

Euskararen Atlas Prosodiko Interaktiboa (EAPI)*

Atlas Interactivo de la prosodia del euskera

Interactive Atlas of Basque Prosody

Aitor Iglesias Chaves

UPV/EHU

aitor.iglesias@ehu.eus

<https://orcid.org/0000-0001-5381-5957>

Aintzane Etxebarria Lejarreta

UPV/EHU

aintzane.etxebarrai@ehu.eus

<https://orcid.org/0000-0001-8418-6276>

Naia Eguskiza Sánchez

UPV/EHU

naia.egusquiza@ehu.eus

<https://orcid.org/0000-0002-8928-7006>

Recibido / Noiz jaso den: 18/06/2021

Aceptado / Noiz onartu den: 08/11/2021

Laburpena

Lan honen helburu nagusia da *Euskararen Atlas Prosodiko Interaktiboa* (EAPI) ezagutaraztea. EAPI euskarazko prosodiaren lehen tresna multimedia eta interaktiboa da. Atlas honetan erabiltzen diren datu-baseak EDAK eta EAS dira, UPV/EHUko EUDIA ikerketa-taldeak sortutako corpusek osatutakoak. Atlasak prosodiari buruz egindako analisiak erakusten ditu: azentuaren kokapena euskararen aldaera eta silaba guztietan; oinarrizko esaldien intonazio eredu esanguratsuenak, euskara estandarrean eta euskalkietan; eta bat-bateko testuen eta testu irakurrien azterketa prosodikoa egiteko bidea irekitzen du. Atlas honekin, prosodiari dagokionean, euskara nazioarteko beste hizkuntza batzuen parean kokatzen da.

Gako hitzak

Atlas; prosodia; euskara, interaktiboa, intonazioa; azentua.

Sumarioa

1. SARRERA. 2. JUSTIFIKAZIOA. 2.1. Atlasen beharra. 2.2. Helburuak. 3. PLATAFORMAREN EZAUGARRIAK. 4. INFORMAZIO PROSODIKOA ETA SOZIO-GEOGRAFIKOA. 5. DATUEN EZAUGARRIAK. 5.1. Corpora. 5.2. Transkripzioa, etiketaztea, segmentazioa eta datuen anotazioa. 6. ERABILERA ETA APLIKAZIOAK.

* Lan hau ondoko erreferentziadun proiektuen barnean egin da: PPGA20/12 eta IT 1208/19.

Resumen. El objetivo principal de este trabajo es dar a conocer el *Atlas interactivo de la prosodia del vasco / Euskararen Atlas Prosodiko Interaktiboa* (EAPI). EAPI es la primera herramienta multimedia e interactiva de la prosodia en lengua vasca. Este atlas cuenta con una base de datos compuesta principalmente por los corpus EDAK y EAS creados por el grupo de investigación EUDIA de la UPV/EHU. El atlas muestra los análisis realizados sobre la prosodia: la ubicación del acento en todas las variedades y sílabas del euskera; los modelos de entonación más significativos de las oraciones básicas, tanto del euskera estándar como de los dialectos; y abre el camino para el análisis prosódico de textos espontáneos y leídos. Con este atlas se pone el euskera a nivel internacional a la par de otras lenguas desarrolladas en el campo de la prosodia.

Palabras clave. Atlas; prosodia; euskera; interactivo; entonación; acento.

Abstract. The main objective of this work is to present the *Interactive Atlas of Basque Prosody / Euskararen Atlas Prosodiko Interaktiboa* (EAPI). EAPI is the first multimedia and interactive tool for prosody in Basque. This atlas has a database made up mainly of the EDAK and EAS corpora created by the EUDIA research group of the UPV/EHU. The atlas shows the analyses carried out on prosody: the location of the accent in all the varieties and syllables of Basque; the most significant intonation patterns of basic sentences, both in standard Basque and in dialects; and opens the way for the prosodic analysis of spontaneous and read texts. With this atlas, Basque is placed on an international level on a par with other languages developed in the field of prosody.

Keywords. Atlas; prosody; Basque; interactive; intonation; accent.

1. Sarrera

Euskararen Atlas Prosodiko Interaktiboa euskararen prosodia erakusten duen tresna multimedia eta interaktiboa da (aurrerantzean zenbaitetan EAPI¹). EAPI argitaratu bitartean ez da egon euskarazko tresnarik, beraz, proiektu berritzailea da; proiektu dinamikoa da, eta denboran zehar ondoz ondoko etapetan aberastu daiteke eta aberastu egingo da. EAPI publiko orokorrari eredu prosodikoaren forma esanguratsuen hautaketa erakusten dion dibulgazio proiektua da. Atlas hau testu mota ezberdinen prosodia ikusteko eta entzuteko aukera ematen duen webgune multimedia eta interaktiboa da. Oinarrian datu-base bat du (ikusi 5. atala), nagusiki EUDIA ikerketa-taldeak batutako *Euskara Dialektalaren Ahozko Korpusa* (EDAK, Aurrekoetxea et al., 2011) eta *Euskararen Atlas Soziolinguistikiko-dialektologikoa* (EAS) corpusek osatuta. Corpus hauek, bereziki, aldaera diatopiko, diastratiko eta diafasikoen ikerketa segmentalak eta suprasegmentalak egiteko balio dute. Hala ere, belaunaldien arteko euskararen barietate aztertzeke xedearekin datu gehiago jaso behar izan dira hizkuntzaren mailak (fonetiko-fonologikoa, morfologikoa, sintaktikoa eta semantikoa) kontuan hartuta. Datu hauek guztiak modu bateratu batean transkribatu, segmentatu eta etiketatu egin dira. Hauek dira jasotako datuak:

- Azentuaren aldakortasuna barietate guztietan.
- Oinarrizko esaldien intonazioa.
- Emozioak eta jarrerak barietate adierazgarrietan.
- Bat-bateko testuak eta testu irakurriak.

¹ <http://eudia.ehu.es/eapi/>

Prosodia, ahozko hizkuntzaren elementu suprasegmentalen multzoa da. Etorkizuneko ikerketek, prosodian fokua jarrita eta atlas interaktibo hau erabilita, ahozko hizkuntzaren elementu suprasegmentalen azterketen emaitzak argitara eman ahal izango dituzte. Prosodiari dagokionez, hiru azterketa mota egiten dira: fonologikoa, fonetikoa eta pertzepzio-azterketak. Alde batetik, bariedade desberdinetako azentuaren kokağunea deskribatzen da eta hizkuntza estandarerako arau-irizpideak ezartzen dira (Gaminde et al., 2014) betiere bariedade ezberdinetako azentu-moldeak eta dagoen heterogeneotasuna kontuan hartuta (Aurrekoetxea et al., 2013; 2014). Bestalde, euskararen oinarritzko esaldien (adierazpenekoak, bai/ez galderak, NZ galderak, hautazkoak, oihartzun galderak eta zalantzazkoak) intonazio-eredu adierazgarrienak zehazten dira, bai hizkuntzaren aldaeretan, bai estandarerako proposatzen dugun eredian ere (Etxebarria & Eguskiza, 2018). Horrekin batera, euskaraz adierazitako jarreraren, emozioen eta bokatiboen parametro prosodiko garrantzitsuenak deskribatzen dira (Gaminde et al., 2017b, Etxebarria et al., 2018). Azkenik, euskarazko testu-mota ezberdinen prosodia ikertzeko bidea ireki berri da (Etxebarria et al., 2017, Gaminde et al., 2017), bereziki korapilatsua dena, irizpideak erabat definituta ez daudelako. Horixe da etorkizunari begira jadetse nahi den beste helburuetako bat.

2. Justifikazioa

Aurretik esan dugun moduan, EAPI argitaratu den arte, euskarak ez du atlas prosodiko interaktiborik izan, baina horrelako eredu bat ez da zerotik abiatzen; izan ere, euskaraz atlas linguistikokoak egon badaude eta beste hizkuntza batzuetan ere atlas prosodiko interaktiboak aspaldi garatu ziren, atal honetan zehaztuko den moduan.

Atlas linguistikoa mapa-multzo bat da, non inkestetako datu linguistikokoak aurkezten diren. Euskarari dagokionez, Bonaparte (1863) izan zen datu linguistikokoak mapa batean irudikatzen aitzindaria; ez zuen atlasik egin, baina jasotako datuak mapa batean islatu zituen. Izenburuan atlas hitza agertu ez arren, hizkuntza atlas bezala har daitekeen lehen lana Azkueren *Erizkizundi hirukoitza* da (1925), bere osotasunean 1984 arte argitaratu ez zena (Etxaide, 1984). Lan horretan, 350 mapa daude. Atlas terminoari erreferentzia bikaina egiten dioten lehen lanak Allièresenak (1978, 1986) dira. Bertan, Bourciezek (1895) bildutako datuak lantzen dira, geroago Aurrekoetxea, Videgain & Iglesiasek (2004, 2005) osorik argitaratutakoak. Ordena kronologikoari jarraituz, Aranzadi Zientzia Elkartek egindako liburuki biak ditugu (1983, 1990), kasu honetan atlas etnolinguistikoa. Gaminderen (1984, 1985a, 1985b, 1985c, 1985d eta 1985e) aditzari buruzko lanak ere aipatu behar dira, hasieran proposatutako definizioari hertsiki lotuz gero, atlas linguistiko gisa ere har daitezkeenak. Azkenik, Gotzon Aurrekoe-

txea eta Xarles Videgainen eskutik, Euskaltzaindiak (Euskaltzaindia, 2008-2019) egindako atlasa azpimarratu behar da, orain arteko handiena, dagoeneko 10 liburu-ri dituen eta oraindik osorik argitaratu gabe dagoena.

2.1. Atlasen beharra

Euskararen atlas linguistikoaren sintesi honetan ikus daitekeen bezala, euskaraz ez dago atlas prosodikorik, prosodiari buruzko lan ugari argitaratu diren arren (Elordieta & Hualde, 2104; Etxebarria et al., 2016; Gaminde et al., 2017a etab.). Hori dela eta, EAPIk orain arte euskararen azterketan hutsik egon den espazio bat bete nahi du, zehazki prosodiarena, XXI. mendean kokatua, non datuen digitalizazioa eta interakzioa guztiz funtsezkoak diren nazioarteko ikerketa aurreratu- en abangoardian egon ahal izateko.

Testuinguru horretan, esan beharra dago prosodiari buruzko ikerketa gara- tuagoak dituzten hizkuntzek azentuari eta intonazioari buruzko webgune es- pezifikoak dituztela (portuges², hizkuntza erromanikoak³, AEBko baritateak⁴, koreera⁵ etab. besteak beste). Hitzen azentuari dagokionez, nazioarteko errefe- renteetako bat Stress Typ2⁶ webgunea da, Harry van der Hulst eta Rob Goedemans zuzendari- ek zuzendua; biak dira 510 hizkuntzaren sistema metrikoari buruzko informazioa duen datu-base tipologiko honen bultzatzaile nagusiak, eta horien artean euskara ere bada. Euskararen inguruko eremu erromanikoko hizkun- tzetan, aspaldi hasi ziren atlas prosodikoak eraikitzen AMPERren proiektuarekin (Atlas Multimedia de la Prosodia del Espacio Románico). Webgunean aipatzen den bezala (Martínez Celdrán & Fernández Planas, 2003-2018; 2005), nazioar- teko proiektu bat da, prosodiari buruzkoa, XX. mendearen amaieran Stendhal Grenoble Unibertsitateko Centre de Dialectologieko Michel Contini dokto- rearen bultzadaz sortu zena. AMPERrek hiru parametro prosodikoak aztertzen ditu: oinarritzko maiztasuna (F0), iraupena eta intentsitatea. Bere helburu nagusia hiz- kuntza erromanikoen prosodia aztertzea eta emaitzak mapetan islatzea da. Mapa horiek internet bidez kontsulta daitezke, bisualki eta pertzepzioz. Proiektu honek adar ugari ditu⁷, Europa eta Latinoamerikako hainbat herrialdetan implementa-

² <http://ww3.fl.ul.pt/LaboratorioFonetica/InAPoP/demo/index.htm>

³ <http://prosodia.upf.edu/iari>

⁴ <http://aschmann.net/AmEng>

⁵ <http://www.linguistics.ucla.edu/people/jun/ktobi/k-tobi-V2.html>

⁶ <http://st2.ullet.net/>

⁷ AMPER España e Iberoamérica [<http://stel.ub.edu/labfon/amper/cast/index.html>]; AMPER-Can (Canarias y Cuba) [<http://labfon.webs.ull.es/proampercan/index.html>]; AMPER-Andalucía y Extremadura [http://www.amprae.es/proyecto_amprae/];

tzen ari baitira. *Atles interactiu de l'entonació del català* ere aipatzekoa da (Prieto & Cabré, 2007-2012). Horiek guztiak nazioarteko erreferentziak dira, eta nahitaezko kontsulta edozein ikerketa egiterako orduan.

Prosodiari buruz egin diren lanek, batez ere alboko hizkuntzetan egindako intonazioaren gainekoek, erraztu dituzte euskarari buruzko ikerketak arlo honetan, nahiz eta oraindik beharrezkoa ikusten den testuen prosodia aztertzea eta denon eskura dauden tresna interaktiboak sortzea.

EAPI euskararen prosodiari buruzko informazio esanguratsuen batzen duen lehenengo atlasa da, eta, beraz, erreferentzia izan behar dena gai honen gainean egingo diren ikerketetan.

2.2. Helburuak

Proiektu honen helburu nagusia euskararen atlas multimedia bat sortzea da. Atlas honetan Euskal Herriko hainbat herritako eredu prosodikoaren forma adierazgarrien aukeraketa grafikoki irudikatzeaz gain, entzun ere ahal izango dira. AMPER (Fernández Planas, 2008) bezala, dibulgazio proiektua da, publiko ororentzat interneten eskainia. Oso garrantzitsua da prosodiari buruzko lanak multimedia euskarrian jartzea, ikertzaileei aztertutakoaren errealitatera iristeko beharrezkoa den proba enpirikoa ematen baitie (Fernández Planas, 2005). Atlasak ikertzaileei datu-base bat eskaintzeaz gain, mapa multimedia bat eta batutako datuak entzuteko aukera ere eman behar die. EAPIren helburu nagusira heltzeko honako bost azpihelburu bete behar dira lehenago:

- 1) Atlaserako webgune bat sortu.
- 2) Atlas osatzen duten datuen anotazio, etiketatze eta transkripzio tresnak garatu.
- 3) EUDIA ikerketa-taldeak dituen corpusetako datuen anotazioa, etiketatzea eta transkripzioa atlasean inplementatzeko.
- 4) Datu berrien bilketa atlasaren corpusa osatzeko, batez ere, hizkuntzaren profesionalen (esatariak, kazetariak...) ahozko datuak.

AMPER-Cat (catalán y castellano) [http://stel.ub.edu/labfon/amper/cast/index_ampercat.html];
 AMPER-Astur (asturiano, gallego-asturiano y castellano de Asturias) [<http://www.unioviado.es/labo-fone/>];
 AMPER-Gal (gallego) [<http://ilg.usc.es/amper/>];
 AMPER-Port (portugués) [<http://pfonetica.web.ua.pt/AMPER-POR.htm>];
 AMPER-Ital (italiano) [<http://www.lfsag.unito.it/amper/index.html>];
 AMPER-Rom (rumano) [<http://amprom.uaic.ro/indexamper.php>];
 AMPER-Friül (friül) [<http://stel3.ub.edu/labfon/amper/friul/index.html>].

- 5) Parametro prosodikoetatik (iraupena, intentsitatea eta intonazioa) abiatuta, hizkuntza-eredu geoprosodikoak bilatu euskararen eredu prosodikoak (azentua, intonazioa) eskaini ahal izateko profesionalei.

3. Plataformaren ezaugarriak

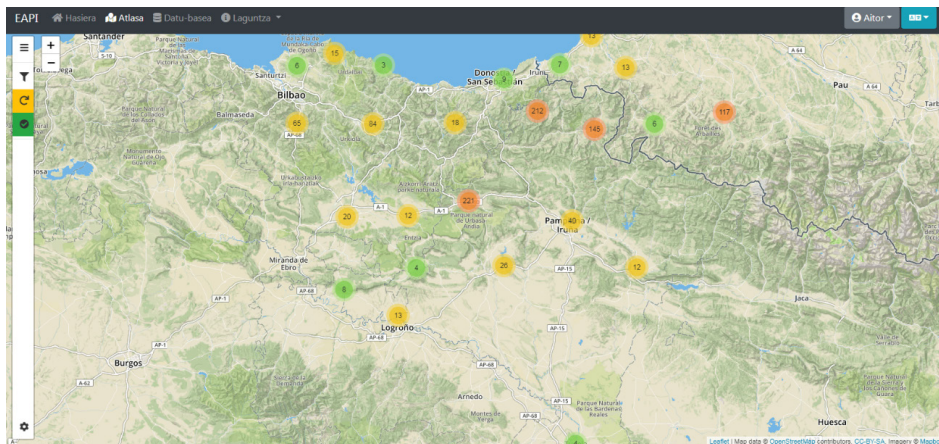
Atal honetan atlasaren alderdi teknikoak deskribatzeaz gain, internautek zelan erabili ahal duten webguneak eskaintzen dituen aukera guztiak ere aurkeztuko da.

Diziplina askotako adituek –linguistek, hezkuntzako irakasleek eta informatika ingeniariak– sortu dute atlasaren webgunea. Webgunearen egitura erabaki dute, zer atal izan behar duen, eduki motak, zer objektu izango duen eta hauen formatua. Webguneak funtsean atal nagusi bi ditu, datu-basea eta mapa. Datu-basearen testuak lau multzotan banatzen dira: azentua, intonazioa, bat-bateko testuak eta testu irakurriak. Testu bakoitzak bere audioa du, *.wav* formatuan; transkribatutako testua, *.text* edota *.textgrid* formatuetan; eta, azkenik, sonogramaren irudia, *.png* formatuan. Datu-basea maparatzen da bere objektu guztiekin; eta, horrela, atlasa sortzen da. Diseinatutako webgune dinamiko honetan erabiltzaileak aldi berean entzun ahal ditu audioak eta espektrogramak eta transkripzioak ikusi. Gainera, nahi izanez gero, guztiak jaitsi ahal ditu, bai audioak bai espektrogramak, bai testuen transkripzioak. Webguneari EUDIA ikerketa-taldeak UPV/EHU duen zerbitzaile batean eman zaio ostatu. Ezaugarri teknikoetan sakondu barik, esan dezakegu kode irekia duen ondoko software hau erabili dela eraikitzeko:

- PHPMailer – Licensed under LGPL.
- Bootstrap – Licensed under MIT license.
- jQuery – Licensed under MIT license.
- Popper.js – Licensed under MIT license.
- Leaflet – Licensed under BSD 2-Clause license.
- Leaflet-sidebar-v2 – Licensed under MIT license.
- Leaflet.markercluster – Licensed under MIT license.
- FontAwesome – Licensed under Font Awesome Free license.
- html5shiv – Licensed under MIT and GPL2 licenses.

Diseinu honen ondorioz webguneak *interface* bi ditu, bat edozein erabiltzailearentzat eta bestea editatzeko baimena duten erabiltzailearentzat (baimena izanez gero, hasierako orrian «saioa hasi» estekaren bidez sartu beharko da).

Erabiltzaile arrunt batek hasierako orrian hizkuntza hautatzeko aukera izateaz gain, beste hiru orrialdetan nabigatzeko aukera ere izango du: atlasean, datu-basean eta laguntza orrialdean. Atlasean sartuz gero, *OpenStreetMap* softwarearen gainean egindako mapan kolore ezberdinetako hainbat zirkulu ikusten dira zenbaki batez markatuak (1. irudia). Zenbaki horrek adierazten du gune horretan zenbat testu entzun eta ikusi ahal diren.



1. irudia. Atlasa.

Mapa gehienetan legez, zoom egin daiteke; adibidez, probintzia baten dauden audioak ikusi nahi badituğu zooma txikia erabiliko duğu, baina herri zehatz baten audioak aztertu nahi badituğu zooma handituko duğu audio bakoitzari dagokion txintxeta agertu arte. Txintxeta horretatik datuak ikusten diren orrialdera joan daiteke eta bertan testu bakoitzaren audioa, espektograma, transkripzioa eta herriari eta hiztunari dagozkion datuak ere eskuragarri daude (2. irudia).

Informatzailearen datuak:

Probintzia:	Bizkaja
Eskualdea:	Busturialdea
Herria:	Gernika
Generoa:	Andrazkoa
Adina:	18-25
Amaz hizkuntza:	A - Euskara etxean ikasitakoa
Ikasketak euskaraz:	Bai
Grabazioaren urtea:	2014

Galderak:

Galdera mota:	Testua
Esataria:	Arrunta
Testu mota:	bat-batekoa
Testu generoa:	Narrazioa

Erantzunak:

b	a	tx	a	n	o	go	rr	i	tx	u	b	a	i	c	a	(v)	H
ba	txa	no	go	rr	txu	ba	tc	(v)	H								
ba txanogerritua balen (v) H																	

Donbora (s) 0 1.181

Erantzun fonetiko eta fonologiko soinuaren imalarekin.

2. irudia. Datuen orria.

Arestian esan den bezala, erabiltzaile batek bilaketa datu-basetik zuzenean egiteko aukera ere badu. Edozein erabiltzailek datu-basea bere osotasunean du audio bakoitzaren datu guztiekin. Hiru datu nagusi daude datu-basean: informatazaila, herria eta audio mota. Datu-baseak duen bilaketak egiteko sisteman iragazkiak jar daitezke nahi den datu-mota guztietan.

Azkenik, laguntzaren orrialdea dugu. Orrialde honek hiru atal ditu. Lehenengo atalean, cookie eta JavaScript buruzko eta proiektuan erabili den kode irekiko softwareari buruzko informazioa dago. Bigarren atalean, egileei, hau da, EUDIA ikerketa-taldearen gaineko erreseina txiki bat dago; eta, amaitzeko, azken atalean, kontaktatzeko bidea.

Editatzeko baimena duen erabiltzaileak nabigatzeko sistema bera izango du, baina ohiko aukerez gain menu zabalgarri bat ere izango du datuak (informatzaila, herria eta audio mota) kudeatzeko modu erraz batean. Horretarako, informatikari buruzko ezagutza minimoak dituen erabiltzaile batek kudeatu dezakeen *interfazea* sortu da. Datu berriak sartzeko informatzaila baino ez du aukeratu behar (edo sortu berria balitz), galdera mota aukeratu eta audioa (.wav formatuan) eta Praatek egindako irudia (.png formatuan) jaregin. EAPIk herri berri bat sartzeko aukera bi ematen dizkigu: latitudea eta longitudea sartuz edo geolokalizazioa ere lortu ahal da maparen gainean herria dagoen lekuan klikatuz. Ikusten denez, oso moldagarria da eta munduko edozein puntu aukeratu ahal da datuak sartzeko; eta, beraz, atlas plataforma honek edozein hizkuntzari buruzko datuak har litzake.

4. Informazio prosodikoa eta sozio-geografikoa

Audioek eskaintzen duten informazio prosodikoa hiru multzotan banatzen da: azentuari buruzkoak, intonazioari buruzkoak eta testuak. Azentuari buruzko audioetan zenbat silaba dituzten eta zer sintagma mota den agertzen da. Intonazioaren arloan hurrengo esaldi motak bereizten dira: adierazpen esaldiak, Bai/Ez galderak, N/Z galderak, hautazkoak, zalantzazkoak eta oihartzun galderak. Azkenik, testuetan, bat-bateko testuak eta testu irakurriak bereizten dira. Informatzailari buruz honako datu hauek eskaintzen dira: generoa, adina, grabazioaren urtea, ama-hizkuntza eta ikasketak euskaraz egin dituenetz. Herriaren kasuan, zein probintziatakoa den aipatzen da eta eskualdea edo kantoia.

5. Datuen ezaugarriak

Puntu honetan atal bi jorratuko dira. Alde batetik, atlas osatzen duten corpusen berri emango da; eta, bestetik, datuen transkripzioa, etiketazioa, segmentazioa eta anotazioa nola egin den azalduko da.

5.1. Corpora

EUDIA ikerketa-taldeak aurreko ikerketa batzuen bidez corpus nagusi bi ditu *Euskara Dialektalaren Ahozko Korpusa* (EDAK) eta *Euskararen Atlas Sozio-linguistikoko-dialektologikoa* (EAS) (Aurrekoetxea & Ormaetxea, 2006; Aurrekoetxea, 2010). Lan hauetan heldu eta gazteen datuak batu dira bariazio diatopiko, diastratiko eta diafasikoen azterketa segmentalak eta suprasegmentalak egiteko. Baina atlasa osatzeko asmoz, corpus bi hauetatik gain, datu berriak biltzeko beharra ikusi da hizkuntzaren mailak kontuan hartuta (fonetiko-fonologikoa, morfologikoa, sintaktikoa eta semantikoa). Bariazio linguistikoa batzeko ohiko teknikak erabiliz (Aurrekoetxea & Iglesias, 2010) egin da informatzaileen datuen bilketa. Horrela, belaunaldi ezberdinetako informatzaile natiboen zein euskara L2 dutenen datuekin euskararen hizkuntza bariazioa aztertzeko plan bat garatu ahalko da.

Behin webgunearen egitura zehaztuta, datuen aukeraketa egin da. Hasteko, arestian aipatutako proiektuetatik, EAS eta EDAK, hartu dira datuak. Proiektu hauetatik datuak hautatzeko eta ustiatzeko, atlasa eratzeko irizpideak kontuan hartuta, datuen azterketa linguistikoa eta estatistikoa egin behar izan da. Ikerketa-taldeak duen datu kopururik handiena helduen eta gazteen azentuari eta intonazioari dagokie, baina bestelako datuak ere badaude informatzaile gazteengandik batutakoak: irakurritako testuak, bat-bateko testuak eta oinarrizko emozioen eta jarreraren testuak ere bai. Datu hauek guztiak aztertu dira egitura prosodiko adierazgarrienak dituztenak hautatzeko eta forma prototipikoetatik aldentzen direnak baztertzeko.

Ahozko euskararen ezaugarri adierazgarrienak osatzeko asmoz, EUDIA ikerketa-taldeak zituen corpusez gain, beste lagin esanguratsu batzuk batu behar izan dira. Datuen grabaketari ekin aurretik, datuak zelan batu behar diren prestatu behar da eta Roseanok (2017) aipatzen duen moduan, metodologiari dagokionez, faserik garrantzitsuena da. Atlas osatzeko batu den lagin berriaren helburua izan da barietate sozial zein geografiko bakoitzaren forma adierazgarrien hautaketa egokia egitea.

Euskal Herriko lurralde guztietako gazte elebidunen datu berriak jaso dira teknika bi hauek erabiliz: galdetegiak eta *map-task* metodologia. Galdetegiak, batez ere, bariazio linguistikoari buruzko erantzun homogeenak jasotzeko helburuarekin erabili dira; izan ere, ikerketaren alderdi geolinguistikoa aztertu nahi denean, ezinbestekoa da metodologia hau hizkuntzaren barietate guztiak zehaztasunez aztertuak izan direla ziurtatzeko. Ikerketa honen eremua euskararena izan denez, geografikoki nahikoa txikia beste hizkuntza batzuen aldean, asko erraztu du helburu hau lortzea. *Map-task* teknika datu prosodikoak jasotzeko erabili da; izan ere, teknika honek bat-bateko testuen datuen batuketa errazten du. *Map-task* bere modalitate bietan erabili izan da: *map-follower* eta *map-*

giverrean. Atlasean ezelango baldintzarik ez duen bat-bateko zein irakurritako testuen beste lagin bat ere jaso da, bost minutu inguruko iraupena duena. Lagin berri hauetan guztietan hizkuntza bere naturaltasunean batu nahi izan da, helburua komunitate baten ohiko hizkera erreala islatzea delako (Martínez Celdrán & Fernández Planas, 2005). Grabazioak gela intsonorizatueta egin dira Marantz PMD620 eta ZOOMH4 grabagailuekin, 44100 Hz.ko frekuentzian.

Datu berri hauek aurreko corpusak osatzeko balio dute atlasaren datu-basea ahalik eta zehatzena izateko eta baita barietate minoritarioak babesten laguntzeko eta jakiteko zeintzuk diren barietate zabalduenak hauek izango baitira modelo estandarra inplementatuko dutenak. Esan bezala, datuen jasotze-fasea kontu handiz egin behar da, ezpabere, ikerketaren emaitzak ezin izango dira fidagarriak. Atlaseko datu-basearen adierazgarritasun maila handia da. Alde batetik informatzaileena, kontuan hartuta alderdi geografikoa eta soziolinguistikoa; eta, bestetik, batutako datuek aztertu nahi izan diren eremu guztiak ordezkatzeko dituzte ereduak eskainiz.

EAPIren asmoa ez da itxiko den proiektu bat izatea, zabalik egon behar da corpus berriak inplementatzeko datu berriekin. Hori dela eta, aurretik aipatutako corpusez gain, atlasak laster jasoko ditu hurrengo subcorpusak:

- Oinarrizko emozioak (poza, tristura, haserra, nazka, beldurra eta harridura) jasoko dituen subcorpusa barietate guztietan.
- Oinarrizko jarrerak (abegikorra, ironikoa, errietalaria, mespretxuzkoa eta intentziozkoa) jasoko dituen subcorpusa barietate guztietan.
- Irakurritako eta bat-bateko testuen subcorpusetan esatari profesionalen grabatutako audioak erantsi.

5.2. Transkripzioa, etiketatzea, segmentazioa eta datuen anotazioa

Datuen batuketa eta hautaketa egin ostean transkripzioa, etiketatzea, segmentazioa eta anotazio prosodikoa egin da. Euskararen inguruan, zoritxarrez, gabezia handiak ditugu. Lanak sakabanatuta daude eta ez dago irizpide bateraturik datuen transkripzio eta anotazioan; horregatik, nekez alderatu daitezke datuak arlo fonetikoan zein fonologikoan ondorioak ateratzeko eta honek zailtasunak ere ematen ditu hizkuntzaren irakaskuntzan aplikatu ahal izateko. EAPIn transkripzio zein anotaziorako erabili diren iritziak bateratu dira batutako datu guztietan.

Arestian aipatu dugun legez, atlasak prosodiari buruzko corpus bat du lau subcorpusetan banatuta.

1) Azentuari buruzkoa.

Euskaren barietate linguistikoak jasota daude, baita estandarra erabiltzen duten informatzaileena ere.

2) Intonazioari buruzkoa.

Oinarrizko esaldien intonazioa jasotzen duen corpusa da: adierazpen esaldiak (sintagma batekin eta birekin), bai/ez galderak, N/Z galderak, hautazkoak eta oihartzun galderak.

3) Irakurritako testuei buruzkoa.

Komunikabideetako profesionalak ez diren informatzaileek irakurritako testuak jasotzen dira.

4) Bat-bateko testuei buruzkoa.

Komunikabideetako profesionalak ez diren informatzaileek ekoiztutako testuak jasotzen dira.

Azentuaren subcorpusean itemak silaba ezberdinetako hitzetan eta talde klitikoetan banatuta aurkezten dira. Datu hauekin azentuak dituen arau fonologikoak eman ahal dira euskaren bariedade guztietan eta euren arteko alderaketak egiteko. Arauaren salbuespenak diren hitz edo talde klitikoaren adibideak (hitzak lexikalki markatuak edo morfema pluralak) ere aurki daitezke.

Intonazioaren subcorpusean oinarrizko esaldien informazioa dago bariedade linguistiko guztien analisi fonetiko eta fonologikoa egiteko. Item guztiak fonologikoki transkribatuta eta prosodikoki anotatuta egoteaz gain, audio artxiboak ematen duen informazio guztia eskura garri dago espektrogramen bidez; audioa bera ere entzun eta jaitsi daiteke.

Irakurritako eta bat-bateko testuen subcorpusetan ere bariedade guztietako informazioa dago. Item guztiak fonetikoki transkribatuta eta prosodikoki anotatuta daude eta hauetan ere espektrogramak eta audioak eskura garri.

Datuen transkripzioa kudeaketaren une kritikoa da (Abasolo et al., 2019). Grabazioen bidez batutako datuak kodifikatu behar dira, baina aurretik erabaki behar da zein kodifikazio mota erabili behar den. Transkripzio guztietan informazioa galtzen da ikertzailearen interesen arabera egiten direlako (Linell, 2005); eta, horregatik, oso argi izan behar da lanaren helburua zein den, horren arabera egokitu behar delako egingo den transkripzioa (Briz et al., 2002). Hala ere, lehen aipatu den moduan, nahiz eta transkripzio zehatzena egin sekula ez da lortuko audio batek duen informazio guztia, horregatik atlas prosodikoetan ezinbestekoa da audioak entzun ahal izatea. Sailkapen zabal bat egiten badugu transkripzioak era bitakoak izan daitezke: fonologikoak edo fonetikoak. Euskal dialektologian ohikoa dugu transkripzio fonologikoa testu etnolinguistikoetan (Gaminde, 1994; Etxebarria, 2016). Transkripzio fonetikoak mota bikoak izan daitezke, zabalak edo zehatzak. Transkripzio zabalek hizkuntza batean ezaugarri fonetiko nabarmenak eskaintzen dituzte, zehatzenean aldiz, soinuaren informazio gehiago eskaintzen dute. Transkripzio zehatz baten adibidea Etxebarria (2017) dardarkariei buruz egindako lanean dugu, non bederatzi dardarkari bereizten dituen. Beraz transkripzio fonetikoak zenbat eta zabalagoa izan gehiago hurbilduko da transkripzio fo-

nologikora. Lan honetan transkripzio mota biak erabili dira kontuan hartuta zein datu jaso diren eta informazio helburuak zeintzuk diren. Transkripzio fonetikoak egiteko Nazioarteko Alfabeto Fonetikoa (IPA *International Phonetic Alphabet*) erabili da. Azentuaren subcorpusean, hitzak eta talde klitikoak transkribatzeko transkripzio fonologikoa erabili da azentua markatuz. Intonazioaren subcorpusean, esaldi laburretarako ere transkripzio fonologikoa erabili da maila bitan, esaldien osagaiak kontuan hartuz eta silabaka. Testuak lau mailatan daude transkribatuta fonetikoki: lehenengo transkripzioan testuaren talde prosodikoak kontuan hartuta; bigarrenetan talde prosodiko bakoitzaren osagaiak kontuan hartzen dira; hirugarrenean silabak, eta, azkenik, soinu bakoitza. Transkripzio eta anotazio prosodiko guztiak *Praaten* jaso dira *TextGrid* fitxategien bidez. Silabafikazioa *Praat* (Boersma & Weenink, 2015) programan lortu da modu semiautomatiko batean beren-beregi prestatutako *script* batekin. Audioa segmentatu ostean egin da transkripzioa maila ezberdinetan material mota eta anotazio prosodikoa kontuan hartuta. Muga tonuak eta eten motak (isiluneak, bokalen luzapenak edo soinuen txertaketak) anotatu dira eta baita tonuaren gorakadak eta beherakadak. Esan bezala, anotazio sistema materialen arabera egokitu da (hitzak, esaldiak, testuak...).

6. Erabilera eta aplikazioak

EAPI euskararen prosodiari buruzko informazio esanguratsuen batzen duen lehenengo atlasa da, eta, beraz, erreferentea izan behar dena gai honen gainean egingo diren ikerketetan publiko jartzen duelako datu-base prosodiko handia. Atlas honek euskara prosodiaren ikerketa arloan oso garatuta dauden beste hizkuntza batzuen, AMPERekoak adibidez, parean jartzen du nazioarteko mailan.

Hezkuntzan erreferentea izan behar da ezinbesteko lanabesa delako euskara estandarren zein dialektalaren azentua eta intonazioaren jabeakuntzarako. Ikerketa askotan euskararen ohiko eredu prosodikoak galtzen edo aldatzen ari dira eta argi dago ondoko hizkuntzen ereduak gailentzen ari direla bai euskara estandarrean eta baita euskara ama-hizkuntza duten hiztunengan, maila dialektalean hain zuzen ere (Etxebarria & Eguskiza, 2018). Euskara hizkuntza gutxitua da, komunikabideetan eragin eskasa duena eta estandarrek ere ez du eredu prosodiko zabalik. Euskararen prosodia egoera kritikoan dago irakaskuntzan ez dagoelako lanabesik prosodia irakasteko. EAPIk hutsune hau bete behar du irakasleek euskara irakasteko behar dituzten azentu eta intonazio ereduak emanez irakaskuntza arauturako, Lehen Hezkuntzatik unibertsitateraino, zein helduen alfabetatze eta berreuskalduntzerako bigarren hizkuntza ikasteko. Atlasean dauden datuen azterketak hizkera ezberdinen pertzepzio eta imitazio ariketak sortzeko aukera emango du. On-line baliabide bat denez, eskola-

ren osagarria ere izango da etxean ikasleek, familiek, hezitzaileek... erabili ahal izango dute lanetan laguntzeko.

Azentu eta intonazio subcorpusten datuen azterketa linguistiko eta estatistikorekin prosodiaren bi elementu hauen eremuak kartografiatu ahal izango dira eta azaldu zer barietate duen euskarak. Gainera, analisi sozioprosodikoa egiteko aukera ere ematen du informatzaileen generoa eta ama-hizkuntza kontuan hartuz. Datuen azterketa estatistikoa egiteko eta kartografiatzeko EUDIA ikerketataldeak ere garatu du on-line eta doakoa den programa bat: *Diatech* (Aurrekoetxea et al., 2016). Euskal hizkuntzalaritzan oso garrantzitsua da ezaugarri linguistiko bakoitzaren eremu geografikoa zein den jakitea. Horretarako, atlasean dagoen datu-baseaz, *Praat* programan egindako azterketa fonologikoa duena, baliatuko gara *Diatechen* automatikoki kartografiatzeko. Alde batetik, datu-basean dauden ezaugarri bakoitzeko mapa bat sor daiteke; eta, bestetik, mapa sintetikoak ere egin daitezke sailkatze helburuen araberako analisi estatistiko ezberdinak erabiliz. Kartografiatze honekin euskararen eremu prosodiko-geografiko garrantzitsuenak ikusiko dira eta hizkuntza ezaugarri bakoitzak bere daturik esanguratsuenak aurkeztuko ditu bere hedapen geografikoa. Datu guztien kartografiatze lanak egin ostean, prosodiari buruzko sailkapen dialektala egin ahal izango da. Sailkapen hau, datu kopurua kontuan hartuta, euskaraz egin den zabalena izango da; eta, gainera, informatzaile gazteekin lortutakoa, orain arte ez dena egin modu hain sakonean hizkuntza honetan.

Datu-base honekin EAPI garrantzitsua izan behar da ere hizkuntza teknologietan; ahotsa ezagutu eta sintesia egiteko motoreak garatzeko, hizkuntzaren patologia tratatzeko, desgaitasunak, ahotsaren lesioak errehabilitatzeko; eta, azkenik, fonetika forentserako; izan ere, datu-base handi honen bidez jakin ahal izango da geografikoki edozein ahots nongoa den.

Laburbilduz, EAPIk euskararen ikerketa arloan zegoen hutsunea bete du euskaraz egin den lehen atlas prosodikoa delako eta komunitate zientifikoaren esku jartzen duelako prosodiaren arloko ikerketak egiteko ezinbesteko tresna bat.

Erreferentziak

- Allières, J. (1978). Petit Atlas linguistique basque français 'Bourciez I' (20 cartes commentées). *Fontes Linguae Vasconum*, 27, 353-386.
- Allières, J. (1986). Petit atlas linguistique basque français 'Bourciez II' (30 cartes commentées). *Fontes Linguae Vasconum*, 47, 5-45.
- Aranzadi Zientzia Elkarte. (1983). *Euskal Herriko Atlas Etnolinguistikoa 1*. Donostia: Aranzadi Zientzia Elkarte.
- Aranzadi Zientzia Elkarte. (1990). *Euskal Herriko Atlas Etnolinguistikoa 2*. Donostia: Aranzadi Zientzia Elkarte.

- Aurrekoetxea, G., Gaminde, I., Gandarias, L., & Iglesias, A. (2014). Prosodic variation in the Basque language: intonational areas. En A. Díaz-Negrillo & F. J. Díaz Pérez (Eds.), *Specialisation and Variation in Language Corpora*. Berna: Peter Lang AG, International Academic Publishers.
- Aurrekoetxea, G., Gaminde, I., Gandarias, L., & Harignordoqui, E. (2011). EDAK: euskal azentuaren etiketatze bidean lehen emaitzak. En G. Aurrekoetxea & I. Gaminde (Eds.), *Prosodiaz eta Heskuntzaz I. Jardunaldiak/I Jornadas sobre Prosodia y Educación*. Bilbao: UPV/EHU.
- Aurrekoetxea, G., Gaminde, I., Iglesias, A. & Gandarias, L. (2013). Prosodic Variation in the Basque Language: Stress Areas. En E. Carrilho, C. Magro & X. Álvarez (Eds.), *Current Approaches to Limits and Areas in Dialectology*. (247-266). Newcastle: Cambridge Scholars Publishing.
- Aurrekoetxea, G., Videgain, X., & Iglesias, A. (2004). 'Bourciez' bildumako Euskal Atlas: 1. Lexikoa. *ASJU XXXVIII-2*.
- Aurrekoetxea, G., Videgain, X., & Iglesias, A. (2005). 'Bourciez' bildumako Euskal Atlas: 2. Gramatika. *ASJU XXXIX-1*.
- Bonaparte, L. L. (1863). *Carte des sept Provinces Basques*. London: Stanford's Geographical Establishment.
- Bourciez, E. (1895). *Recueil des idiomes de la région gasconne*. (Eskuizkribua). Bordeaux: Bibliothèque Universitaire des Lettres.
- Elordieta, G. & Hualde, J.I. (2014). Intonation in Basque. *The phonology of intonation and phrasing*, 405-463.
- Etchebest, X. (2017). *Zuberotar dardarkarien bariazio sozio-fonetikoa*. (Doktoretza tesia. UPPA & UPV/EHU).
- Etxaide, A.M. (1984). *Erizkizundi hirukoitza / Triple cuestionario / Triple questionnaire (Euskera 1925)*. IKER 3. Bilbo: Euskaltzaindia.
- Etxebarria, A. & Eguskiza, N. (Eds.) (2018). *Bariazioa esaldien intonazioan*. Bilbao: UPV/EHU.
- Etxebarria, A., Eguskiza, N., Gaminde, I., & Romero, A. (2018). Ahots seduzitzaileen ezaugarriak generoaren ikuspegitik. En L. Unamuno, A. Romero, A. Etxebarria & A. Iglesias (Eds.), *Linguistic variation in the Basque language and Education/Euskararen bariazioa eta bariazioaren irakaskuntza* (43-59). Leioa: UPV/EHU.
- Etxebarria, A., Gaminde, I., Romero, A. & Iglesias, A. (2016). Desarrollo de la competencia prosódica en la lectura en voz alta: importancia de las pausas. *OCNOS Revista de estudios sobre lectura*, 15, 110-118.
- Etxebarria, A., Sanz, A., Garay, U. & Romero, A. (2017). Unibertsitateko ikasle elebidunen prosodia gaitasuna Txanogorritxu ipuinaren kontaketa: muga-tonuak. En A. Iglesias & A. Ensunza (Eds.), *Gotzon Aurrekoetxea lagunarterik hara* (111-122) Bilbao: UPV/EHU.
- Etxebarria, J. M. (2016). *Gorbeia inguruko etnoipuin eta esaundak II*. Bilbo: Deustuko Unibertsitatea & Euskaltzaindia.
- Euskaltzaindia (2008-2019). *Euskararen Herri Hizkeren Atlas* (10 liburuki). Bilbo: Euskaltzaindia.

- Gaminde, I. (1985e). *Aditza Ipar Goi Nafarreraz III*. Iruña: UEU.
- Gaminde, I. (1994). *Bilboko Larraskituko euskaraz: Lexikoa eta Testuak*. Bilbo: Bilboko Udala.
- Gaminde, I. (1985a). *Aditza Biskaieraz II*. Iruña: UEU.
- Gaminde, I. (1985b). *Aditza Biskaieraz III*. Iruña: UEU.
- Gaminde, I. (1985c). *Aditza Ipar Goi Nafarreraz I*. Iruña: UEU.
- Gaminde, I. (1985d). *Aditza Ipar Goi Nafarreraz II*. Iruña: UEU.
- Gaminde, I., Aurrekoetxea, G., Etxebarria, A., Garay, U., & Romero, A. (2014). *Ahoskera lantzeko argibideak eta jarduerak. Laguntzarako materiala: teoria eta praktika*. Leioa: UPV/EHU.
- Gaminde, I., Etxebarria, A., Romero, A., & Eguskiza, N. (2017a). Características de la competencia prosódica de jóvenes bilingües vascos en la lectura en voz alta: las cumbres tonales. *Revista de Lingüística Teórica y Aplicada*, 55, 35-72.
- Gaminde, I., Iglesias, A., Eguskiza, N., & Etxebarria, A. (2017b). Tasun akustikoen eragina jarrenen bereizkuntzan. *Ikastorratza*, 19, 126-149.
- Martínez Celdrán, E. , & Fernández Planas, A. M. (Coords.). (2003-2018). *Atlas Multimedia de la Prosodia de l'Espai Romànic*. http://stel.ub.edu/labfon/amper/cast/index_ampercat.html [24-02-2020]
- Martínez Celdrán, E., & Fernández Planas, A. M. (2005). Estudio metodológico acerca de la obtención del corpus fijo en el proyecto AMPER. *Estudios de fonética experimental*, 14, 29-66.
- Prieto, P., & Cabré, T. (Coords.). (2007-2012). *Atles interactiu de l'entonació del català*. <http://prosodia.upf.edu/atlesentonacio> [10-02-2020]

