

## Euskararen normalizazioa osasun-arloan

### 3.5. Itzulbide: osasun-txostenen itzulpen automatiko neuronal

Olatz Perez de Viñaspre Garralda

*Ixa taldea, UPV/EHU*

*olatz.perezdevinaspre@ehu.eus*

#### 1. Sarrera eta arloaren egoera

---

Euskal Autonomia Erkidegoan (EAE) euskara eta gaztelania hizkuntza ofizialak dira, eta txosten ofizial guztiak bi hizkuntzetan argitaratzen dira. Osakidetza EAEko osasun zerbitzu publikoa denez, Erakundeak bi hizkuntzetan argitaratzen ditu txosten orokorrak; baina txosten klinikoei dagokienez, gaztelania izan da orain arte erabili den hizkuntza (ia) bakarra. Hala ere, eta Osakidetzaren proiektu estrategikoaren arabera, aurreikusitako ekintzetako bat da “dokumentazio klinikoan eta historia klinikoan euskararen presentzia pixkanaka handitzea” (1.4.2. ekintza, Osakidetzaren 2017-2020ko erronka eta proiektu estrategikoei buruzkoa), gaur egunean oraindik errealitatea ez bada ere.

Osakidetzako langile egonkor elebidunen portzentajea urtetik urtera goraka doa (% 2-4 urteko, gaur egunean ia-ia % 50era heldu arte), eta agerikoa da langile gazteenen kasuan portzentaje hori askoz altuagoa dela. Hala eta guztiz ere, profesional falta eta kanpotik sortutako arazoak direla medio, Osakidetzako langile guztiak euskaldunak izango diren eguna urrun dugu oraindik; ondorioz, historia klinikoan euskarak gaztelaniarekin partekatu beharko du tokia. Horretaz gain, Europatik datozen berriei kasu eginez gero, pertsona guztien historia kliniko laburtua Europako edozein hizkuntzatan egongo da eskuragarri etorkizunean. Testuinguru horretan, ezinbestekoa ematen du txosten elebidunen arloa jorratzen hastea, eta, jakina, euskarari dagokion esparrua aitortu eta lantzen hastea.

Tamalez, osasun-arloko langileek denbora oso mugatua dute pazienteak artatzeko eta dagokion txostena idazteko. Beraz, txosten horiek bi hizkuntzetan idaztea ez da irtenbide bideragarria profesionali laguntza eskaintzen ez bazaie. Gaur egunean teknologiak eskaintzen dizkigun aukerekin, itzulpen automatikoko sistema batlaguntza tresna paregabea izan daiteke.

Dokumentazio elebiduna automatikoki sortzeari dagokionez, antzeko esperientziak egin dira justiziaren arloan. Kasu zehatz horretan, aurrez erredaktatutako testu kopuruaz eta testuen formaltasunaz baliaturik, alde aurreko itzulpenetan oinarritutako estrategia bat baliozkotu ahal izan zen, eta, itzulpen-memoriak erabiliz, beharren arabera emaitzak lortu ziren. Dokumentazio klinikoaren esparrua eta justiziarena oso desberdinak dira: testu juridikoen formaltasuna ez dator bat txosten klinikoetan erabilitako testu librearekin; esate baterako, txosten klinikoetan siglak eta akronimoak erabiltzen dituzte masiboki (hazietako askoez daude ezta adostuta ere), edo amaitu gabeko esaldiak (aditzik gabeko esaldiak). Gainera, testu horien aldagarritasuna handia da, bai estiloetan bai terminologiaren espezializazioan. Ezaugarri horiek testuak prozesatzeko teknika aurreratuagoak eskatzen dituzte, edo, zehazkiago, itzulpen automatikoko teknikak.

Itzulpen automatikoak, berriki, iraultza bizi izan du sare neuronaletan oinarritzen diren teknikei esker. Itzulpen automatiko modernoaren hastapenetan, erregetan oinarritutako itzultzaileak garatzen ziren; hau da, itzulpenak egiteko, hizkuntzalari zein itzultzaileek aztertutako itzulpen erregelak hiztegi elebidunekin konbinatzen ziren. Hala ere, itzultzaile horiek oso antzekoak ziren hizkuntzetan arrakasta izan bazuten ere (gaztelaniaren eta katalanaren arteko itzulpena, adibidez), sintaktikoki ezaugarri ezberdinak zituzten hizkuntzen arteko itzulpenak ez ziren behar bezalakoak. Izan ere, eskuz idatzitako erregela horiek hizkuntzen paradigma guztiak kontutan hartu behar zituztenez, horrek arazoak zekartzen hizkuntzaren anbiguotasunari aurre egiteko. Itzultzaile horiek sortu zituzten, besteak beste, “lehenengo **landarea**” (“*primera planta*”), “azkeneko **batasunak**” (“*últimas unidades*”) edota “autobusak goiz eta **berandu**” (“*autobuses mañana y tarde*”) itzulpen traketsak.

XX. mendeko 90eko hamarkadatik aurrera, digitalizatutako corpus kopurua areagotzearekin batera, corpusetan oinarritutako metodoak agertu ziren (berezi metodo estatistikoak). Metodo horiekondoko aldagaietan oinarritzen dira itzulpenak egiteko: corpus paraleloetako hitz segiden agerpenen maiztasuna eta haien probabilitatea. Corpus paraleloak bi hizkuntzatan (edo gehiagotan) idatzitako testu bildumak dira, zeinak esaldi mailan lerrotatuta dauden. Elhuyarrenweb-corpusen atarikocorpus paraleloa kontsultatzeko interfazean<sup>1</sup>, adibidez, esaldiak lerrotatuta ikus ditzakegu:

### 1. irudia: Elhuyarren web-corpusen atarikocorpus paraleloaren pantaila-irudia

Emaitzak: 176 itzulpen-unitate, 36 domeinu	
Euskara	Gaztelera
Clavenet-eko kodeak (sarrerakoa nahiz sinadurakoa) blokeatuko dira, hiru aldiz <b>oker</b> sartzen badira.	Las claves de Clavenet (tanto la de entrada como la de firma) se bloquean cuando se introducen erróneamente tres veces.
Nola ziurta dezaket ez dudala <b>okerrik</b> egin datuak aginduan sartzean?	¿Cómo puedo asegurarme que no he cometido errores en la introducción de datos en la orden?
<b>Okerrik</b> diseinuan.	Errores de diseño.
Ligatzean ez <b>okerrik</b> egin (emakume eta gizonezkoak)	Cuando liganos no te liganos (mixto)
Iñaki Zubeldia: Zure liburu konkretuetara etorri, zure gazte literaturako lehen obra, <b>oker</b> ez banago, "Zikoinen habian sartuko naiz" liburua da, 1986an idatzia.	Iñaki Zubeldia: Volviendo a tus libros, si no estoy confundido, tu primera obra de Literatura Joven fue el libro "Zikoinen habian sartuko naiz", que escribiste en 1986.
Gaiaren arabera ikuspegi eta tratamendua ere ezberdina da, jakina, baina, <b>oker</b> ez banago, guztietan ezaugarri nagusietako batzuk irudimen handia, hizkuntzarekin jolasteko grina, umorea eta haurrenganako hurbiltasuna dira.	Según el tema que se trate, varían la perspectiva y el tratamiento, por supuesto, pero, si no es equivocado, las características principales de todos ellos son una gran imaginación, la pasión por jugar con el lenguaje, el humor y la proximidad a la infancia.
Etxaldea, fatxada luzituta eta leilatu <b>okerrik</b> dituzten eraikuntzez osatuta dago.	El caserío está formado por edificios de fachada enlucida y empinados tejados.
Ez ziren hemeretzigarren mendeko eraikuntza eta zabalguneko zoruaren ezaugarriak aintzat hartuta idatzitakoak izan, eta susmatzeko da <b>okerra</b> izan dela halaber ordenantzen kudeaketa; sustatzaileen aldetik ez ezik baita erakundeen aldetik ere.	No fueron consideradas las características constructivas ni las características del suelo ensanche en su redacción, y es de sospechar que su gestión ha sido equivocada; no sólo por parte de promotores sino por parte de la administración.
<b>Okerreko</b> sarrerak itzultzeko eskaera	Solicitud devolución de ingresos indebidos

Azkenik, 2015. urtetik aurrera, itzulpen automatikoaren kalitateak berebiziko iraultza izan zuen sare neuronaletan oinarritutako arkitekturei esker. Arkitektura horiek gure burmuineko neuronetan dute inspirazioa: neurona bakoitzak ekuazio lineal simple bat kalkulatu du, eta milaka neurona konbinatuz, ataza oso konplexuak ikasteko gai dira. Harrigarria da nola oso antzekoak diren sare neuronaletan oinarritutako sistemen ikaste prozesua eta pertsonen ikaste prozesua. Izan ere, testuingurua erabiltzen dute esanahia ezagutzeko, guk geuk hitz berri bat ikastea gauzatzen dugun prozesu berdina. Hau da, demagun “Yolixpa” hitza ezezaguna zaigula, eta honakoa irakurtzen dugula:

“Mexikora egin genuen bidai hartan probatu genuen Yolixpa lehen aldiz”.

Esaldi horretatik, “Yolixpa”ri buruz janari edo edari bat izango dela ondoriozta dezakegu, baita Mexikokoa dela ere. Jarraian, beste esaldi hau irakurtzen dugu:

“Gogoratzen dudana ajerik okerrera izan nuen Yolixpa edan eta gero”.

Bi esaldiak irakurrita (hiztegiko definizioa begiratu gabe), “Yolixpa” Mexikoko edari alkoholdu bat dela jakin dezakegu; eta, antza, sare neuronalek modu antzekoan ikasten dituzte hitzen eta esaldien errepresentazioak eta itzulpenak.

1 Web-corpusen ataria: <http://webcorpusak.elhuyar.eus/cgi-bin/kontsulta2.py> (2020ko otsailaren 29an atzitu)

Itzultzaile neuronalek corpus kopuru benetan erraldoiak behar dituzte horren emaitza onak lortzeko (milioika eta milioika hitz) eta, ondorioz, ez da beti posible itzultzaile neuronal dezente bat sortzea baliabide nahiko ez badaude. Ildo horretatik, Koehn eta Knowles egileek (2017) itzulpen automatiko neuronalaren 6 erronka plazaratu zituzteneuren lanean, zeinetatik<sup>3</sup> bereziki esanguratsuak dira arlo klinikora egokitze eta euskararekin lan egiteko:

Corpus erraldoien beharra: esate baterako, aurretik aipatutako lanean (Koehn eta Knowles, 2017), ingelesa-gaztelania hizkuntza parerako, 385,7 milioi hitzeko corpus paraleloa erabiltzen dute emaitza onenak lortzen dituen sistema entrenatzeko. Halere, halako corpus handiakeskuratzea ezinezkoa izan daiteke hizkuntza pare eta domeinu zehatz gehienetan.

Domeinura egokitze zailtasunak: corpus erraldoien beharrarekin dauka zerikusia puntu honek. Izan ere, domeinu zehatzetan eskura egoten diren corpusak urriak izaten dira (edo, behintzat, ez erraldoiak), eta horrek sare neuronalaren ikasteko ahalmena asko murrizten du.

Inflexio altuko hitzekin arazoak (deklinabideak eta aditzak): bai euskarazko deklinabide eta aditzekin, zein gaztelaniazko aditzekin gertatzen da. Inflexioa hitzen alterazio bat dat da, deklinabidearen bidez aditzak jokatzeko sortzen dena. Arazo hau hitz arraroekin ere ematen da; izan ere, sare neuronalean entrenamenduak egiteko, hiztegi baten tamaina definitu behar da, eta hiztegi horretan sartzen ez diren hitzak ez dira ikasten. Horrek ikas ahalmenean eragina dauka, sare neuronalek ikas dezaketen hitz kopurua mugatzen duelako. Kontuan izan ordenagailuak ulertzen duela “etxe”, “etxea”, “etxera” eta “etxetik” 4 hitz ezberdin direla.

Mundu osoko ikerlarien komunitate handi batek lanean dihardu erronka horiei (eta gehiagori) soluzioak proposatu eta soluziook probatzeko, eta egunetik egunera hobekuntzak nabarmenak dira.

Itzulpen automatikoan emandako aurrerapenek eta Osakidetzaren barnean langileek euskaraz bizi eta lan egiteko duten uneko grinak bideragarria egin dute txosten klinikoak itzuliko dituen itzultzaile automatiko bat garatzeko ikerketa abiaraztea.

## 2. Itzulbide proiektua

---

Itzulbide proiektua **ikerketa** proiektu bat da, Osakidetzak eta UPV/EHUko Ixa taldeak elkarlanean garatzen hasi zirena 2019ko udan.

### 2.1. Helburuak

Itzulbideren helburu nagusia da txosten klinikoak euskaratik gaztelaniara eta gaztelaniatik euskarara itzuliko dituen itzultzaile automatiko bat garatzea. Itzultzaile horrek gaur eguneko sektoreak eskain dezakeenaberdiru edo hobetuko badu, sare neuronalean oinarritutako itzultzailea beharko du izan. Horretarako, ezinbestekoa zaigu txosten klinikoaren corpus paraleloa (elebiduna) osatzea. Izan ere, itzultzaile neuronalek corpusetatik ikasten dute, eta haien fidagarritasunak eragin zuzena dauka itzulpenen kalitatean. Zer esanik ez, corpus horren tamainak ere eragina dauka, eta, horregatik, corpusak ahalik eta handiena izan beharko du.

### 2.2. Abiapuntua

Proiektuarekin hasi aurretik, Ixa taldean oinarritako itzultzaile bat garatu genuen txosten klinikoak itzultzeko, mota horretako corpus paralelorik izan gabe (Soto et al., 2019). Lan horretan, besteak beste honako baliabideak erabili genituen itzultzailea entrenatzeko:

Euskara-gaztelania corpus paralelo orokor erraldoi bat: bai euskara bai gaztelania, hizkuntzak eurak, ikasteko baliagarriak zaizkio itzultzaileari. Azken finean, txosten klinikoetan ere euskararen eta gaztelaniaren sintaxi zein ortografia arauak mantentzen dira.

Alor klinikoko hiztegi espezializatuak: corpora balira bezala, hiztegiko sarrerak eman dizkiogu itzultzaileari, terminologia espezializatu ikas dezan. Kontuan izan, itzultzaileak entrenamenduan

Olatz Perez de Viñaspre Garralda

ikusitako hitzak bakarrik ezagutzen dituela, eta horregatik dela garrantzitsua entrenamenduan ahalik eta terminologia zabalena barnebiltzea.

Gaztelaniazko txosten klinikoan corpus elebakar handi bat erabili duguback-translation (Senrich et al., 2016) deritzon teknika aplikatzeko; hau da, helburu hizkuntzaren corpora (gaztelaniazkoa gure kasuan) aurretik lortutako itzultzaile onenarekin (corpus orokorra eta hiztegien lortutakoa, gurean) jatorri hizkuntzara (euskara, gure kasuan) itzultzeko. Horrela, sasi-corpus berriak jatorrizko hizkuntza ondo izango ez badu ere, helburua ondo ikasiko du, eta, beraz, itzulpena sortzeko prozesua ondo egingo. Frogatuta dago teknika horrek asko hobetzen dituela itzultzaile neuronalen emaitzak, eta azkenaldian garatu diren itzultzaile gehienek hura erabiltzen dute corpora handitu eta emaitzak hobetzeko.

Kontuak horrela, eta euskarazko corpus elebakar erraldoirik ez dugunez, euskaratik gaztelaniarako itzultzailea izango litzateke bideragarria gaur eguneko baliabideekin. Itzulpen automatikoen adibide batzuk ikus ditzakezue jarraian:

(1-eu) fibrinolisiaren ondoren egoera klinikoak ez da askorik aldatu

(1-es) *tras fibrinólisis el estado clínico no ha variado significativamente*

Lehenengo adibidean, itzulpena pertsona batek egina ematen du; ez dauka inongo akats gramatikarik, ez eta esanahiaren ikuspegitik ere. Egitura zailagoak ere arazorik gabe itzultzeko gai dela ikus dezakegu bigarren adibide honetan:

(2-eu) hipertentsioaren aurkako tratamenduarekin hasi da, tentsioak neurri egokian mantenduz; hipergluzemiarako joera antzeman da egonaldian

(2-es) *se ha iniciado tratamiento antihipertensivo, manteniendo tensiones adecuadas; se ha detectado tendencia a la hiperglucemia durante su estancia*

Hala ere, dena ez da urrea, eta 3. zein 4. adibideetan corpusaren falta nabarmena da.

(3-eu) egun batzuetatik hona **eginkariak** beltzak direla iruditu zaio

(3-es) *desde unos días, se le ha parecido que son negros*

Izan ere, 3. adibidearen kasuan, ematen du “eginkariak” hitza ez duela ezagutu (ziurrenik, ez zen entrenamenduan agertu ere egingo), eta itzulpenean nahasmena da nagusi.

(4-eu) ondorengo egunetan ez zuen botaka egin, eta **beherakoak** gutxituz eta giltzurrunen funtzioa hobetuz joan ziren

(4-es) *en los días posteriores no presentaba vómitos, con disminución de la tos y mejoría de la función renal*

Azkeneko adibidean, beherakoak eta eztula nahasten dituela ikus dezakegu (larria oso, testuinguru klinikoan sintomak nahastea). Antza, corpusean ez du nahikoa kasutan “beherakoak” eta “eztula” hitzak ikusi, eta, ikusi dituenean, beharbada testuingurua oso antzekoaz, nahasgarria.

### 2.3. Planteamendua

Itzulbide proiektua aurrera eramateko, 3 faseko prozesua abiatu dugu:

#### Corpusa lortu

Corpusa lortzeko fasea, fase garrantzitsuentzat har dezakeguna. Osasun-arloko langileen inplikazioa eta laguntza behar-beharrezkoak zaizkigu txosten klinikoan euskara-gaztelania corpus paraleloa sortzeko. Ezinbestekoa da adituek sortutako corpora izatea, eta ez itzultzaile profesionalak itzulia.

Izan ere, aditu horiek dira euskaraz nola idatzi nahi duten erabakiko dutenak (estiloa), eta itzulpen egokiaren garantiak emango dizkigutenak (semantikoki itzulpena zuzena dela). Horretaz gain, Osakidetzako itzultzaileek sarritan aipatutako arazo bat ekidingo genuke: terminologia espezializatua zehaztasunez beteta dago, eta zenbaitetan bi termino sinonimotzat har badaitezke ere, beste zenbaitetan ñabardurak egoten dira bien artean; ondorioz, itzulpenaren zehaztasuna (edota zuzentasuna) galdu egiten da. Ezin dugu ahaztu terminologia berria etengabeen sortzen ari dela, bereziki arlo klinikoan, eta termino berri horiek izendatzeko pertsona egokienak adituak eurak direla. Adituak, itzulpenak egitearekin batera, euskaraz idazteko ohitura hartzen hasiko dira, eta horri lotutako zailtasun eta beharrei aurre egiteko aukera sortuko da.

Corpusa espezialitate multzotan bildu dugu. Espezialitateak multzokatzeko, idazteko estiloan jarri dugu arreta. Adibidez, Pediatrian, Lehen Mailako Arretan eta Larrialdietan antzera idatziko balitz, eurek denek (edo gehiagok) espezialitate multzo bat osatuko lukete. Espezialitate multzo horiek aztertu nahi ditugu, eta, beharrezkoa balitz, bakoitzerako itzultzaile espezifikoki bat entrenatu. Espezialitateez gain, txosten motak ere kontuan izan ditugu: alta-txostenak, ebolutiboak eta abar.

Kopuru minimoak zehaztu ditugu espezialitate multzo horietako bakoitzerako: 220.000 hitz hizkuntzako. Horrekin ez dugu esan nahi espezialitate multzo guztietarako kopuru hori lortzea denik helburua, baizik eta egindako esperimenduak esanguratsuak izateko kopuru minimo hori beharko dugula. Izan ere, beste espezialitate multzoetako testuak beti erabiliko ditugu, baina optimizazioak erreferentziatzeko multzoarekin egingo.

Ezin dugu ahaztu fase honetatik sortuko den baliabidea urrea izango dela etorkizuneko edozein itzultzaile automatiko entrenatzeko. Horregatik, corpusa ahalik eta handiena izatea lortu behar dugu, eta, horretarako, proiektuan ahalik eta parte-hartzaile gehien izatea da bide eraginkorrena.

### **Itzultzailea entrenatu (edo sortu)**

Fase honetan, sektoreko teknika arrakastatsuenak aplikatu eta ataza zehatz honetara egokituko ditugu. Horretarako, aurreko fasean sortutako corpusa gehituko diegu hastapeneko itzultzailearen sorkuntzan erabiliko baliabideei. Hasierako itzultzaile hori garatu genuenetik, sektorean aurrerapauso nabarmenak eman dira, eta aurrerapauso horiek Itzulbide proiektuan ere eragin positiboa izango dutela aurreikusten dugu. Horretaz gain, sortuko diren arazo espezifikoeiako soluzioak proposatuko ditugu, hala nola terminologia espezifikoen erabilera.

Aurretik aipatu bezala, espezialitateak zein txosten motak kontuan hartuko ditugu itzultzailea entrenatzeko esperimenduetan, eta aukerarik emankorrenarekin egingo dugu aurrera. Hau da, baliteke itzultzaile bakarra espezialitate zein txosten mota guztietarako nahikoa izatea, edota hainbat itzultzaile behar izatea hainbat espezialitate zein txosten motatarako.

### **Postedizioa eta ebaluazioa**

Azkenik, hirugarren fasean sortutako itzultzailea ebaluatuko dugu. Sortutako corpuseko 20.000 hitz erreserbatuta izango ditugu ebaluaziorako. Fase honetan, osasun arloko profesionalen laguntza beharko dugu berriz. Ebaluazioa modu kualitatiboan egin ahal izateko, itzulitako esaldien zuzentasuna neurtu beharko da, eta akatsak identifikatu eta sailkatu. Adibidez, aurretik aurkeztutako 4. adibidea hizkuntzaren aldetik gramatikalki zuzena eta jatorria da, baina itzulpenaren esanahia ez dator erabat bat jatorrizkoarekin, akatsa semantikoa izanik.

Ebaluazio kualitatiboaz gain, postedizio kanpaina bat abiatuko dugu parte-hartzaile batzuekin. Itzulpen automatiko bat oinarri hartuta, hari zuzenketak (edizioak) egitean datza postedizioa. Hori horrela izanik, aurreko fasean lortutako itzultzaile automatiko emankorrena martxan jarriko dugu, parte-hartzaileek hura erabili eta zuzenketak egin ditzaten. Itzultzaileak zuzenketa horietatik nola ikas dezakeen ere ikertuko ditugu fase honetan, erabilerarekin kalitatea hobea dezan.

Ebaluazio eta postedizio lanetarako ez dugu lehenengo fasean behar izan dugun parte-hartzaile kopuru handia beharko, eta talde txikiago batekin nahikoa izango da.

### 3. Ondorioak

---

Itzulbide ikerketa proiektu bat da, 2019ko udan martxan jarri zena, UPV/EHUko Ixa taldearen eta Osakidetzaren lankidetzari esker. Itzulpen automatikoa (euskararen eta gaztelaniaren artean) txosten klinikoetan aplikatzeko metodoak ikertzean datza proiektua.

Gaur egungo teknologiak euskaratik gaztelaniara itzultzeko kalitate hobea eskaintzen digu (*back-translation* deritzon teknikari esker), eta, izatez, noranzko horrek du premiarik handiena euskararen normalizazioan aurrerapausoak emateko. Izan ere, profesional euskaldunek trabak aurkitzen dituzte euskaraz idazteko orduan, jarraian datorren profesionalak testu hori ulertuko ez duen beldurrez.

Itzulbide proiektutik lortu dezakegun itzultzaile automatikoak ez ditu osasun alorrak exijitzen dituen kalitate neurriak beteko, baina lehen pausu garrantzitsua izango da ibilbide horretan. Gaur egungo itzultzaile automatikoek corpus kopuru erraldoiak behar izaten dituzte, eta itzultzailea corpus handiagoa lortzeko lagungarria izan daiteke, profesionalen zuzenketekin.

Itzulpen automatikoa gai polemikoa izan daiteke, bereziki testuinguru kritikoetan; adibidez, osasungintzaren eta hizkuntza minorizatuen kasuan. Hala ere, gaur egunean inortxok ez luke etorkizun eleaniztun bat irudikatuko itzultzaile automatikoen laguntzarik gabe, eta, ziurrenik, horrela izango da osasungintzan ere. Itzulbide proiektuak aukera ematen digu osasun alorreko itzultzaile automatikoaren garapenean subjektu aktibo izateko, eta prozesu osoan behar diren neurri eta erabakiak hartzeko.

Orain da momentua itzultzaile automatiko batekin elkarbizitzeko arauak azertu, hausnartu eta erabakitzeko. Ibilbide horren hasieran gaude; berazjar ditzagun denon artean oinarriak, osasungintzan euskara normalizatzeko tresna baliagarria izan dadin Itzulbide.

### 4. Erreferentziak

---

1. Soto, X., Perez-de-Viñaspre, O., Labaka, G., & Oronoz, M. (2019). Neural machinetranslationofclinicaltextsbetweenlongdistancelanguages. *Journal oftheAmericanMedicalInformaticsAssociation*, 26(12), 1478-1487.
2. Koehn, P., & Knowles, R. (2017, August). Six Challengesfor Neural MachineTranslation. In *ProceedingsoftheFirstWorkshop on Neural MachineTranslation*(pp. 28-39).
3. Sennrich, R., Haddow, B., & Birch, A. (2016, August). Improving Neural MachineTranslationModelswithMonolingual Data. In *Proceedingsofthe54th Annual MeetingoftheAssociationforComputationalLinguistics (Volume 1: Long Papers)* (pp. 86-96).