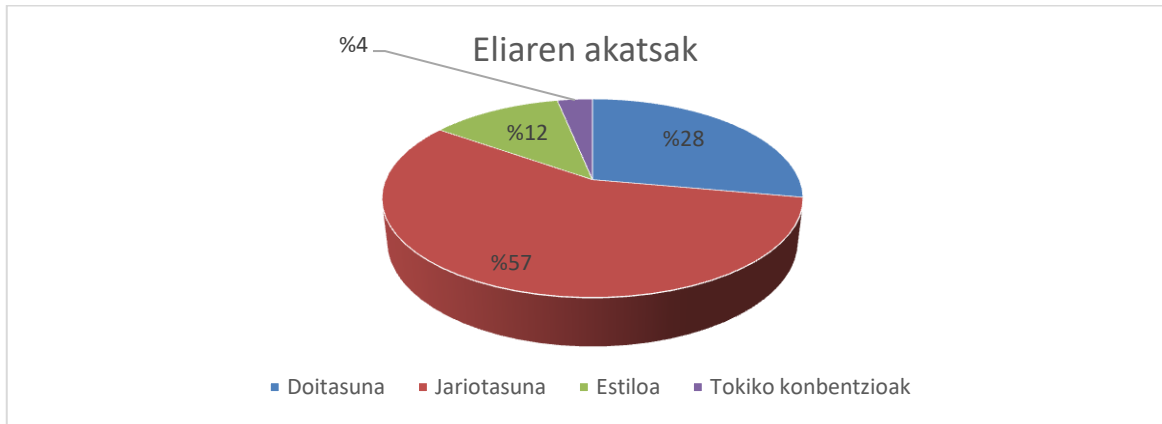
⁹ Eranskinen atalean daude testuen izenburuak eta euskal Wikipedian argitaratutako bertsioen estekak.

Goiko taulan bereiz daitekeenez, kategoria bakoitzean bildutako akatsen eta testuen artean ez da joerarik nabari. Testu batzuetan akatsen banaketa antzekoa izan arren (adibidez, doitasunari dagokionez, 16. testuak eta 17.-ak 17 akats dituzte), beste batzuetan alde nabarmenak daude. Doitasunaren kategoriara bueltatuz, oro har, akatskopurua testuko 6tik 38ra bitartekoa da. Bestalde, jariotasunaren zutabearen ere egoera antzekoa dela esan daiteke. Izan ere, erroreak 16tik 88ra artean dabilta. Azpimarratzekoa da, doitasunarekin alderatuz, testu batzuen eta besteen arteko erroreen aldea handiagoa dela. Hori horrela da, akats bera behin baino gehiagotan errepikatu arren, guztiak zenbatu ditugulako. Horrenbestez, 11. testuak 88 akats ditu IAk aditz-denbora okerra erabili duelako testu osoan zehar eta, hortaz, gaizki jokaturako aditz beste akats zenbatu ditugu. Estiloaren eta ortotipografiaren kasuan, berriz, egoera pixka bat aldatzen da: alde batetik, testu gehienek 3-9 estilo-akats inguru dituzte; eta, bestetik, tokiko konbentzioen esparruan erroreak 1etik 7ra artekoak dira, hau da, beste bi kategoriekin konparatuta aldea jada ez da hain handia.

Horrez gain, hiru grafiko ere txertatu ditugu IAk egindako itzulpenetan identifikatu ditugun akatsen ikuspegi orokor bat izateko. Grafikoetan, zenbatutako akats guztietatik kategoria bakoitzak zein ehuneko hartzen duen ikus daiteke, bai eta maila nagusi bakoitzaren barruan errore gehien eragiten dituzten azpi-kategoriak zeintzuk diren ere.

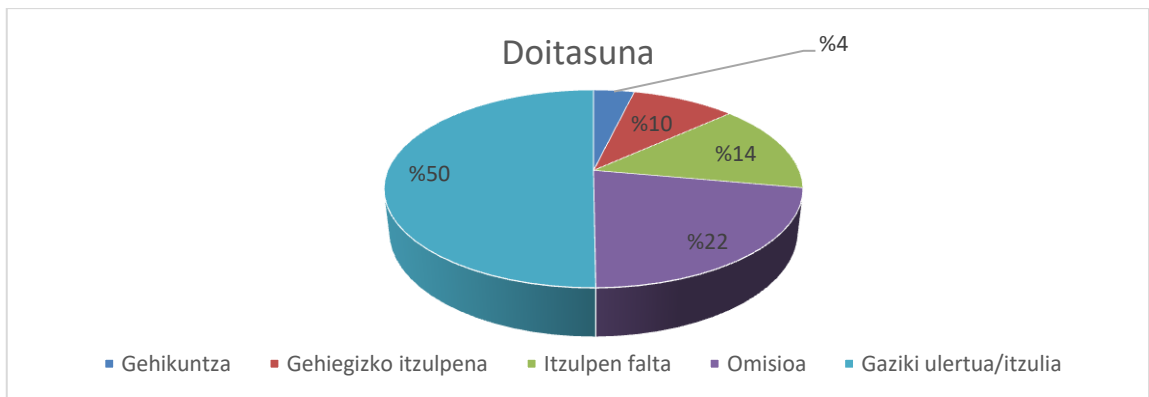
4. grafikoan ikus daitekeenez, aztertutako lau kategoria nagusien arteko aldeak nabarmenak dira; batik bat, doitasunaren eta jariotasunaren artean. Elia eta *Content Translation* gailuaren bitartez itzulitako dokumentuetan aurkitu ditugun akats guztien erdia baino gehiago (% 57) jariotasunari lotutakoak dira. Zerrendan hurrengoa doitasuna da, lagin osoaren % 28 hartzen duelarik eta, ondoren, estiloa % 12rekin. Azkenik, tokiko konbentzioak bereiz daitezke, gure kasuan, ortotipografia, % 4arekin.

Irudia 4: Eliaren akatsak guztira



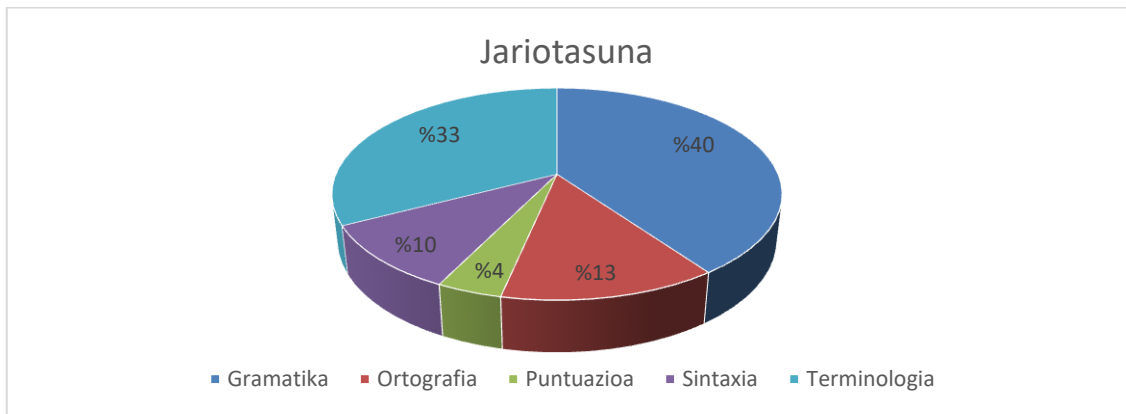
Gainera, errore-kopuru handieneko kategorien baitan bildutako azpikategorien artean ere desberdintasunak daude (5. eta 6. irudiak). Alde batetik, doitasunari dagokionez, gaizki ulertu edo itzulitako segmentuak izan dira arazo gehien sortu dituztenak (% 50). Omisioek, bestalde, errore guztien % 22 hartzen dute, eta itzuli gabeko egiturek % 14. Zerrendan azkenak, gehiegizko itzulpenak eta gehikuntzak dira, % 10 eta % 4arekin, hurrenez hurren.

Irudia 5: Eliaren akatsak: doitasuna



Bestalde, jarriotasunari lotutako akatsak jasotzen dituen grafikoan (6. irudia) ikus daitekeenez, arazo gehien eragin dituen azpikategoria gramatika izan da (% 40). Gramatikari gertutik jarraitzen dio terminologiak, akats guztien % 33rekin. Jarraian, % 13arekin, ortografia dago eta, % 10arekin, sintaxia. Amaitzeko, errore gutxien bildu dituen azpikategoria puntuazioa izan da (% 4).

Irudia 6: Eliaren akatsak: jariotasuna



7.3.2. Posteditatutako testuen analisisa: zuzenketa fasea.

7.3.2.1. Azterketa kuantitatiboa

Euskal Wikipediako erabiltzaileek posteditatutako testuen azterketari dagokionez, sailkapena hiru maila nagusitan banatu dugu, hala nola konpondutako akatsak, mantendutako erroreak eta gehitutakoak. Gehitutako eta mantendutako akatsak antolatzeko behean txertatutako taula (2. taula) erabili dugun arren, zuzenketak ez ditugu bertan jaso.

Taula 2: Wikilarien postedizioak (balore absolutuak)

TESTUAK	DOITASUNA		JARIOTASUNA		ESTILOA		ORTOTIPOGRAFIA	
	G. ¹⁰	M. ¹¹	G.	M.	G.	M.	G.	M.
1. TESTUA	1	4	6	18	0	2	10	0
2. TESTUA	7	5	24	5	4	0	1	0
3. TESTUA	2	0	19	8	1	1	1	2
4. TESTUA	3	0	19	4	0	2	0	7
5. TESTUA	6	1	26	8	2	3	0	0
6. TESTUA	7	1	10	10	3	4	0	0
7. TESTUA	2	2	2	13	1	2	0	1
8. TESTUA	6	14	15	27	1	3	0	0
9. TESTUA	2	3	7	9	2	0	2	0
10. TESTUA	6	0	19	7	3	1	1	1
11. TESTUA	0	4	9	6	1	1	6	0
12. TESTUA	2	1	10	7	4	2	1	4
13. TESTUA	2	9	8	14	1	5	0	6
14. TESTUA	1	4	3	17	0	3	0	2

¹⁰ G= Gehitutako akatsak.

¹¹ M= Mantendutako akatsak.

15. TESTUA	1	0	8	2	1	1	22	2
16. TESTUA	2	2	3	1	0	1	14	1
17. TESTUA	3	13	4	13	0	1	1	0
18. TESTUA	3	0	10	2	1	2	0	0

Erabiltzaileek egindako postedizioak aztertu eta emaitzak jaso ostean (2. taula), esan daiteke wikilari guztiek akatsak gehitzeko joera dutela; kopurua gorabehera, ia kategoria guztietan daude lehen ez zeuden erroreak. Bestetik, oro har, testuetan identifikatutako akatsen banaketa nahiko antzekoa da (jariotasunaren kasuan izan ezik). Lehenik eta behin, mantendutako akatsei erreparatuko diegu. Doitasunari dagokionez, errore-kopurua testuko Otik 14ra artekoa da, eta 0-5 artekoa, berriz, estilo-akats kopurua. Eta tokiko konbentzioei lotutako akatsak Otik 7ra artean dabilta. Bestetik, gehitutako erroreen kasuan, doitasunari lotutako akats-kopurua Otik 7ra doa eta akats estilistikoena Otik 4ra. Bukatzeko, ortotipografia akatsak 0-14 artean dabilta, baina 15. testua posteditatu duen wikilariak besteek baino dezente errore gehiago gehitu ditu. Hala, mantendutako akatsak guztira 295 dira, gehitutakoak 342 eta zuzendutakoak, ostera, 949.

Horrez gain, oraingoan ere posteditatutako testuen ikuspegi orokor bat eskaini nahi dugunez gero, hiru grafiko sortu ditugu. Lehenengoan, wikilarien postedizioetan identifikatutako aldaketa guztiak bildu ditugu, eta beste biekkin, ostera, mantendutako zein gehitutako akatsen artean bolumen handiena zein kategoriak duen erakutsi nahi dugu.

Hala, wikilarien postedizio guztien % 76 zuzenketek osatzen dute, hots, erdia baino gehiago. Bigarren postuan, aldaketen % 24arekin, mantendutako akatsak daude, hau da, wikilariak identifikatu gabe utzi dituztenak.

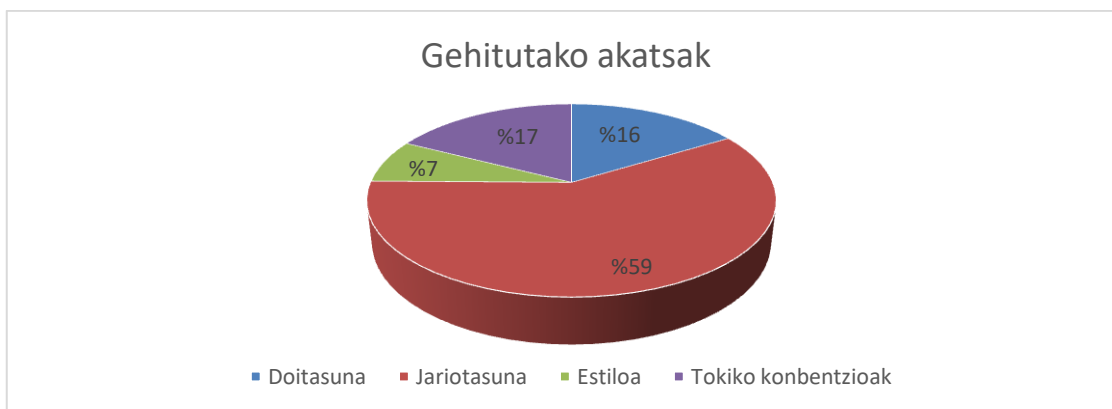
Irudia 7: Wikilarien postedizioak guztira



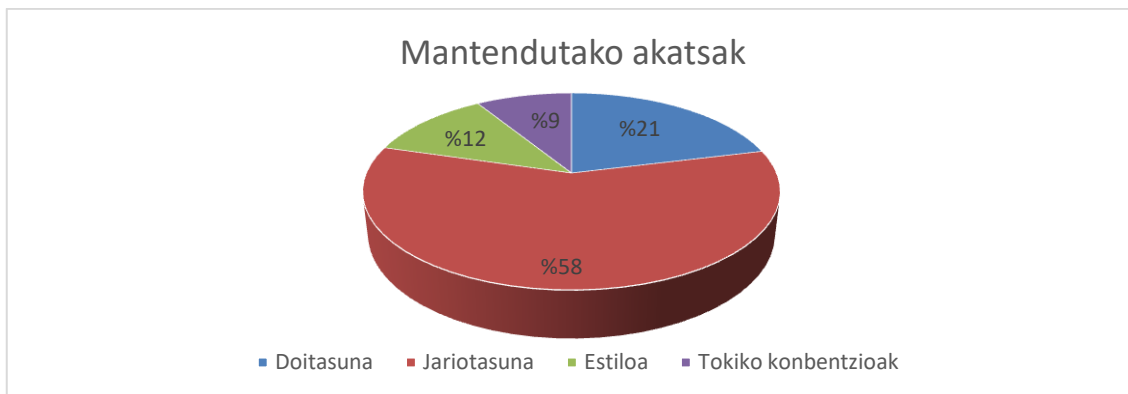
Gehitutako eta zuzendu gabe geratu diren akatsak, bestalde, aurreko kategoriak erabiliz (doitasuna, jarioetasuna, estiloa eta tokiko konbentzioak) sailkatuz gero, emaitzak nahiko parekoak dira (ikus 8. eta 9. irudiak).

Izan ere, gehikuntzen eta mantendutako akatsen kasuan, jarioetasuna izan da arazo gehien eragin dituen: % 59 eta % 58rekin, hurrenez hurren. Bigarrenik, gehikuntzei dagokienez, tokiko konbentzioei lotutako akatsak nagusitu dira (% 17), eta mantendutako akatsen kasuan, doitasuna (% 21). 8. eta 9. irudietan ikus daitekeenez, gehitutako akatsen artean doitasuna da hirugarrena zerrendan, % 16arekin, eta konpondu gabekoetan, ostera, estiloa (% 12). Amaitzeko, estiloak du puntuazio baxuena (% 7) gehitutako akatsetan, eta tokiko konbentzioek (% 9) bigarrenaren kasuan.

Irudia 8: Wikilarien postedizioak: gehitutako akatsak



Irudia 9: Wikilarien postedizioak: mantendutako akatsak



7.3.2.2. Azterketa kualitatiboa

- **Gehitutako akatsak**

Arestian esan bezala, jariotasuna edo hizkuntzaren alderdi gramatikalak dira arazo gehien eragiten dituztenak posteditatzerako orduan, baina lehen azaldutako grafikoetan ikus daitekeenez, badaude baita kontuan hartu beharreko beste akats batzuk ere. Beraz, jarraian, kategoria bakoitzean aurkitu ditugun akatsetariko batzuk zerrendatuko ditugu.

Lehenik eta behin, gehitutako akatsak azalduko ditut. Jariotasunaren kategorian sartu ditugun akatsen artean, akats ortografikoak daude (3. taula). Harrigarria dirudien arren, testu batean baino gehiagotan aurkitu ditugu akats ortografikoak, nahiz eta gaur egungo testu-prozesatzaile gehienek zuzentzailea txertatuta izan. Doitasunarekin jarraituz, omisioak eta gehiegizko itzulpenak aurkitu izan ditugun arren, arazo gehienak ulermenari lotutakoak izan dira (4. taula); wikilariet sorburu-testuaren zentzua aldatzearen ondorioa. Estilo akatsak izan dira gutxien gehitu direnak, oro har, mantendu edo zuzendu egin direlako. Halere, horrek ez du esan nahi akats estilistikorik ez dagoenik, horren erakusle da seigarren taulan jasotako adibidea. Azkenik, akats ortotipografikoak ere identifikatu ditugu, zeintzuen artean ohikoenak erdal elkarte, mugimendu, sari eta antzekoen idazkerari lotutakoak baitiren (5. taula). Dena den, kontuan hartu behar da wikilariet ez dutela zertan ortotipografia araurik jakin; eta, are gehiago, Wikipediako estilo-gidan ere ez da halako kasuetarako aholkurik eskaintzen. Beraz, informazio-hutsune hori bete beharko litzateke.

Taula 3: Gehitutako akatsa (jariotasuna)

6. TESTUA	IA Beheko lerroak jartzean edo giza jarduerak (adib., errepideak, eraikinak) gauzatzen diren eremuetatik hurbil kokatzean, elektrokutatzeiko arriskua areagotzen da.	PE Leineak lur-azaletik edo giza-jardueretatik hurbil kokatzean pertsonak elektrokutatzeiko arriskua areagotzen da.
--------------	--	---

Arestian aipatu bezala, wikilariek batzuetan arazo terminologikoak identifikatzea lortu arren, zuzenketak egiterako orduan akats ortografikoak gehitzen dituzte. Goiko adibidea kasu horietako bat da. Lankideak IAk erabilitako “lerroak” terminoa zuzendu behar zuen, testuinguru horretan erabiltzen den hitz zehatza “linea” delako, baina hori egitean lehen ez zegoen akats ortografiko bat egin du.

Taula 4: Gehitutako akatsa (doitasuna)

3. TESTUA	IA Ofizialtasunaren aldeko mugimendu deitzen zaio Asturiasko gizarte-mugimenduari edo gizarte-sareari, autonomia-estatutuak asturiarraren izaera ofiziala jaso dezan eskatzeko ; berezko hizkuntza duten beste autonomia-erkidego batzuetan gertatzen den bezala.	PE <u>Asturiaseko estatutuan</u> bertako hizkuntzaren lege-onarpenaren alde borrokatzen duen gizarte mugimendu edo sarea da.
--------------	---	---

Beste adibide honetan, zentzua pixka bat aldatu da. Izan ere, sorburu-testuak ez dio autonomia estatutuaren barnean hizkuntzaren lege-onarpenaren alde borrokatzen duen mugimendu bat denik, autonomia estatutuan hizkuntzaren karaktere ofiziala lortu nahi duen mugimendu bat baizik. Eta goiko zuzenketa horretan badirudi mugimendua estatutuaren parte dela, baina ez da hala.

Taula 5: Gehitutako akatsa (tokiko konbentzioak)

15. TESTUA	IA	PE
	2019an, Abebek AAren inguruan lan egiten duen Osasunaren Institutu Nazional lantaldean jardun zuen.	2019an zehar, Abebe " <i>National Institutes of Health (NIH)</i> "-ean Adimen Artifizialeko lantaldeko kide izan zen hainbat aditurekin batera

Akats ortotipografikoekin jarraituz, wikilariak zentroaren jatorrizko izena mantentzea erabaki du. Halaber, hori egitean, lehen ez zegoen akats ortotipografiko bat gehitu du. Izan ere, erdal hitz bat testuan txertatzean, letra etzana erabiltzen da eta ez, ostera, kakotxak. Beraz, kasu honetan, azken horiek soberan daude.

Taula 6: Gehitutako akatsa (estiloa)

2. TESTUA	IA	PE
	Kolonia ugari sortu zituen herri horrek, Cartagok bezala, Afrikako iparraldean.	Afrikako iparraldean kolonia ugariaren sortzailea izan zen, (...).

Azkenik, estiloari dagokionez, argitaratutako bertsioan (posteditatutakoan) aldaketa bat nabari da. Egia da ez dela akats bat, esaldia gramatikalki zuzena delako, baina Eliaren itzulpena hobea da. Izan ere, euskaraz nominalizazioak baztertzekoak dira, hots, aditzak hobesten dira.

- **Mantendutako akatsak**

Mantendu diren akatsei jarraikiz, deklinabidea da zuzendu gabe geratu den akatsetako bat (7. taula). Doitasunaren kasuan, beste behin, gaizki ulertu edo itzultako segmentuak izan dira gehien errepikatu direnak (8. taula). Horrez gain, tokiko konbentzioen harira, 9. taulan jasotakoa bezalako akatsak ez dira gehien azaldu direnak, baina taulan lanean aipatzea erabaki dugu, ezjakintasunak baino arreta-faltak eragindako errorea izan daitekeelako. Eta PE egoki bat aurrera eramateko, arreta mantentzea ezinbestekoa da. Azkenik, estiloari buruz hitz egingo dugu. Ez da akatsik larriena, baina bada garrantzitsua; izan ere, IAk jatorrizko bertsioari gertutik jarraitzeko tendentzia du,

euskara berezi egiten duen ezaugarrietako bat (sintaxia) baztertuz (10. taula), eta wikilariak, sarri, ez dira arazo horiek identifikatzeko gai izaten.

Taula 7: Mantendutako akatsak (jariotasuna)

14. TESTUA	Jatorrizkoa	PE
	(...) more active involvement of research investors in collaboration with research providers developing programs of work.	(...) ikerkuntzan inbertitzaileek parte-hartze aktiboagoa izatea, lan-programak garatzen dituzten ikerketa-hornitzaileekin lankidetzan.

Gehitutako akatsen artean, jariotasunaren kategorian, arazo gramatikalak dira nagusi. Postedizio-azterketan zehar akats gramatikal ezberdinak aurkitu ditugu, baina horien artean ohikoenak deklinabideari lotutakoak izan dira. Alegia, bai Eliak bai eta erabiltzaileak ere, ez dute deklinabide zuzena asmatzerik lortu. Kasu honetan, esaterako, -n beharrean, “ikerkuntza” hitzak -ko deklinabidea izan beharko luke.

Taula 8: Mantendutako akatsak (doitasuna)

8. TESTUA	Jatorrizkoa	PE
	Skeuomorphism is parallel to, but different from, path dependence in technology , where an element's functional behavior is maintained even when the original reasons for its design no longer exist.	Eskeumorfismoa teknologia-bidearen mendekotasunarekiko paraleloa da, baina desberdina, non elementu baten portaera funtzionala mantendu egiten baiten, baita diseinuaren jatorrizko arrazoiak existitzen ez direnean ere.

Sorburu-testuari erreparatuz ikus daiteke elkartu beharreko hitzak ez direla *teknologia* eta *bidea*, *bidea* eta *menpekotasuna* baizik. Hori horrela, ez IAK ezta wikilariak ere ez dute esaldia ulertu.

Taula 9: Mantendutako akatsak (tokiko konbentzioak)

9.	Jatorrizkoa	PE

TESTUA	"I started taking it seriously by the time I went to college when I was 17."	Panpingintza serio hartzen hasi nintzen 17 urterekin, unibertsitatera joan nintzenean".
--------	--	---

Bestalde, tokiko konbentzioen kategorian aurkitu dugun akatsetako bat, nahiz eta goian esan bezala errepikatuena ez izan, aipua txertatzeko kakotxak gaizki erabiltzea izan da. Izan ere, teorian, aurrean eta atzean jarri behar dira, baina IAren bertsioan ez dira ageri eta wikilariak ez du akatsa identifikatu.

Taula 10: Mantendutako akatsak (estiloa)

3. TESTUA	Jatorrizkoa	PE
	Estos últimos dos colectivos tienen incluso dos asociaciones destinadas a la promoción de esta demanda.	Azken bi talde horiek eskaera hori sustatzeko bi elkarte propio ere badituzte.

Azken adibide hau ere, berez, ez da akats bat. Hain zuzen, jatorrizko bertsioak ere erakuslea birritan erabili du: “*estos*” eta “*esta*”. Sorburu-testuan erabili diren bi erakusleen generoa eta numeroa ez da berdina, eta, beraz, esaldia ez da errepikakorra. Euskal bertsioan, ostera, erakusle bera erabili denez bi kasuetarako, “hori”, testua errepikakortasun kutsu bat du. Gainera, gazteleraz erakuslea beharrezkoa da, baina euskaraz ken daiteke.

8. Ondorioak

Atarian esan bezala, GRAL honekin bi helburu lortu nahi genituen: alde batetik, Wikipedian, edo azken horri lotutako eremuetan (EWKW), itzulpen-politika formal zein informalik dagoen ikustea; eta, bestetik, Wikipedian gauzatzen den postedizioa zer nolakoa den aztertzea.

Wikipediak eskaintzen dituen baliabide ezberdinak kontsultatu ostean, ohartu gara euskal entziklopedia askeko komunitateak ez duela itzulpena arautzen duen neurririk. Hutsune horrek, hortaz, bide ematen du Góngora-Goloubintseff (2021, 2022) eta Shuttleworthek (2017) aipatutako continuum-aren teoria fenomeno orokortzat hartzeko. Izan ere, itzulpenari lotutako erabakiak estilo-gidan sartzeak erakusten du Wikipedian itzulpena eta edizioa ez direla bi jarduera bereiz, hots, ez bata ez besteak ez dutela tratamendu berezirik jasotzen.

Bestalde, Wikipediak munduko hizkuntzen errealitatea islatzen du, hau da, bizitza errealean txikiak diren hizkuntzak Wikipedian komunitate txikiak izango dituzte, eta alderantziz. Hori horrela, egoera horri aurre egin eta komunitate txikietako edukiak modu azkar eta eraginkor batean hedatzeko, horietako askok itzulpena jo dute. Hala ere, Wikipedia eta itzulpenaz luze eta zabal hizketan aritu ostean, beharrezkoa deritzogu eztabaida berri bat mahaigaineratzea: itzulpena ez ezik, Wikipedian nagusitzen dena ez al da postedizioa izango? *Content Translation* eta IA konbinatzeari esker, itzulpen-prozesua goitik behera automatizatu da. Hau da, jada ez dira wikilariak itzulpenak egiteaz arduratzen direnak, Eliak emandako testuak posteditatzeaz baizik.

Azterketan ikusi ahal izan dugunez, IAk egindako akats-kopuru handiena jarioasuna deritzon kategorian aurkitu dugu. Hain da horrela, ezen azken horrek bildutako errore guztien % 57 hartzen baituen. Gainera, jarioasunaren baitan, gramatika eta terminologia izan dira arazo gehien eragin dituzten azpi-kategoriak.

Bestalde, erabiltzaileak itzultzaile ez-profesionalak izanik, Eliak egindako akats gehienak zuzentzen dituzten arren (% 76), horietako batzuk identifikatu gabe uzten dituzte (% 24); eta, are gehiago, berriak ere gehitzen dituzte (% 22). Aipatzekoa da, gainera, IAren erroreekin gertatzen den bezala, jarioasunak duela ehunekorik altuena bi kasuetan (% 58 eta % 59, hurrenez hurren).

Postedizio-azterketaren analisiarekin bukatzeko, kontuan hartu behar da, hautatutako testu gehienak EIZIEk bere webgunean Solaunen gidalerroak argitaratu ostekoak direla (gidalerroen argitalpen data: 2022ko iraila), eta gidalerro horiek baliatzeari esker akats batzuk saihestu zitezkeela. Ondorioz, beharbada, entziklopedia askeko erabiltzaileek ez dute gidalerro horien berri. Hori ikusirik, etorkizuneko ikerketei begira interesgarria izango litzateke wikilariak postedizioarekiko duten jarrera aztertzea, eta jarduera horren aurrean hartzen dituzten erabakiak (gidalerrorik kontsultatzen duten, esperientziarik duten, etab.) ezagutzea.

Amaitzeko, gogorarazi nahiko genuke entziklopedia askean gauzatzen den postedizioa oraindik hobetzeke dagoen arren, ezin daitekeela ahaztu wikilariak izugarrizko lana egiten ari direla euskal edukiek ahalik eta hedapen handiena izan dezaten. Eta, era berean, itzulpen-prozesuan zehar Elia eta *Content Translation* plataformak baliatzen dituztenez gero, azken bi horiek ere eginkizun garrantzitsua dutela Wikipedian. Halere, erabiltzaileen lan horrek bere fruituak eman ditzan, garrantzitsua da arreta kantitatean baino kalitatean jartzea. Horrenbestez, postedizioaren munduan *laguntza* eta *traba* ondo bereizi behar dira, eta kontuan hartu batzuetan hobe dela hutsetik hastea, gaizki eginda dagoen zerbait hobetzen ahalegintzea baino.

9. Erreferentzia bibliografikoak

Administratzaileak. (d.g.). In *Wikipedia*. 2023ko martxoaren 20an hemen eskuragarri: [Wikipedia:Administratzaileak - Wikipedia, entziklopedia askea](#).

Alegria, I. et al. (2013). Recirpocal Enrichment Between Basque Wikipedia and Machine Translation. In I. Gurevych eta J. Kim (Ed.), *The People's Web Meets NLP* (101-118 .or.). Springer Publishing. [Reciprocal Enrichment Between Basque Wikipedia and Machine Translation | SpringerLink](#)

Aranberri, N eta Iñurrieta, U. (2021). Postedizio lantegia: ikastaroaren gakoak. *SENEZ Aldizkaria*, (52), 165-187.

Bost euskarriak. (d.g.). In *Wikipedia*. 2023ko martxoaren 20an hemen eskuragarri: [Wikipedia:Bost euskarriak - Wikipedia, entziklopedia askea](#).

Burokratak. (d.g.). In *Wikipedia*. 2023ko martxoaren 20an hemen eskuragarri: [Wikipedia:Burokratak - Wikipedia, entziklopedia askea](#).

Comparin, L., eta Mendes, S. (2017). *Using error annotation to evaluate machine translation and human postediting in a business environment*. EAMT 2017 konferentzia, Praga. 2023ko maiatzaren 17an hemen eskuragarri: https://ufal.mff.cuni.cz/eamt2017/user-project-product-papers/papers/user/EAMT2017_paper_76.pdf

Elhuyar ezagutuz aldatzea. (2021eko abenduak 9). *Wikimedia Foundationek eta Elhuyarrek hitzarmena sinatu dute Elia wikipedietan erabiltzeko*. [Wikimedia Foundationek eta Elhuyarrek hitzarmena sinatu dute Elia wikipedietan erabiltzeko | Elhuyar](#)

Elia Euskarazko Wikipedian: nola erabili. (d.g.). [Elia Euskarazko Wikipedian: nola erabili – Euskal Wikilarien Kultur Elkarte \(wikimedia.eus\)](#)

Estilo gida. (d.g.). In *Wikipedia*. 2023ko apirilaren 4an hemen eskuragarri:
https://eu.wikipedia.org/wiki/Wikipedia:Estilo_gida

Góngora-Goloubintseff, J. G. (2021). *Translation in Wikipedia: A praxeological study of normativity, negotiation and automation across four language communities* [Doktorego tesia, University of Manchester]. The University of Manchester, Faculty of Humanities, United Kingdom.

Góngora-Goloubintseff, J. G. (2022). Assessing the impact of translation guidelines in Wikipedia: A praxeological approach to the study of documented standards across four language communities. *Translation Spaces*, 11(2), 254-276.

Hong Kong Baptist University. (2021). *Understanding Wikipedia's Dark Matter: Translation and Multilingual Practice in the World's Largest Online Encyclopaedia* [Eskuorria].

Itzulpen automatikoak. (d.g.). In *Wikipedia*. 2023ko martxoaren 20an hemen eskuragarri: [Lankide eztabaida:Ibarreusk - Wikipedia, entziklopedia askea](#).

Jones, H. (2013). Wikipedia as a translation zone: a heterotopic analysis of the online encyclopedia and its collaborative volunteer translator community. *Target*, 31(1), 77-97.
<https://benjamins.com/online/target/articles/target.18062.jon>

Sycz-Opón, J. eta Gałuskina, K. (2017). Machine translation in the hands of trainee translators –an empirical study. *Studies in logic, grammar and rhetoric*, 49(62), 195-212.

Lankideak. (d.g.). In *Wikipedia*. 2023ko martxoaren 20an hemen eskuragarri: [Wikipedia:Lankideak - Wikipedia, entziklopedia askea](#).

Lankide aktiboen zerrenda. (d.g.). In *Wikipedia*. 2023ko martxoaren 20an hemen eskuragarri: [Lankide aktiboen zerrenda - Wikipedia, entziklopedia askea](#).

- Laxström, N., Giner, P. eta Thottigal, S. (2015eko ekaina). Content Translation: Computer-assisted translation tool for Wikipedia articles [Txostena]. *Proceedings of the 18th Annual Conference of the European Association for Machine Translation, Antalya, Turkey* (194-197. or.). European Association for Machine Translation. https://helda.helsinki.fi/bitstream/handle/10138/156250/eamt2015_cx_final.pdf?sequence=1
- McDonough Dolmaya, J. (2014). Revisioning history: Translation trends in Wikipedia. *Translation Studies*, 8(1), 16-34. [Revision history: Translation trends in Wikipedia: Translation Studies: Vol 8, No 1 \(tandfonline.com\)](#)
- McDonough Dolmaya, J. (2017). Expanding the sum of all human knowledge: Wikipedia, translation and linguistic justice. *The Translator*, 23(2), 143-157. <https://doi.org/10.1080/13556509.2017.1321519>
- O'Brien, S. (2011). Towards predicting post-editing productivity. *Machine translation*, 25(3), 197-215.
- Patten, A. (2001). Political Theory and Language Policy. *Political Theory*, 29(5), 691-715. https://scholar.princeton.edu/sites/default/files/apatten/files/pt_2001.pdf
- Politikak. (d.g.). In *Wikipedia*. 2023ko martxoaren 20an hemen eskuragarri: [Wikipedia:Politikak - Wikipedia, entziklopedia askea.](#)
- Sarasola, K. (d.g.). Itzulpen automatikoa erabili eta gero... “giza zuzenketa” behar da [Blog bateko sarrera]. *Euskal Wikilarien Kultur Elkartea*. <https://wikimedia.eus/2020/04/itzulpen-automatikoa-erabili-eta-gero-giza-zuzenketa-behar-da/>
- Shuttleworth, M. (2017). Locating foci of translation on Wikipedia: Some methodological proposals. *Translation Spaces*, 6(2), 310-332.

- Solaun, A. (2021). *Wikipedian posteditatzeko gidalerroak* [Gradu Amaierako Lana, UPV-EHU]. Euskal Herriko Unibertsitatea, Letren Fakultatea, Euskal Herria.
- Torres-Simón, E. (2018). The concept of translation in Wikipedia. *Translation Studies*, 12(3), 273-287. <https://doi.org/10.1080/14781700.2018.1534605>
- Txokoa: Elia Itzultzailea ContentTranslationen prest. (d.g.). In *Wikipedia*. 2023ko martxoaren 20an hemen eskuragarri: [Elia itzultzailea ContentTranslationen prest on Wikipedia:Txokoa](#)
- User: Jimbo Wales/Statement of principles. (d.g.). In *Wikipedia*. 2023ko martxoaren 20an hemen eskuragarri: [User:Jimbo Wales/Statement of principles - Wikipedia](#)
- Wikipedia. (d.g.). In *Wikipedia*. 2023ko martxoaren 20an hemen eskuragarri: [Wikipedia - Wikipedia, entziklopedia askea.](#)
- Wikipedians. (d.g.). In *Wikipedia*. 2023ko martxoaren 20an hemen eskuragarri: [Wikipedia:Wikipedians - Wikipedia](#)

10. Eranskinak

10.1. Erdal Wikipediatako artikulak

- 1) Adrienne Avril de Sainte-Croix:
https://fr.wikipedia.org/wiki/Adrienne_Avril_de_Sainte-Croix
- 2) Tesoro de Aliseda: https://es.wikipedia.org/wiki/Tesoro_de_Aliseda
- 3) Movimiento por la oficialidad de la lengua asturiana:
https://es.wikipedia.org/wiki/Movimiento_por_la_oficialidad_de_la_lengua_asturiana
- 4) Solar power: https://en.wikipedia.org/wiki/Solar_power
- 5) Frente del Oriente Próximo (Primera Guerra Mundial):
[https://es.wikipedia.org/wiki/Frente_del_Oriente_Pr%C3%B3ximo_\(Primera_Guerra_Mundial\)](https://es.wikipedia.org/wiki/Frente_del_Oriente_Pr%C3%B3ximo_(Primera_Guerra_Mundial))
- 6) Transmisión de energía eléctrica:
https://es.wikipedia.org/wiki/Transmisi%C3%B3n_de_energ%C3%ADa_el%C3%A9ctrica
- 7) Reloj (complemento): [https://es.wikipedia.org/wiki/Reloj_\(complemento\)](https://es.wikipedia.org/wiki/Reloj_(complemento))
- 8) Skeumorph: <https://en.wikipedia.org/wiki/Skeuomorph>
- 9) Greer Lankton: https://en.wikipedia.org/wiki/Greer_Lankton
- 10) Estudios de medios: https://es.wikipedia.org/wiki/Estudios_de_medios
- 11) Marie Goegg-Pouchoulin: https://fr.wikipedia.org/wiki/Marie_Goegg-Pouchoulin
- 12) Más Madrid: https://es.wikipedia.org/wiki/M%C3%A1s_Madrid
- 13) Medical students' disease:
https://en.wikipedia.org/wiki/Medical_students%27_disease
- 14) Agricultural policy: https://en.wikipedia.org/wiki/Agricultural_policy
- 15) Rediet Abebe: https://en.wikipedia.org/wiki/Rediet_Abebe
- 16) Regina Honu: https://en.wikipedia.org/wiki/Regina_Honu
- 17) William Glasser: https://en.wikipedia.org/wiki/William_Glasser
- 18) Precautionary principle: https://en.wikipedia.org/wiki/Precautionary_principle

10.2. Euskal Wikiledian argitaratutako bertsoiak

- 1) Adrienne Avril de Sainte-Croix:
https://eu.wikipedia.org/wiki/Adrienne_Avril_de_Sainte-Croix
- 2) Alisedako altxorra: https://eu.wikipedia.org/wiki/Alisedako_altxorra
- 3) Asturieraren ofizialtasunaren aldeko mugimendua:
https://eu.wikipedia.org/wiki/Asturieraren_ofizialtasunaren_aldeko_mugimendua
- 4) Eguzkiaren energia: https://eu.wikipedia.org/wiki/Eguzkiaren_energia
- 5) Ekialde Hurbileko Frontea Lehen Mundu Gerran:
https://eu.wikipedia.org/wiki/Ekialde_Hurbileko_Frontea_Lehen_Mundu_Gerran
- 6) Energia elektrikoaren garraio:
https://eu.wikipedia.org/wiki/Energia_elektrikoaren_garraio
- 7) Erlojua (osagarria): [https://eu.wikipedia.org/wiki/Erlojua_\(osagarria\)](https://eu.wikipedia.org/wiki/Erlojua_(osagarria))
- 8) Eskeumorfismo: <https://eu.wikipedia.org/wiki/Eskeumorfismo>
- 9) Greer Lankton: https://eu.wikipedia.org/wiki/Greer_Lankton
- 10) Komunikabideen zientziak:
https://eu.wikipedia.org/wiki/Komunikabideen_zientziak
- 11) Marie Goegg-Pouchoulin: https://eu.wikipedia.org/wiki/Marie_Goegg-Pouchoulin
- 12) Más Madrid: https://eu.wikipedia.org/wiki/M%C3%A1s_Madrid
- 13) Medikuntzako ikasleen nahasmendua:
https://eu.wikipedia.org/wiki/Medikuntzako_ikasleen_nahasmendua
- 14) Nekazaritza politika: https://eu.wikipedia.org/wiki/Nekazaritza_politika
- 15) Rediet Abebe: https://eu.wikipedia.org/wiki/Rediet_Abebe
- 16) Regina Honu: https://eu.wikipedia.org/wiki/Regina_Honu
- 17) William Glasser: https://eu.wikipedia.org/wiki/William_Glasser
- 18) Zuhurtasun printzipio: https://eu.wikipedia.org/wiki/Zuhurtasun_printzipio