

## IGARRITZ: euskarazko testu iragarpenerako web ingurune egokitua

*IGARRITZ: adapted web environment for text prediction in Basque*

<sup>1</sup>Mikel Iruskietia, <sup>1</sup>Iker de la Iglesia, <sup>1</sup>Unai Atutxa, <sup>2</sup>Lierni Ortiz

<sup>1</sup>HiTZ - Ixa taldea (UPV/EHU, Donostia, Gipuzkoa)

<sup>2</sup>Lekeitio BHI

**LABURPENA:** Motrizitate mugatua duten ikasleek, garun paralisi batek sorturiko muga dutenek adibidez, tresna egokituak izaten dituzte testuak idazteko; esaterako, begiradaren jarraipeneko hardware bat, zeinarekin ordenagailuan letrak aukeratu eta sistemak iragartzen dituen hitzak aukeratu daitezkeen. Sistema hauek euskaraz idazteko baliabideak izaten dituzte, edo euskarazko hitz zerrendak sartuta iragarpenak aukeratzeko abagunea ematen dute. Edozein testu iragarpenen xedea da testua idazteko esfortzua murriztea, baita testu luzeagoak azkarrago idatzi ahal izatea ere. Irisarekin idaztea hamar atzamarrekin idaztea baino geldoagoa eta nekezagoa da; horrez gain, testu iragarpenerako tresnak euskaraz beste hizkuntzetan baino okerrago ibiltzen dira, eta sarritan nekez laguntzen dute. Lan honen helburua da euskarazko testu iragarpenerako web ingurune egokitu bat egitea, adimen artifizialeko teknikak erabiliz. Horretarako, euskarazko hizkuntza-ereduetan oinarrituriko web interfazea sortu dugu, IGARRITZ izenekoa. Sortutako testu iragarleak *Transformer* arkitektura erabiltzen du eta HiTZ/roberta-eus-euscrawl-base-cased hizkuntza-eredua berrentrenatu da hezkuntzarako corpusarekin (ikasleen testuak, Gizapedia entziklopedia eta hezkuntzari buruzko testuak: Wikipediakoak eta Berriakoak). Bukatzeko, tresna ebaluatu eta egun eskura dagoen beste sistema batekin konparatu dugu, garun paralisia duen Bigarren Hezkuntzako ikasle batek ekoiztako testuekin. Emaitzen arabera, IGARRITZek euskarazko testu iragarpena hobetu egiten du eta begiradarekin idazten duen pertsonak berak askoz errazago eta gehiago idazten duela adierazi du. Horrez gain, ebaluazio automatikoa egin dugu, eta hor ere emaitzak hobetzea lortu dugu.

**HITZ GAKOAK:** Testu iragarpena, begiradaren jarraipena, euskara, hezkuntza, adimen artifiziala, hizkuntza-ereduak

**ABSTRACT:** *Students with limited mobility, for example, those caused by brain paralysis, have adapted tools for writing texts, such as eye-tracking hardware, to select letters and predict words. For instance, they can use eye-tracking hardware to select letters and choose words predicted by the system. These systems offer resources for writing in Basque, and predictions can be customized by inputting Basque word lists. The primary aim of text prediction is to alleviate the effort involved in typing and to facilitate faster or increased text production. However, writing with Iris is slower and more challenging compared to conventional typing with ten fingers. Furthermore, predictive text functionality in Basque is comparatively less effective than in other languages, offering minimal quality output. Thus, the objective of this study is to develop an adapted web environment for Basque text prediction employing artificial intelligence techniques. To achieve this goal, we have developed a web interface named IGARRITZ based on the HiTZ/roberta-eus-euscrawl-base-cased language model, utilizing a Transformer architecture. It was re-trained with an educational Basque corpus sourced from student texts, educational texts from Gizapedia, Wikipedia, and Berria. Finally, we evaluated the tool and compared it with another currently available system using texts produced by a secondary school student with cerebral palsy. The results indicate that IGARRITZ enhances text prediction in Basque. The student, who writes using eye-tracking technology, reported that the writing process has become significantly easier and more efficient. Additionally, our automatic evaluation demonstrated improved results compared to the existing system.*

**KEYWORDS:** text prediction, eye tracking, Basque, education

1

**\*Harremanetan jartzeko/ Corresponding author:** Mikel Iruskietia, HiTZ (UPV/EHU), Bilboko Hezkuntza Fakultatea, Sarriena auzoa z/g, Leioa.  <https://orcid.org/0000-0002-6121-3902>, [mikel.iruskietia@ehu.eus](mailto:mikel.iruskietia@ehu.eus)

**Nola aipatu / How to cite:** Iruskietia, M. de la Iglesia, I.; Atutxa, U.; Ortiz, L. (2024). <<IGARRITZ: euskarazko testu iragarpenerako web ingurune egokitua>>, Ekaia, Ale berezia, xx-xx. (<https://doi.org/10.1387/ekaia.26214>)

Jasoa: maiatzak 13, 2024; Onartua: ekainak 28, 2024

ISSN 0214-9001-eISSN 2444-3225 / © 2024 UPV/EHU



Obra Creative Commons Atribución 4.0 Internacional-en lizentziapean dago

## 1. SARRERA

Mugikortasun mugatua duten eta goiko gorputz-adarretan mugitzeko zailtasunak dituzten ikasleek zenbait baliabide behar izaten dituzte eskolako lanak egiteko eta curriculuma garatzeko. Mugimendurako zailtasunez gain, ahozko komunikaziorako arazoak badaude, komunikatzen eta irakurketa-idazketan laguntzen dien gailu teknologikoren bat behar dute, dela testuak irakurtzeko, dela pentsamenduak eta ezagutzak idazketaren edo beste komunikazio-sistema alternatibo baten bidez adierazteko. Ahozko komunikazioa erabili ezin denean, halaberrez erabili behar dira Komunikazio Sistema Handigarriak edota Ordezkoak (KSHO) [1]. KSHO sistema motak bitan sailka daitezke [2] [11]: i) laguntza gabeko komunikazio sistemak (esku-alfabetoa eta zeinu-hizkuntza, besteak beste) eta ii) komunikazio sistema lagunduak (komunikazio softwareak).

Esandakoarekin lotuta, psikomotrizitateari eragiten dion garun-paralisia duten DBHko ikasleentzat oso zaila da testu luzeak euskaraz sortzea, ohiko testu-editoreen iragarpenek edo hiztegiatan oinarritutako testu iragarleek produkzio-aukera txikia eskaintzen dutelako. Besteak beste, euskara hizkuntza eranskari izatea zailtasun gehitua delako, eta, ondorioz, hitz erro bakarrarekin hainbat hitz osa daitezkeenez, zaildu egiten da testu iragarpeneko tresnen lana. Desgaitasun hori duten ikasleek tresna egokituak izaten dituzte testuak idazteko; esaterako, begiradaren jarraipenerako programekin, letrak aukeratu eta sistemak iragartzen dituen hitzak aukeratu ditzakegu. Hala ere, ordenagailua begiradaren jarraipenarekin erabiltzeak badu zailtasun nagusi bat; honako hau: prozesu motela da begiekin letraz letra idaztea, eta erabili behar duen pertsonari ahalegin handia eskatzen dio testu labur bat idazteko. Beraz, idazketa azkartzeko eta, aldi berean, ahalegina txikiagotzeko, hitzak aurrerako sistema bat erabiltzea aterabide aproposa izan daiteke. Izan ere, horrelako sistemen bidez, teklatu-pultsazio gutxiagorekin idatz daitezke hitzak. Adibidez, idazteko laguntza horiek mugikorrek, testu-prozesadore batzuek eta abarrek dituztenen antzekoak dira.

Lan honetan, DBH 4. mailako ikasle batek, garun-paralisia duenak, begiradaren jarraipenarekin ordenagailua erabiltzeko teknologia erabiltzen du (*eye-tracking*)<sup>1</sup>, eta balia ditzakeen baliabideak honakoak dira: i) Grid3, Easyclick edota ii) AsTeRICS-Grid [3] (estatuko beste unibertsitate batzuek sortutako sistema, esaterako). Aukera horien artean, ikasleak Grid3 darabil ikastetxean, eta begiradaren jarraipenera egokitutako programa honen laguntzaz maneiatu dezake ordenagailua, pantailako teklatu birtualaren bidez: fitxategiak ireki edo ixteko eta testuak idazteko. Horrelako sistemek euskaraz idazteko baliabideak izaten dituzte, edo, euskarazko hitz zerrendak sartuta,

---

<sup>1</sup> Irisbond edo Tobii sistemak ordenagailua irisarekin erabiltzeko baliabideak dira.

iragarpenak aukeratu daitezke. Alabaina, testu iragarleak euskaraz beste hizkuntzetan baino okerrago ibiltzen dira, eta nekez laguntzen dute edozein testu iragarlek duen xedea erdiesten; alegia, testua idazteko esfortzua murriztea eta testu gehiago idaztea edo azkarrago egitea.

Egun, ikasle honek ez du euskarazko idazkera arin eta erosoa errazten dion sistema eraginkorrik; izan ere, idazteko dituen zailtasunek eta idazten duen abiadurak testuen kalitateari ere eragiten diote, ahalegin handia egin behar duenez, esaldi laburragoak edota gutxi landuak erabiltzeko joera baitu. Barbarako, asmatutako gertaera bat idazteko eskatu diogunean, 11 minutu eta 20 segundoan egitura sinpledun bi esaldi idatzi ditu; 19 hitz (103 letra) orotara.<sup>2</sup> Esfortzua txikiagotuz gehiago eta azkarrago idazteko aukera eskaintzeaz gain, iragarpen-sistema egoki batek teklatze- edota ortografia-akatsak saihesten ditu, eta horrek ere ikaslearen komunikazioari mesede egingo dio. Bestalde, euskarazko kalitatezko testuak aurreratzeko sistema izateak, aipatutako ikasle horientzat ez ezik, antzeko egoeran leudekeen beste ikasle askorentzat ere aurrerapena ekarriko lukeelakoan gaude, teklatura erabiltzeko zailtasunak dituzten ikasleek idazketa erraztu eta bizkortzeko sistema bat eskura izan baitezakete.

Lan honen helburua da euskarazko testu iragarpenerako eta irisarekin idazteko programekin bateragarria den web ingurune egokitu bat egitea, adimen artifizialdun teknikak erabiliz. Horretarako, euskarazko hizkuntza-ereduetan oinarrituriko web interfazea sortu dugu, IGARRITZ izenekoak. Sortutako testu iragarleak *Transformer* arkitektura erabiltzen du [4]; zehazki, RoBERTa kodifikatzaile arkitekturan [5] oinarritutako HiTZ/roberta-eus-euscrawl-base-cased hizkuntza-eredua [6] erabili dugu. Hizkuntza-eredu hori erabili dugu euskarazko hitzen distribuzioa eta hitzen arteko erlazioak ikasteko, euskaraz osatutako corpus handi batekin aurrentrenatu baitugu. Aurrentrenamendu (*pre-trained*) horretan ikasitako erregistroak askotarikoak dira, eta DBHn erabiltzen direnetatik bestelakoak. Beraz, hizkuntza-eredua berregokitzeko eta iragarpen-hitzak ikaslearen beharrekin ahalik eta hobekien bat etortzeko, eredua berrentrenatu egin dugu, berariaz bildu eta egokiagoa den hezkuntzari buruzko corpusarekin (Gizapedia eta hezkuntzari buruzko testuak: Wikipediakoak eta Berriakoak), etengabeko aurrentrenamendua (*continual-pretraining*) deritzon teknika erabiliz.

Duen funtzioa kontuan izanik, IGARRITZ honela sailkatu dugu lan honen [7] irizpideen arabera: komunikazioa erraztea eta areagotzea, zehazki, euskarazko testu idatzia iragartzea irisarekin. Bestalde, prezioa edo erabilera doakoa da, eta atzitzeko bidea online da. Materialari dagokionez, baliabide interaktiboa da: erabiltzaileari hurrengo hitza iragartzen dio IGARRITZek, eta, letra bat sartu ahala,

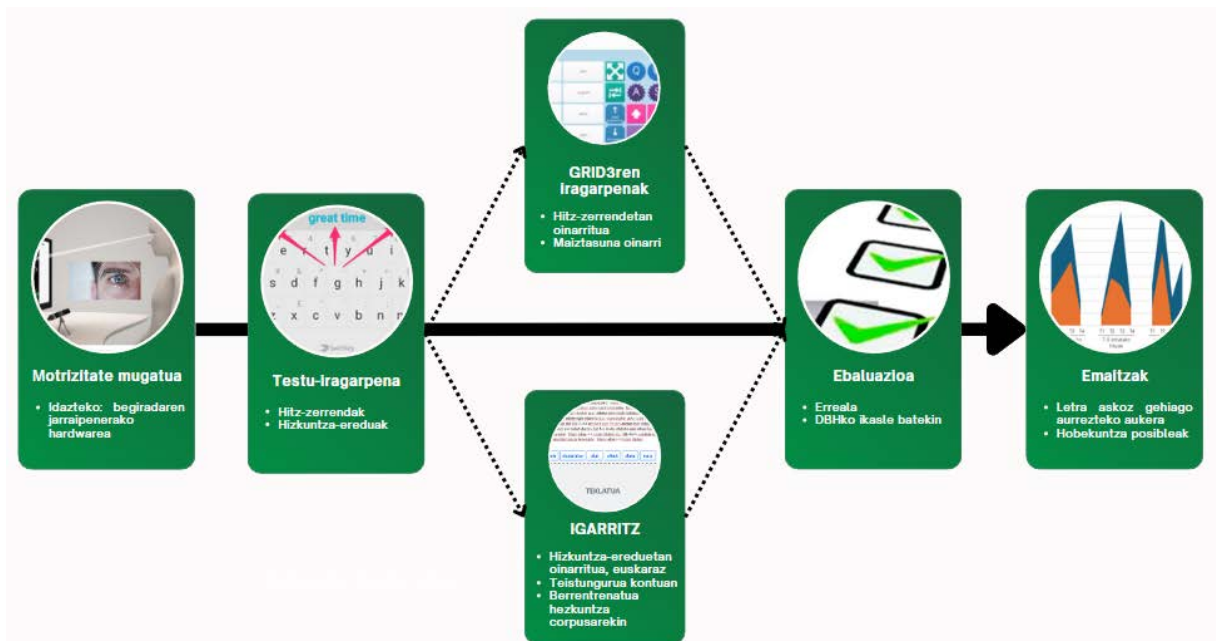
---

<sup>2</sup> Batzuetan, darabilen gailua kalibratu behar izateak ere eragiten dio. Adibide honetan, behar izan dituen 11 minutu pasatxo horietatik, minutu 1 eta 5 segundo gailua kalibratzeko behar izan ditu idazketa-prozesuaren erdian.

iragarpena egokitzen du, idatzitako testuingurura eta sartutako letrara egokituz. Softwarea hizkuntza-teknologian oinarrituta dago, hots, euskarazko hizkuntza-ereduak eta hezkuntzari lotutako corpusak erabili ditugu hurrengo hitza zein den iragartzeko. Erabilitako baliabideen izaerari dagokienez, IGARRITZ euskarazko testuak idazteko testu iragarpena egiten duen tresna da, ezgaitasun motorra dutenentzat.

Esan bezala, testua iragartzen duten tresnek giza esfortzua eta mezuaren ekoizpen-denbora murriztu egin behar dute; beraz, iris bidezko teklatu-pultsazioak gutxitu egin behar ditugu. Emaitzen arabera, IGARRITZek euskarazko testu iragarpena hobetu egiten du, eta irisarekin idazten duen pertsonak berak askoz errazago eta gehiago idazten duela adierazi du. Sortutako IGARRITZ testu iragarlea ebaluatu, eta egundaino ikasleak erabilitako testu iragarlearekin konparatu dugu, aipatutako ikasleak ekoiztutako testuekin. Horrez gain, ebaluazio automatikoa egin dugu, eta hor ere emaitza onak lortu ditugu, esanguratsuak, gainera (ikus 1. irudian prozesu horien laburpena). Bukatzeko, IGARRITZ testu iragarlea CLARIAH-EUS webgunean [12] probatu daiteke, hemen: <http://igarritz.clariah.eu/>.

### 1. irudia. IGARRITZ aplikazioaren sorkuntza



Hurrengo ataletan, lan honetan ezarri dugun helburua lortzeko jarraitu dugun metodologia azalduko dugu. Horretarako, lehendabizi, begiradarekin idazteko sistemez arituko gara, eta sistema horietan euskarak dituen mugak aztertuko ditugu (2. atala). Ondoren, testuak iragartzeko hizkuntza-ereduak izango ditugu mintzagai, eta, jarraian, IGARRITZ sortzeko erabili dugun hizkuntza-eredua zein den zehaztuko dugu (3. atala), baita hizkuntza-eredua elikatzeke erabili ditugun corpusen berri emango ere (4. atala). Gero, hurrengo atalean, IGARRITZ webzerbitzuaz arituko gara (5. atala); hasteko, webzerbitzua eta horren erabilera azalduko ditugu, eta, ostean, IGARRITZ sisteman egin ditugun egokitzapenak eta horiek egin izanaren arrazoen azalpena ere emango ditugu. Azkenik, ebaluazioaz arituko gara (6. atala). Ebaluazio bikoitza izan da burutu duguna. Alde batetik, IGARRITZ automatikoki nola ebaluatu dugun ikusiko dugu; eta, beste alde batetik, kasuan oinarrituta egin dugun eskuzko ebaluazioa aztertuko dugu. Azken honetan, GRID3 tresnaren iragarpenak (ikasleak orain arte erabiltzen izan dituenak) eta IGARRITZ-enak alderatuko ditugu.

## 2. BEGIRADAREKIN IDAZTEKO SISTEMAK ETA EUSKARA

Orain arte, testu iragarpenerako, Grid3 programak barneratua duen sistema erabili izan du ikasleak.<sup>3</sup> Grid3 KSHOrako programa bat da, eta idazketa nahiz ordenagailua erabiltzen laguntzen duen eye-trackerrarekin (Irisbond edo Tobii) bateragarria da. Programa horrek memorian gordetzen ditu erabilitako hitzak, azkarrago idazteko, eta norbere mezu laburrak gorde daitezke, behar denean erabiltzeko. Horrez gain, idatzitakoa entzun daiteke, TTS sistemak integratuta dituelako, eta, halaber, bestelako programa zehatz batzuetara sartzeko lasterbideak erabilerrazak dira.

Euskara, ordea, ez dago Grid3 programaren hizkuntza aukeren artean. Edonola ere, aukera bat da 10000 hitzeko kanpo hiztegi batekin elikatzea, iragarpenak egin ditzan. Hasiera batean, Euskal Hiztegiaren Maiztasun Egitura (EHME) [8] corpusean oinarrituz, euskarazko 10000 hitz erabilienekin elikatuta erabili izan du ikasleak Grid3. Erabiltzailea teklatuan idazten joan ahala, sistemak zerrenda horretan oinarritutako 6 iragarpen erakusten ditu, horiek ordena alfabetikoaren arabera proposatuz. Hitzaren igartze-tasari begira, iragarpenen ordena alfabetikoa izateak ez duela laguntzen ikusi dugu; adibidez, erabiltzaileak “a” letra idatzi ondoren, beti proposatzen zaion lehen hitza “abendua” da, honek zentzurik eduki ez arren. Arazo hori gainditzeko, hainbat ahalegin egin ditugu Grid3n, bat esaterako, hitz zerrenda bat sartzea, hitzaren maiztasuna kontuan hartuz. Sortutako zerrenda berrian

---

<sup>3</sup> Esana dugun legez, ikasleak GRID3 sistema erabiltzen du begiradaren bidez ordenagailua maneiatu ahal izateko; eta, zerbait idaztera doanean, GRID3k berak proposatutako iragarpenak erabi en ditu. IGARRITZ erabiltzean, ez gara ariko erabiltzen GRID3ren iragarpenak, baina bai GRID3 sistema, ikasleak ordenagailua maneiatzeko funtsezkoa baita.

hitz errepikapenak sartu dira: EHMEko 100 hitz erabilienak 13 aldiz idatzi dira zerrendan, ondorengo 100 hitz erabilienak 12 aldiz... 10000 hitzetako testu mugara iritsi arte. Modu horretan, Grid3k hitz errepikatuenak lehenago proposatzen ditu, orden alfabetikoaren arazoari aurre eginez. Errepikapen horiek sartu behar izateak, ordea, Grid3k bere sisteman 2800 hitz inguru soilik erregistratuta edukitzea dakar. Gainera, iragarpenak, aurretik idatzitakoaren testuingurua kontuan hartu gabe jarraitzen du, corpusaren maiztasunari bakarrik begiratzen baitio. Beraz, GRID3 tresnaren iragarpenak ikaslearen beharrezanetara apur bat ekartzea lortu arren, mugak oso nabarmenak dira, eta ikasleak dituen beharrezanetatik oso urruti daude.

### 3. TESTUAK IRAGARTZEKO HIZKUNTZA-EREDUAK

Hiztegiekin testuak iragartzeko metodologia baino sistema aurreratuagoa proposa daiteke hizkuntza-ereduak erabiliz, horiek testuingurua kontuan hartzen dutelako, hau da, saioan bertan aurretik idatzitakoa kontuan hartzen du iragarpeneko hautagaiak egokitzeko eta proposatzeko. Hizkuntza-ereduak hizkuntzaren ezaugarriak gordetzen dituzten eredu konputazionalak dira, testuak irakurtzeko, analizatzeko edota sortzeko. Hizkuntza-ereduekin aipatutako hitzen arteko erlazioak ikasteko eta eredu estatistikoak lortzeko, testu multzo handiak behar ditugu. Eredu jakin bat sortzeko prozesu honi entrenamendua esaten diogu, eta zenbat eta datu gehiagorekin entrenatu, orduan eta hizkuntzaren irudikapen estatistiko aberatsagoa lortzeko aukera dugu. Hortaz, esan genezake ereduak honako hauek atzeman ditzaketela: i) corpuseko hizkuntzaren erregulartasunak eta ii) hitzen arteko harreman estatistikoak (harreman estatistiko horiek esanahi- eta estilo-ñabardurenak ere izan daitezkeelarik), distantziak distantzia .

Hizkuntza-ereduaren zereginetara dagokienez, zeregin nagusia da testuinguru zehatza duen testu baten hurrengo balizko hitzei probabilitateak esleitzea. Ataza horri esker, hizkuntza-ereduak, probabilitate handiko hitzak erabiliz, koherentetzat jo daitezkeen testua iragar dezake. Horietako bat ala beste aukeratzeak hurrengo hitzen probabilitateak ere aldatuko dituenek, hitz bakoitzaren ondorengo probabilitate berri horiek kalkulatzeko ahalmenari esker, esanguratsuak izan daitezkeen hitzak iragar ditzake, testuingurua kontuan izanik eta testuingurura etengabe egokituz.

Horren ondorioz, aurrez definitutako hiztegiaren edo hitz-zerrendetan oinarritutako sistemek ez bezala, hizkuntza-ereduek aukera ematen dute iragarpen adierazgarri eta koherenteagoak sortzeko. Horrez gain, hizkuntza-ereduak corpus espezializatuekin entrenatzen edo berrentrenatzen baditugu, erregistro horretara egokitzen da testu iragarpena; besteak beste, hiztegia, hitzen ordena eta testuaren estiloa imitatuz.

Azken urteetako aurrerapenei esker, ikasketa automatikoko metodo zaharren zenbait muga gainditu dira *Transformer* arkitekturarekin sortzen diren hizkuntza-ereduetan [4]. Izan ere, *Transformer* arkitekturaren arreta-mekanismoari esker [10], iragarpenak egiteko, ereduak hitz-sekuentzia jakin bateko hainbat hitzek duten garrantzia haztatzen du. Zehazki, RoBERTa arkitektura noranzko bikoitzeko arreta-mekanismoaz baliatzen denez, ohiko hizkuntza-ereduak ez bezala, aurreko eta osteko testuingurua kontuan izan dezake testuaren erdiko hitzak iragartzeko [9]; aurreko esaldi bat zuzentzerakoan edo aldatzerakoan, esaterako, testu berrira hobeto egokitzeko egin ohi dira horrelakoak.

#### 4. ATAZARI EGOKITZEKO ERABILITAKO CORPUSAK

IGARRITzen iragarpenak sortzeko erabilitako oinarrizko hizkuntza-eredua, HiTZ/roberta-eus-euscrawl-base-cased, askotariko testu motekin entrenatua izan da eta, horren ondorioz, sarritan DBHko ikasle batek erabiltzen duen hizkerarekin bat ez datozen hitzak eta esaldiak iragar ditzake. Horregatik, oinarrizko hizkuntza-eredu horrek sortutako testuak atazara doitzeko, hezkuntzari buruzko testuak bildu ditugu: Gizapedia, ikasleen idazlanak eta hezkuntzari buruzko testuak (Wikipediakoak, Elkarrekoak eta Berriakoak) (ikus 1. taula).<sup>4</sup>

Rola	Corpusa	Dokumentuak	Tokenak	URLa
Hizkuntza-eredua	Euscrawl	12.528k	423M	<a href="https://www.ix.eus/euscrawl/">https://www.ix.eus/euscrawl/</a>
Doiketarako berrentrenamendua	Gizapedia	5.688	714k	<a href="https://gizapedia.org/">https://gizapedia.org/</a>
	Wikipedia	1.962	3,3M	<a href="https://eu.wikipedia.org/wiki/Kategoria:Hezkuntza_Programako_artikuluak">https://eu.wikipedia.org/wiki/Kategoria:Hezkuntza_Programako_artikuluak</a>
	Berria	354	161k	<a href="https://www.berria.eus/ikasgela/gaiak/hezkuntza">https://www.berria.eus/ikasgela/gaiak/hezkuntza</a>
	Elkar	43	1,1M	-
	DBHko ikasleen	1.469	194k	-

<sup>4</sup> Online ez dauden eta erabili ditugun corpusak ataza zehatzak soilik egiteko baimena izan dugu.

	Idazlanak			
--	-----------	--	--	--

### 1. taula: Hizkuntza-ereduko corpusaren eta berrentrenamendurako corpusen estatistikak

Gure hasierako hizkuntza-eredua, bildutako corpora erabilia, ikasle baten hizkerara doitzeko, eredua entrenatzen jarraitu dugu, bere aurrentrenamenduan erabili dugun ataza berbera jarraituz, hots, etengabeko aurrentrenamendu teknika erabiliz. Ataza honetan, testuaren hitzak maskaratu edo ezkutatu egiten dira, eta maskaratutako leku horietan dauden hitzak iragarri behar ditu ereduak. Entrenamenduan, Huggingface liburutegiaren defektuzko balioez gain, honako hiperparametro hauek erabili ditugu hauek dira:  $2e-5$  ikasketa-tasa ( $\gamma$ ),  $0.01$  pisuen gainbehera ( $\lambda$ ) eta AdamW optimizazio algoritmoa erabili dugu  $\beta_1=0,9$ ,  $\beta_2=0,9$  eta  $\epsilon=1e-6$ .

## 5. IGARRITZ WEBZERBITZUA

Hizkuntza-ereduekin eta corpusarekin azaldutako metodologia erabilia, IGARRITZ webzerbitzua sortu dugu, eta nahi duenak erabiltzeko moduan jarri dugu (ikus 1. irudia); Firefox esploratzailea erabiliz hemen proba daiteke: <http://igarritz.clariah.eus>.

IGARRITZek bi atal nagusi ditu: hitzen iragarpenaren eremua eta, bere gainean, hitza aukeratzeko hautagaien testu-kaxa. Iragarpenari dagokionez, sistemak hainbat hitzen aurreikuspena egiten du. Batetik, ordura arteko testua edota idatzitako letrak kontuan izanik, balizko hurrengo 6 hitz erakusten ditu. Proposaturiko hitzen artean, erabiltzaileak behar duen hitzik ez badago, letra bat idatz dezake teklatuarekin edo teklatu birtualarekin. Halaber, IGARRITZek teklatu birtuala jartzeko eremu bat du, zeina kendu edo jarri egin daitekeen, erabiltzaileak hala nahi badu. Sagua edo begirada, aukeratu nahi den hitzaren edo letraren gainean  $0,6$  segundo mantenduz edo klik eginez aukeratzen da, eta kaxa denbora pasatu ahala koloreztatu egiten da, erabiltzaileak zer aukeratzen ari den jakin dezan.

Bestalde, iragarpenak irakurtzean begirada mantenduz gero, nahi ez den hitzen bat aukeratzeko arriskua dagoenez, IGARRITZ-en eragiketak desegiteko botoiak jarri ditugu. Horrela, erabiltzaileak idatzitako letrak ez dira galduko, eta aurreko egoerara bueltatzeko aukera izango du. Botoi horiek aukeratzeko, begirada denbora gehiagoz mantendu behar da, zailagoa izan dadin botoiok nahi gabe aukeratzea.



Bukatzeko, IGARRITZek, idatzitakoa kopiatzeko eta beste programa batean itsasteko edota idatzitakoa ezabatzeko aukerak eskaintzen ditu. Testua nahi gabe ezabatuz gero, atzera egiteko aukera dago eragiketa deseginez (ikus 1. irudia).

## 2. irudia. IGARRITZ aplikazioa eta interakzioen azalpena.



### 5.1. Webzerbitzuko idazketa-sisteman egindako egokitzapenak

Begiradaren jarraipenarekin ordenagailua maneiatzeko, begirada detektatzeko hardwarea dago (*eye-trackerra*: Irisbond edo Tobi, besteak beste), eta, horrez gain, prozesu hori errazten duten programak (Grid3, Easyclick...) ere kontuan izan behar ditugu. IGARRITZ tresna begiradaren jarraipenarekin idazteko diseinatu badugu ere, Diseinu Unibertsalaren kontzeptua gogoan izan dugu, ahalik eta testuinguru gehienetara egokitzeko. Hori argi ikusten da egin dizkiogun aldaketetan eta tresnak eskaintzen dizkigun aukeretan. Aldaketa eta aukera horiek, batez ere, bi fasetan egin ditugu: i) erabiltzailearekin lanean ari den irakaslearen gomendioak jarraituz egin diren aldaketak, hots, aditu baten aholkuz. ii) Erabiltzailearekin zenbait proba egin eta gero egin diren aldaketak.

### 5.2. Erabiltzaileari begira egindako aldaketak (proba aurrekoak)

Honako aldaketak egin ditugu:

- Aukeratzea. Baliabide honi esker, klik egin gabe eta begirada edo sagua botoiaren gainean mantenduz aukeratu ahal izango ditugu iragarritako hitzak. Aukeratzeko, ikasleak berak irisarekin klik egiteko konfigurazio-denbora errespetatu dugu, hots, 0,6 segundo mantendu behar dugu begirada edo sagua iragarpenean, aukeratu ahal izateko.
- Aukeraketa nabarmentzea. Begirada non dagoen jakiteko, IGARRITZeko botoiak kolorez betetzen hasten dira, ezkerretik eskuinera, begirada bertan dagoen bitartean. Botoiak guztiz betetzen direnean egiten dute klik.
- Kopiatu botoia. Testu-kaxan idatzitako testu guztia kopiatzen du, eta beste nonbaitera eraman ahal izatea errazten du.
- Ezabatu botoia. Testu-kaxan dagoen testu guztia ezabatzen du. Botoi hau nahi gabe ez aukeratzeko, denbora gehiago egon behar da begirada edo sagua bere gainean.
- Berregin eta desegin botoiak. Historian aurrera edo atzera joateko botoiak dira. Ezabatu botoiarekin, egindako eragiketak zein iragarpen aukeraketak desegin litezke. Beraz, iragarpen bat oker aukeratuz gero, desegin botoiari emanda, erabiltzaileak idatzi dituen letrak ez dira galduko, eta berriro agertuko zaizkio iragarpen guztiak.
- Botoien tamaina. Iragarritako hitza aukeratzeko, markoa handitu egin da, irisarekin errazago aukeratu ahal izateko.
- Kanpo teklatura jartzeko eta menua ezkutatzeko botoia. IGARRITZ beste teklatu edo sistema batzuetan integratzeko eta pantailan espaziorik ez galtzeko, erabiltzaileak teklatuaren hutsunea ken lezake.
- Puntuazioa zuzendu. Iragarritako hitza aukeratzean, espazioa ere jartzen du. Puntuazioa idaztean, aurretik jarritako espazioa ezabatu egiten da, eta aurreko hitzari lotzen zaio puntuazio-ikurra. Horrek guztiak erabiltzaileari espazioa gehitzen edo ezabatzen ibili beharra aurrezten dio.

### **5.3. Erabiltzailearekin egon ondoren egindako aldaketak**

Honako aldaketak egin dizkiogu IGARRITZi:

- Iragarritako hitzen kokagunea. Probabilitate handiena duten hitzak erdian agertuko dira. Modu horretan, begirada gutxiago mugitzeaz gain, iragarritako hitzak gaintik irakurtzeko begirada nora eraman behar duen jakingo du eta hitz esanguratsuena aukeratzeko ahalegin gutxiago egingo du.

- Botoien tamaina. Botoien horizontaltasuna mugatu eta erdian bertikalago jarri ditugu, begirada bazterretara asko mugitu gabe hitzak irakurri eta aukeratzeko.
- Historiala gorde. Begiradaren kontrola ez denez erabatekoa, gertatu izan da ezabatu botoiari nahi gabe eman izana eta idatzitako guztia galdu izana. Hori saihesteko, ezabatzeko botoiari emateko denbora gehiago ezarri diogu eta, horrez gain, historialean atzera eta aurrera egiteko aukera ere bai. Horrela, zerbait nahi gabe eginez gero, egindakoa erraz berreskura daiteke.

## 6. EBALUAZIOA ETA EMAITZAK

### 6.1. Ebaluazio automatikoa

Hitzak iragartzeko, ezkutatuta edo maskaratuta dauden hitzak igartzeko entrenatu dugu IGARRITZek darabilen hizkuntza-eredua, eta, beraz, aipatu behar da tresnaren helburua ez dela testu osoak sortzea, baizik eta testuaren une jakin batean hitz probableen zerrenda proposatzea. Ondorioz, testu sorkuntzan erabiltzen diren ebaluazio metrika eta prozesu ohikoenak ez dira aproposak gure tresna ebaluatzen. Metrika horien zenbait adibide dira: BLEU, Rouge eta perplexity neurriak, besteak beste. Hori horrela, IGARRITZek darabilen hizkuntza-ereduaren iragarpen-gaitasuna neurtzeko, gure atzari lotutako ebaluazio-prozesu egokitua diseinatu dugu.

Horretarako, hizkuntza-ereduaren ezaugarriez gain, atazan, tresnaren erabileran eta zenbait kasuistikan oinarritu gara. Ebaluaziorako testu multzoan dauden hitz guztien asmatze-tasa neurtzea nekeza denez, testu bakoitzean gutxienez hiru letraz osatutako hitzen % 15 maskaratu dugu. Gutxieneko letra kopurua ezarri dugu hitz luzeak iragartzen zailagoak direlako. Bestalde, testuak moztu egin ditugu, adibide berriak sortzeko eta testua erabat burutu gabe dagoela irudikatzeko. Azken kasu horietan, azken hitza beti maskaratu dugu, baita bestelako hitzen % 10 ere. Bestelako hitzak ere enmaskaratzeak behartzen du hizkuntza-eredua erregularitasun sendoagoak eta orokortuagoak ikastera eta ebaluatzen, eta ez da soilik mugatzen asmatu behar duen uneko hitzera.

Ebaluazio instantziak prestatuta, maskaratutako hitzen iragarpena eta ebaluazioa egin dugu. Erabiltzailearen erabilpena erreplikatzeko, hitz guztien artean probabilitate handiena duten sei hitz hartu ditugu. Hitz-multzo horretako zerrendan jatorrizko hitza ez balego, hitzaren hurrengo letraren arabera bahetu ditugu hitzak, beste sei hautagai lortzeko. Jatorrizko hitza hautagaien zerrendan agertu arte edota hitza guztiz idatzi arte errepikatu dugu prozesu hori.

Horrekin, alde batetik, ereduaren zehaztasuna neurtu dugu. Bestalde, ereduak zenbat hitz aurreztuko lukeen ere neurtu dugu. Azken neurri hori oso garrantzitsua eta esanguratsua da ataza honetan, helburu nagusia ikaslearen testuen idatzizko produkzioa handitzea delako.

## 6.2. Kasuan oinarritutako eskuzko ebaluazioa

Ikasleak, Grid3 tresnaren iragarpenak erabiliz (euskarazko hitz zerrendak sartuta), 4 testu idatzi ditu, eta bideoan grabatu dugu prozesu osoa<sup>5</sup>. Ondoren, bideoen laguntzaz, testu horietan erabilitako hitzak bildu egin ditugu, eta ikasleak egindako modu berean idatzi ditugu IGARRITZ erabilia, hots, hitzen testuingurua erabat mantenduz. Kasu batzuetan, ikasleak hitz bat osorik idatzi du, baina, gero, ezabatzea erabaki; horrelakoetan, hitza kontatu egin dugu. Ostera, letra solteak edo hitza osatzera iristen ez diren letra-errenkadak ezabatu baditu, horiek ez ditugu bildu, ez baitugu horrelakoak aurreikusterik. Bi arau horiek kontuan izanda, ebaluazioa burutzeko, bi motatako hitzen azterketa egin dugu:

- A. Testu iragarleak osorik aurreikusten dituen hitzak: ikasleak idatzitako hitza iragarleak zehatz-mehatz aurreikusi duenean, zenbatgarren letran proposatzen duen jaso dugu, baita zenbat letra aurrezteko parada ematen duen ere. Adibidez, “irakasleak” hitza idatzi nahi izan dugunean, IGARRITZek “ir” idatzita ematen digu nahi izan dugun emaitza; beraz, kasu honetan, 2 letra idatzi behar izan ditugula etiketatu dugu, eta 8 letra aurreztu ahal ditugula.
- B. Hitza osorik idazteko moldaketaren bat behar dutenak: iragarleak ez ditu hitz guztiak asmatzen, baina, asmatzen ez dituen horien artean, batzuetan iragarpenek antz handia dute idatzi nahi den hitzarekin, eta, iragarpenari moldaketaren bat eginda, nahi dugun hitza idaztera hel gintezke. Horrelakoetan, moldaketak eginda zenbat letra aurreztu ahal ditugun jaso dugu. Posible da iragarleak egindako iragarpen batek baino gehiagok moldaketak egiteko aukera ematea; kasu horretan, letra gehien aurrezteko aukera ematen duena erabili dugu. Adibidez, “oinarritzeko” hitza ez du iragartzen IGARRITZek, baina, “o” idatzi dugunean, “oinarrituta”, “oinarriturik” eta “oinarrituz” agertu dira 6 aurreikuspenen artean. Kasu honetan, “oinarrituz” hitza hartu dugu, “uz” ezabatu ondoren “zeko” gehitzeko. Horrela, 12 letra dituen hitzetik, “o” eta “zeko” idatzi ditu (5 letra), eta “uz” ezabatzeko birritan egin behar denez klik, orotara 5 letra aurreztu dituela jaso dugu. Beste bi aurreikuspenak ez ditugu hautatu, letra bat gehiago ezabatu beharko genukeelako.

---

<sup>5</sup> Testuetako bat, 2. testua, ikasleak aurreko egunean hasitako gutun bat da; bideoan gutunaren bigarren zatia soilik grabatu denez, zati hori hartuko dugu testutzat. Bestalde, 3. eta 4. testuak testu beraren parte dira, baina, ikasleak bi egunetan idatziak direnez, bakoitza testu bat bailitzan hartu dugu.

Eskuzko ebaluazioari dagokion hitzen azterketa bi eratarata egin dugu: orokorrean eta testuka. Horrez gain, ondoren, hitzen letra-kantitatea hartu dugu aldagaitzat, eta lau testuetan IGARRITZek egindako aurreikuspenak zer-nolakoak izan diren aztertu dugu. Horretarako, hitzak lau multzotan sailkatu ditugu: 0-3 letra, 4-6 letra, 7-9 letra eta 10 letra edo gehiago dituztenak.

Modu honetako ebaluazioa egin dugu, idazmenaren ekoizpenean lagundu beharreko ikasleari idazteko lan gehiago ez ematearren eta ikaslearen erabilerak ez duelako iragarpen ahalmena neurtzen. Beraz, benetan neurtu duguna da tresnen iragarpen-ahalmena, eta ez ikaslearen erabilera. Hori bai, ikasleari zuzenean galdetu diogu eta ikasleak erantzun du askoz azkarrago idazten duela IGARRITZekin, nahiz eta oraindik orain funtzio batzuk egokitu behar ditugun erosoago erabil dezan.

### 6.3. Emaitzak

Lan honetan lorturiko emaitza nabarmenena IGARRITZ sistema sortu izana da <http://igarritz.clariah.eus>.

IGARRITZ-en erabilpena antzeratzen duen ebaluazio automatikoarekin bi eredu testatu ditugu; alde batetik, oinarriko ereduak, HiTZ/roberta-eus-euscrawl-base-cased, eta, beste aldetik, hezkuntzari buruzko testuekin doitutako ereduak. Oinarrizko hizkuntza-eredua erabilita, 70,9 mila letra aurrezten dira (% 19,57 zehaztasunarekin (*accuracy*)); hezkuntzako corpusarekin doitutako hizkuntza-ereduarekin, ordea, 74,5 mila letra aurrezten dira (% 19,89 zehaztasuna), 3,6 mila letra gehiago, hots, % 5 gehiago. Doitutako ereduak lehenago iragartzen du hitz aproposa, eta, horren ondorioz, erabiltzaileari letra gehiago aurrezten dizkio, bi ereduak antzeko zehaztasuna izan arren.

IGARRITZ-en eskuzko ebaluazioari dagokionez, 1. taularen arabera, emaitzak honakoak dira: Aztertu ditugun 4 testuetan, IGARRITZekin emaitza askoz hobea lortu ditugu ikasleak orain arte erabilitako GRID3 tresnaren iragarpenekin baino. Testuen datu orokorrei erreparatzen badiegu, asmatutako hitzen bidez aurrez genitzakeen letrak % 50,07 dira IGARRITZekin; GRID3ren iragarpenekin, aldiz, % 18,09. Azken honekin alderatuta (% 13,9), moldaketekin aurrez genitzakeen letra kopurua jaitsi egiten da IGARRITZekin (% 7,77), baina joera hori positiboa da, asmatutako hitzekin hitzak aurrezten joan ahala, normala baita moldaketa gutxiago egiteko beharra izatea. Orokorrean ez ezik, testu guztietan hobetu ditu emaitzak IGARRITZek. Hala ere, kontuan hartu beharrekoa da testuen artean dagoen aldea. Esaterako, 2. testuan, asmatutako hitzei dagokionez, 238 letratatik 149 aurreztu ahal ditugu, hots, % 62,61, eta, 4. testuan, ordea, ehunekoa dezente jaisten da, % 26,67ra, hain zuzen. 2. testua aurrez idatzitako gutun baten jarraipena denez eta IGARRITZi

aurreko zati hori erantsi diogunez hasieran, ondoriozta genezake aurreko testuingurua lagungarria izan daitekeela oso iragarpenak hobetzeko. Halaber, esana dugunez, 4. testua ere beste testu zati baten jarraipena da, baina honakoan, ordea, ez ditugu emaitza hain onak lortu. Beraz, testuingurua lagungarria dela antzeman badezakegu ere, edukiak eta gaiak eragin zuzena dutela dirudi.

**2. Taula.** Eskuzko ebaluazioaren emaitzak; GRID3ren iragarpenekin eta IGARRITZekin aurrez daitezkeen letrak

	Testuko letrak	Aurreztutako letrak		Moldaketetan aurreztutakoak	
		GRID3	Igarritz	GRID3. (bir.)	Igarritz (bir.)
<b>1. testua</b>	108	20	55	16	16
<b>2. testua</b>	238	45	149	54	9
<b>3. testua</b>	233	56	107	22	20
<b>4. testua</b>	90	0	24	4	7
<b>Orotara</b>	<b>669</b>	<b>121</b>	<b>335</b>	<b>93</b>	<b>52</b>
<b>Aurreztutakoa</b>		<b>% 18,09</b>	<b>% 50,07</b>	<b>% 13,90</b>	<b>% 7,77</b>

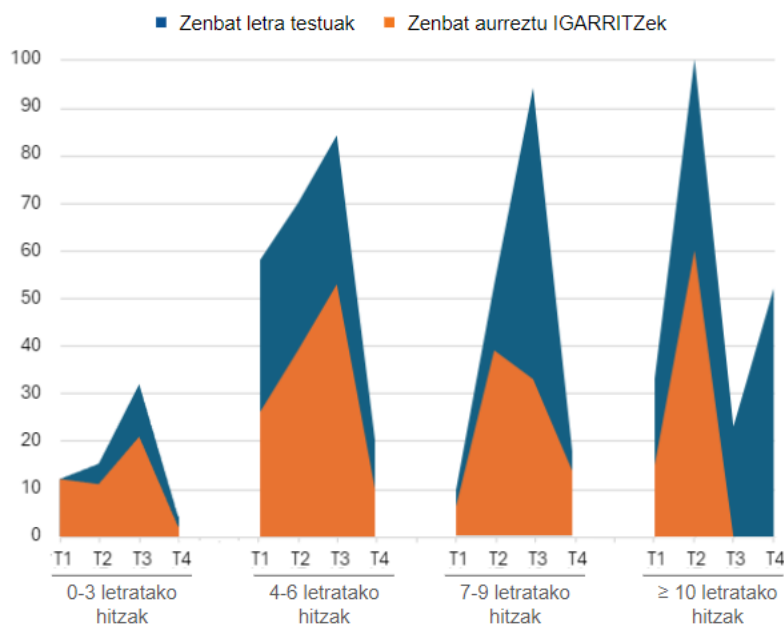
Zenbat letra aurrez daitezkeen aztertu ondoren, IGARRITZ-en iragarpen eta aurrezte-aukerak zer hitz motatan izan diren aztertu dugu; ikus 2. taula. Letra kantitatea kontuan izanda, ikus genezake 0-3 letra dituzten hitzak direla maiztasun txikiena dutenak (% 9,42). Alabaina, horietan du IGARRITZ-ek iragarpen-tasa eraginkorrena, letren % 73,02 aurreztea ahalbidetzen baitugu. Gainontzeko 3 multzoetan, letra kantitatea orekatuagoa izan da, testuko 4-6 letrako hitzetan letren % 34,38 aurreztu dira, 6-9 letra dituztenetan % 25,26 eta 10 letratik gorakoetan % 30,94. Aurrezteko aukerei erreparaturaz gero, 4-6 eta 6-9 letrako hitzetan aurreztutako datuak oso antzekoak dira, hurrenez hurren letren % 54,78 eta % 52,66 aurrezteko parada izan baitugu. Haatik, ehunekoa dezente jaisten da 10 letratik gorako kasuan, % 35,75era, hain zuzen. Gainera, testuka gertatu dena begiratzen badugu, argi eta garbi ikusten dugu testuen artean alde esanguratsua egon dela. Emaitzetan ikusten denez, 1. testuan, 10 letratik gorako hitzek 32 letra biltzen dituzte, eta IGARRITZ-ek 14 aurrezteko aukera eman digu (% 43,75), eta 2. testuan emaitzak are hobekak dira, 100 letratatik 60 (% 60) aurrezteko aukera ematen baitugu. Baina emaitzak erabat bestelakoak dira beste bi testuetan; batean zein bestean, hitzek 23 eta 52 letra bildu arren, ez dugu izan letraren bat aurrezterik testu moldaketarik egin ezean. Testuen arteko alde hori oso ondo ikusten da 3. irudian. Irudi horretan, ikus genezake testuaren letra kopuruak gora egiten duenean, IGARRITZ-ek letrak aurrezten jarraitzen duela, hots, asmatze-tasari eutsi egiten dio, forma bereko mendi urdin eta laranjak irudikatuz. Alabaina, batzuetan ez da bete hori, eta ikus genezake nola multzo bereko hitzetan grafiko urdin eta laranjak forma bera izatetik erabat bestelakoa

izatera igaro dela. Multzo bereko hitzen artean horrelako aldea ikusi ahal izateak berriro ere ideia bera azpimarratzera garamatza: IGARRITZek aurrez duen testuinguruak eta idazten ari garen eduki edo gaiak eragin zuzena dute iragarpenak eraginkorrak izateko.

**3. Taula** Hitz moten arabera IGARRITZekin aurrez daitekeen letra kopurua, moldaketak kontuan izan gabe

	1-3 letra		4-6 letra		6-9 letra		10 ≤ letra	
	zenbat	aurreztu	zenbat	aurreztu	zenbat	aurreztu	zenbat	aurreztu
<b>1. testua</b>	12	12	57	25	7	4	32	14
<b>2. testua</b>	15	11	70	39	53	39	100	60
<b>3. testua</b>	32	21	84	53	94	33	23	0
<b>4. testua</b>	4	2	19	9	15	13	52	0
<b>Orotara</b>	<b>63</b>	<b>46</b>	<b>230</b>	<b>126</b>	<b>169</b>	<b>89</b>	<b>207</b>	<b>74</b>
<b>%</b>	<b>9,42</b>	<b>73,02</b>	<b>34,38</b>	<b>54,78</b>	<b>25,26</b>	<b>52,66</b>	<b>30,94</b>	<b>35,75</b>

**3. irudia:** Aurreztutako letrak 4 testuetan



**7. ONDORIOAK ETA ETORKIZUNeko LANA**

Lan honetan euskarazko testu iragarpena egiteko eta begiradarekin idazteko programekin bateragarria den lehen web ingurune egokituia sortu dugu adimen artifizialeko teknikak erabiliz: IGARRITZ (ikus: <http://igarritz.clariah.eus/>). Testu iragarpena egiteko *Transformer* arkitektura erabiltzen du [4], eta RoBERTa kodifikatzaile arkitekturan [5] oinarritutako HiTZ/roberta-eus-

euscrawl-base-cased hizkuntza-eredua [6]. IGARRITZ behar duen DBHko ikasle batek tresna probatu eta hobetzen lagundu digu. Ikaslearen beharrezanetara egokitzeko, tresnak darabilen hizkuntza-eredua berariaz sorturiko hezkuntzari buruzko corpusekin berrentrenatu dugu, etengabeko aurrentrenamendua (*continual-pretraining*) deritzon teknika erabiliz. Hizkuntza-teknologian oinarrituta dagoen tresna da IGARRITZ. Ezgaitasun motorra dutenentzat diseinatu da, euskaraz idazteko; eta begiradaren bidezko komunikazioa errazteko eta areagotzeko sistema da, doakoa eta online atzitzen dena. Horrez gain, baliabide interaktiboa da; erabiltzaileari hurrengo hitza idazteko hainbat hautagai iragartzen dizkio IGARRITZek, eta, letrak sartu ahala, iragarpenak egokitzen ditu, idatzitako testuingurura eta sartutako letretara egokituz.

Egindako ebaluazioei begira, ebaluazio automatikoari esker, IGARRITZek iragarpen hobekak zein hizkuntza-eredurekin egiten dituen ebatzi ahal izan dugu. Emaizten arabera, hezkuntzako corpusarekin doitutako hizkuntza-ereduak lehenago iragartzen du hitz aproposa, eta, horren ondorioz, erabiltzaileari letra gehiago aurrezten dizkio, oinarritzko hizkuntza-ereduak antzeko zehaztasuna izan arren. Bestalde, IGARRITZ-en iragarpen-ahalmena eskuz ebaluatu dugunean, ikusi dugu IGARRITZekin emaitza askoz hobekak lortu direla ikasleak orain arte erabilitako GRID3 tresnaren iragarpenekin baino, eta hobekuntza hori aztertutako testu guztietan egon dela. Hala ere, eskuzko ebaluazio honetan, testu batetik bestera aldea dagoela ikusi dugu, eta alde horren zio nagusiak bi izan daitezkeela: i) aurrez testu zati bat (testuingurua) idatzita izateak aurreikuspenak hobetu egiten dituela eta ii) testuaren gaiak eta edukiak eragin zuzena dutela aurreikuspenetan. Nabarmendu nahi dugu lan honetan ebaluatu duguna tresnen iragarpen-ahalmena dela, eta ez ikaslearen erabilera erreala. Edonola ere, ikaslearekin izan dugun harreman zuzenari esker, zuzenean behatu dugu erabilera; gainera, zuzenean galdetu zaio haren iritziaz, eta ikasleak erantzun du askoz azkarrago idazten duela IGARRITZekin, nahiz eta oraindik orain funtzio batzuk egokitu behar diren erosoago erabil dezan. Horrez gain, aipatu nahi dugu ikaslearen tutoreak ere adierazi digula IGARRITZek asko laguntzen diola eta etorkizuneko ikasturteetako eginkizun eta gaietara gehiago egokitzea interesgarria litzatekeela.

## **Eskertza**

Lan honek Eusko Jaurlaritzako Hezkuntza Saileko laguntza jaso du. Horrez gain, eskerrak eman nahi dizkiogu Iñaki Orberi eta Lorea Aretxagari ikerketa hau sustatzeko emandako laguntza guztiarengatik, baita IGARRITZ erabili duen ikasleari eta bere familiari. Horrez gain, lan hau CLARIAH-EUS azpiegituraren lanen barruan egin da, eta HiTZ zentroak ere lagundu du, baliabideak



eta webgunea eskainiz. Bukatzeko, ISEI-IVEIri, ELKARri eta Gizapediari ere eskerrak eman nahi dizkiegu, beraiek sortutako testuak IGARRITZ berrentrenatzeko uztearren.

## BIBLIOGRAFIA

- [1] Eusko Jaurlaritza. (2022) *Hezkuntza-sisteman “komunikazio sistema handigarriak edo/eta ordezkioak” (KSHO) inplementatzeko protokoloa*. Eusko Jaurlaritza Hezkuntza Saila. [https://www.euskadi.eus/contenidos/informacion/inn\\_edu\\_inc\\_neapoyoeducativo/eu\\_def/adjuntos/Protocolo\\_SAACS\\_e.pdf](https://www.euskadi.eus/contenidos/informacion/inn_edu_inc_neapoyoeducativo/eu_def/adjuntos/Protocolo_SAACS_e.pdf)
- [2] Lloyd, L. & Karlan, G. (1984). Non-speech communication symbols and systems: where have we been and where are we going? *Journal of mental deficiency research*, 28 (Pt 1), 3–20. <https://pubmed.ncbi.nlm.nih.gov/6232388/>
- [3] Klaus, B., Aigner, B., & Veigl, C. (2019). AsTeRICS Grid—a flexible web-based application for Alternative Communication (AAC), environmental and computer control.
- [4] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł. & Polosukhin, I. (2017). Attention is all you need. *Advances in neural information processing systems*, 30.
- [5] Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., Levy, O., Lewis, M., Zettlemoyer, L. & Stoyanov, V. (2019). Roberta: A robustly optimized bert pretraining approach. *arXiv preprint arXiv:1907.11692*.
- [6] Artetxe, M., Aldabe, I., Agerri, R., Perez-de-Viñaspre, O., & Soroa, A. (2022). Does Corpus Quality Really Matter for Low-Resource Languages?. *arXiv preprint arXiv:2203.08111*.
- [7] Luna, M.R. (2013): Tecnología y discapacidad: Una mirada pedagógica. *RDU, Revista Digital Universitaria*, 14(12), 1-19, <http://www.ru.tic.unam.mx/handle/123456789/2177> [20XX-XX-X].
- [8] Euskal Hiztegiaren Maiztasun Egitura. (Eguneratua: 2024-01-10). Euskal Herriko Unibertsitatea. <https://www.ehu.eus/ehg/ehme/datu2hitz.htm>
- [9] Wang, A., & Cho, K. (2019). BERT has a mouth, and it must speak: BERT as a Markov random field language model. *arXiv preprint arXiv:1902.04094*.
- [10] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, & Illia Polosukhin (2017). Attention is All you Need. In *NIPS* (pp. 5998–6008).
- [11] Garay-Vitoria, N., & Abascal, J. (2006). Text prediction systems: a survey. *Universal Access in the Information Society*, 4, 188-203.
- [12] Altuna, B., Iruskietta, M., Estarrona, A., Farwell, A., Arriola, JM., Alkorta J., and Arregi., X. (2022). CLARIAH-EUS: Building a Cross-border CLARIAH Node for the Basque Language. *Digital Humanities Conference November 23-25*. Budapest.