

# Euskal hizkeren azterketarako bi baliabide informatiko

Larraitz Uria, Ricardo Etxepare\*  
IXA taldea, IKER UMR-5478\*

Euskal Herrian leku batetik bestera eta garai batetik bestera modu, forma edota egitura desberdinak erabili izan ditugu euskaraz gauza berdinak adierazteko. Adibidez, gaur egun, Euskal Herriko leku batzuetan *Jon artzaina da* esaten da; beste batzuetan, *Jean artzain da* (garai batean, artikulurik gabeko forma hori hedatuagoa zen gaur egun baino). XVI. mendean aginterazko forma sintetikoak zuzenean ezeztatuak izan zitezkeen (*ze eikek mauruti oanean, eder eztanik kalean* (Lakarra, 1996)), egungo euskaran ezinezkoa dena. Euskalkiak edo hizkerak modu sistematiko batean biltzeak aukera emango liguke euskal hizkeren panoramika bat izateko, hau da, euskal hizkeren bilakaera ezagutzeko, euskara leku eta garai jakin batean nolakoa izan den jakiteko. Hala, gure ikerketa-lanaren xedeetako bat euskal hizkerak aztertzeko baliabide informatikoak garatzea da, ikerketa-alor honetan tresna informatikoek aukera handiak eskaini ahal dizkigutela uste baitugu. Artikulu honetan euskara ez-estandarra aztergai duten bi proiekturen berri eman nahi dugu: 2009an abiarazi genuen BASYQUE aplikazioa eta diseinu-fasean dagoen euskarazko corpus historikoaren arakatzaille sintaktikoa.

GAKO-HITZAK: Dialektologia · Aldakortasun sintaktikoa · Hizkuntzalaritza Konputazionala · Corpusak.

## Two NLP based resources for the analysis of Basque dialects

In the Basque Country, in different locations and different periods, several structures and forms have been used to say the same thing. Collecting information about Basque dialects across different places and periods in a systematic way, offers us the possibility to build a panoramic view of the Basque dialects and their evolution, that is, to know how our language has been spoken in a given place and in a given period. And this is, precisely, one of our research goals: we intend to develop Natural Language Processing (NLP) based resources focused on the analysis and processing of Basque dialects across time and space, since we believe that current technological applications provide interesting possibilities in this research field. In this article, we present the BASYQUE application and the syntactic parser we intend to develop for the analysis of historical texts.

KEY WORDS: Dialectology · Syntactic alternation · Computational Linguistics · Corporal.

*Jasotze data:* 2016-06-28    *Onartze data:* 2016-11-30

## 1. Sarrera

Euskararen banaketa dialektala euskarazko lehenengo testuetatik antzematen da. Euskal Herri gehieneko lekukotasun idatziak dauzkagu XVI. menderako, eta hasiera-hasierako idazki horiek gure hizkuntzaren barne-aldakortasunaren testigantzak dira. Euskalkien inguruko ikerketak ere orain dela urte asko hasitako lanak dira. *Euskalkiak* webgunean<sup>1</sup>, euskal hizkerak aztertzen ibili ziren hiru aitzindari aipatzen dira: Arnaut Oihenart (1592-1667), Aita Manuel Larramendi (1690-1766) eta Luis Luziano Bonaparte (1813-1891).

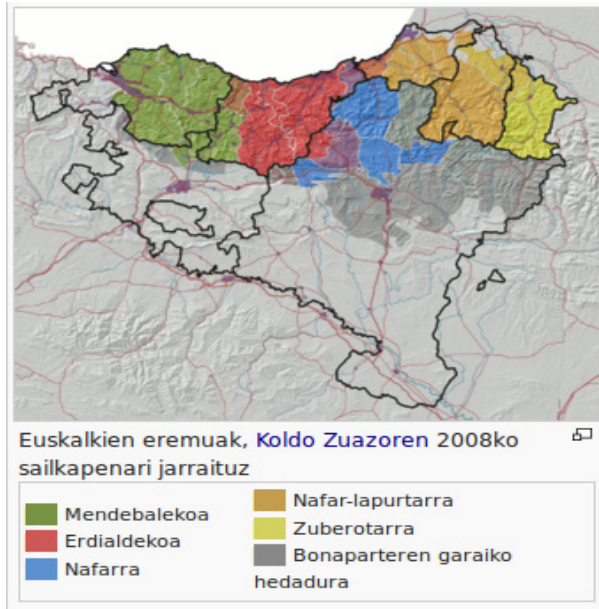
Arnaut Oihenart zuberotarra (1592-1667) izan zen euskalkien jatorriaren gaineko iritziak plazaratzen lehena. Hain zuzen ere, 1638an Parisen argitaratutako *Notitia Utriusque Vasconiae* liburuan aurkeztu zituen Oihenartek euskalkiak. Sailkapen berritzailea aurkeztu zuen: euskalkien mugak herrialdeen arabera bereizi beharrean (bere garaikide zen Axularrek egin zuen moduan, esate baterako), antzinako leinuekin lotu zituen eta lau euskalki bereizi zituen: akitaniarra (Ipar Euskal Herrian egiten zena), baskoiena (Nafarroan egiten zena), barduliarrena (Gipuzkoan eta Araban egiten zena) eta autrigoiena (Bizkaian egiten zena). Oihenarten sailkapenaren arabera, beraz, euskalkiak leinuak zeuden garaikoak dira, erromatarrak hona etorri aurrekoak.

Euskalkiak sailkatzeko bigarren ahalegina Aita Manuel Larramendi jesuitak (1690-1766) egin zuen. *El Imposible Vencido. Arte de la Lengua Vascongada* (1729) gramatika hiru euskalki bereizi zituen (1729: 12): bizkaitarra, giputza eta nafar-lapurterra. Arabako euskara izaera nahasikotzat zeukan eta bizkaitarrarekin batera sailkatu zuen, hainbat ezaugarri hartatik aldentzen zutela aitortu arren. Larramendirena, dena dela, sailkapen aski erromesa da.

Luis Luziano Bonaparte Printzea (1813-1891) izan zen benetan euskararen eremua bere hedadura eta zehaztasun osoan egiazki ezagutzen eta aztertzen saiatu zen lehena, Euskal Herriko lehen dialektologotzat jo dezakegu. Bonaparte ez zen euskalkien jatorriaz arduratu. Bere helburua bere garaian egiten ziren euskalkiak definitu eta mapa batean erakustea zen. Azterketa-lana eta haren emaitza den lehenengo mapa dialektala elkarrekin inprimatu ziren 1863an. Euskara 8 euskalkitan zatitu zuen Bonapartek, hiru talde nagusitan banatuta: mendebalekoa (bizkaiera litzatekeena), erdigunekoa (gipuzkera, lapurtera, iparraldeko goi-nafarrera eta hegoaldeko goi-nafarrera) eta ekialdekoa (zuberera, ekialdeko behe-nafarrera eta mendebaldeko behe-nafarrera). Zortzi euskalki horien barruan, gainera, 25 azpieuskalki eta 36 hizkera bereizi zituen (Bonaparte, 1863 eta 1869).

---

1. <[www.euskalkiak.eus](http://www.euskalkiak.eus)>.



1. irudia. Koldo Zuazoren sailkapena (2008).

Bonaparteren garaian hasi eta orain artean, ikertzaile ugari ibili da euskal dialektologiari buruzko azterketak egiten. Egun, ikertzaileek gai honen inguruan ikertzen jarraitzen dute, euskal hizkuntzalaritzaren alorrean interes handia duen gaia delarik euskalkiena. Guk, proiektu honetan, Koldo Zuazok 1998an eta 2008an aurkezturiko euskalkien sailkapena hartu dugu oinarritzat (1. irudia), euskal hizkeren sailkapen eguneratua baita, azterketa luze eta sakon baten emaitza<sup>2</sup>.

Euskalkiak, beraz, aspaldidanik datozen hizkerak dira. Euskara batua izeneko hizkuntza-eredua sortu baino lehenago, euskalkiak ziren euskararen agerbide bakarrak<sup>3</sup>.

Euskararen estandarizazioa planifikatzeko eta gauzatzeko ahaleginak lehenago-koak diren arren (1918an sortu zen Euskaltzaindia), 1968an, Euskaltzaindiak Arantzazun egindako VII. Biltzarrean, sortu zen ofizialki euskara batua eta ekin zitzaion, zinez eta formalki, euskara idatziaren batasuna bideratzeko zereginari. Euskara batua diseinatzeko, erdialdeko euskalkietan (gipuzkera, lapurtera eta goi-nafarrera) oinarritutako eredu konposizionala hartu zen abiapuntutzat, aldi berean euskalki guztietako ekarpenak beretuko zituen eredu finkatuz. Gaur egungo euskara batua izeneko hizkuntza-ereduaren oinarriak finkatu eta idazteko arau batzuk gomendatu zituen Euskararen Akademiak. Garai hartan zegoen

2. Hala ere, kontuan izan behar da egoera dialektala, gaur egun ere, ez dela hain garbia; batetik, bizirik dirauten euskalkietan euskara batuaren eragina handia da leku batzuetan; bestetik, gizartearen nahiz norbanakoen erabilera nahasiek konplikatu egiten dute sailkapen argi eta zehatz bat egitea; eta gainera, jende askoren euskalduntzea gertatu da. Gertakizun horien ondorioz, zaila da euskararen egungo egoera dialektala argi eta garbi zehaztea, bai espazio geolektalei baita maila idiolektalei dagokienez ere.

3. Euskalkien historiaz gehiago jakin nahi duenak informazio zehatza eta ongi laburbildua aurkituko du webgune honetan: <[http://euskalkiak.eus/euskalkien\\_jatorria.html](http://euskalkiak.eus/euskalkien_jatorria.html)>.

euskalkien aniztasunaren aurrean, Euskaltzaindiak erabaki zuen euskara batua beharrezkoa zela gure hizkuntzaren berreskurapenerako eta biziraupenerako, leku guztietako euskaldunok elkarri ondo ulertzeko hizkuntza-eredu bat izateko. Beraz, hizkera desberdinetako hiztunen arteko ulergarritasuna bermatzea izan da kode estandarraren xede nagusia, euskal hiztunen kopurua handitzeko eta gure hizkuntza-komunitatea trinkotzeko. Eta euskara batuaren gauzatzeak ekarri zuen, gainera, euskal hizkuntzaren barruan dauden kode desberdinen banaketa funtzionala (Zuazo, 2000; Euskaltzaindia, 2004).

Euskara batuaren eta euskalkien erabileraren inguruan joera desberdinak egon izan dira, hainbat gogoeta egin dira, askotariko iritziak plazaratu dira eta bibliografia oparoa argitaratu da (King, 1983; Larringan, 1993; San Martin, 1998; Villasante, 1998; Zuazo, 2005 eta 2000, besteak beste). Euskaltzaindiaren sorreratik jada, eztabaida eta desadostasun ugari egon izan da eredu estandarraren eta euskalkien erabileraren inguruan (Maia, 2001). Lekuaren eta testuinguruaren arabera forma eta mintzabide egokiak hautatu beharraz ari dira gehienbat euskalariak gaur egun, eta bertan euskalkiek bete-beteko eginkizuna dute. Euskalkien erabileraren aldeko iritziak indarra hartzeak izan du eraginik Hizkuntzaren Azterketa eta Prozesamenduaren (HAP) alorrean ere. Izan ere, orain dela urte gutxi arte euskara batuaren tratamendu automatikora mugatu bagara ere, hasiak gara hizkeren prozesamenduan urratsak egiten, batuarekin batera tokian tokiko komunikazio-moduei ere arreta berezia eskaintzea ezinbestekoa dela sinetsita.

Hizkuntza-bariazioaren alorrak piztu duen interesa ikusita eta azken urteotan teknologia berrien garapen handiaren eraginez, euskal hizkerak, hizkeren berezitasunak nahiz hizkeren arteko desberdintasunak aztertzeke hainbat ikerketan eta proiektu garatu dira gure hizkuntza-komunitatean. Eta lan horien emaitza dira ondoko webgune, corpus edota aplikazioak, besteak beste<sup>4</sup>:

- Koldo Zuazok eta bere lantaldeak sortutako *Euskalkiak* webgunean<sup>5</sup>, euskal hizkeren jatorriari eta ezaugarriei buruzko informazio osatua/zabala eskaintzen digute, ikus-entzunezko testigantzekin.
- *Euskal Herriko Ahotsak*<sup>6</sup> proiektuan Euskal Herriko ahozko ondarea biltzen eta herri-hizkerak katalogatzen ari dira, gure herrietako adineko jendea elkarriketatuz euren euskara eta bizipenak jasotzeko eta herritarren artean zabaltzeko helburuarekin.
- Euskal Herriko Unibertsitateko *Bas&Be (Aldakortasun sintaktikoa aztergai: euskara eta beste)*<sup>7</sup> ikerketa-taldearen xede nagusia aldakortasun sintaktikoa aztertea da, hizkuntza arteko eta hizkuntzaren barneko aldaerek erakusten dutena. Hizkuntza asko dituzte aztergai, baina batez ere euskararen eta euskal hizkeren arteko aldakortasun sintaktikoan dute interesa.

---

4. Testurik zaharrentzat eta bestelako iturri digitalentzat, ikusi ere Bilbao eta Gómez (2014).

5. <<http://euskalkiak.eus>>.

6. <<http://www.ahotsak.eus>>.

7. <<http://basdisyn.net/Bas&Be/>>.

- *Eudia*<sup>8</sup> Euskal Herriko Unibertsitateko beste ikerketa-talde bat da, eta talde honen xede nagusiak dira euskararen bariazioa ikertzea, bariazio hori sortzen duten faktoreak (sozialak zein bestelakoak) zein diren zehaztea eta bariazioa aztertzekeo lagungarriak izango diren tresnak sortzea.
- *Euskararen Herri Hizkeren Atlasak (EHHA)*<sup>9</sup> ahozko euskararen aldaerak mapetan jasotzea du helburu.
- Baionako IKER UMR-5478 taldean hasitako *Norantz proiektuaren*<sup>10</sup> helburua gaur egungo Iparraldeko euskaldunen gramatika-oinarriak, joera handienetan, zein bidetatik doazen ikustea da, informazio hori datu-base batean bildurik erakustea. Eta horrela, euskal hizkeren gramatikaren bilakaeraren behatoki baten zimenduak finkatzea.
- Deustuko Unibertsitatean sortu dute *Bonaparteren ondarea* ataria<sup>11</sup>. Bertan, Bonaparte Ondarea osatzen duten euskalki guztietako testuen edizio digitalizatua eskaintzen da batetik, eta, bestetik, testu horiek maila morfosintaktikoan analizatzeko aukera.
- *Lazarragaren eskuizkribu digitalizatua*<sup>12</sup> arabar idazleak berak gutxi gorabehera 1567tik 1602ra bitartean idatzitako lanak biltzen dituen paper sorta da. UPV/EHUko Monumenta Linguae Vasconum ikerketa-taldea ari da testu horiek editatzen eta aztertzen; zehazki, testuaren irakurketa sakonki lantzen eta oharrez hornitzen ari dira, besteak beste testuaren ulermena laguntzeko helburuarekin.
- *Klasikoen Gordailua*<sup>13</sup> eta *Euskal Testuen Gordailua*<sup>14</sup> aipatu beharreko beste bi atari dira, biek eskaintzen baitute testu-bilduma handia modu errazean eskuratzeko eta arakatzeko aukera. Euskal literaturako testu klasikoen bildumak dira, XX. mendea bitartean garai eta leku desberdinetako idazleek idatzitako lanekin osatutako bildumak (lan gehienak literaturaren alorrekoak dira, baina ez bakarrak). Horregatik, corpus horiek oso baliagarriak dira euskal hizkeren azterketarako, euskalkien azterketa sinkroniko nahiz diakronikoak egin ahal izateko abiapuntu eta oinarri sendo bat osatzen baitute. 2007an, corpus-arakatzaila sortu zen, Euskal Klasikoen Corpora (EKC)<sup>15</sup>, testu-bilduma horietako batzuetan bilaketak egiteko aukerak eskaintzen dituena. Testu horiek, baina, ez daude ez lematizatuta, ez etiketatuta, eta, ondorioz, mugatuak dira corpusaren gainean egin daitezkeen bilaketa motak. Aurrerago azalduko dugun moduan, gure xedea, hain zuzen ere, horrelako testu zaharren azterketa egitea da, maila sintaktikoan.

---

8. <<http://www.ehu.eus/eu/web/eudia/home>>.

9. <<http://www.euskaltzaindia.eus/ehha>>.

10. <<http://www.norantz.org/>>.

11. <<http://bonaparte.deusto.es>>

12. <<http://www.ehu.eus/monumenta/lazarraga> / [https://eu.wikipedia.org/wiki/Lazarra-garen\\_eskuizkribua](https://eu.wikipedia.org/wiki/Lazarra-garen_eskuizkribua)>.

13. <<http://klasikoak.armiarma.eus/>>.

14. <<http://www.vc.ehu.es/gordailua/HomePage.html>>.

15. <<http://www.ehu.eus/ehg/kc/>>.



adierazpen hori baieztatzen duten bi baliabide aurkezten ditugu artikulu honetan: BASYQUE aplikazioa<sup>18</sup> eta euskarazko corpus historikoaren arakatzaille sintaktikoa.

BASYQUE Iparraldeko hizkeren artean dagoen aldakortasun sintaktikoa aztertzeko aplikazioa da. Bertan jasotzen den informazioak azterketa dialektologiko sinkronikoak nahiz diakronikoak egiteko aukera eskaintzen digu, hau da, momentu jakin batean edo urte jakin batzuetan zehar hainbat herritako hiztun euskaldunek egitura morfosintaktiko jakin batzuk nola erabiltzen dituzten jaso eta leku batetik bestera dauden desberdintasunak alderatzeko eta aztertzeko aukera eskaintzen digu<sup>19</sup>. Eta horixe izan da, zehazki, euskal hizkeren azterketa eta prozesamenduaren alorrean gure abiapuntua. Orain, lan hori oinarri hartuta eta lanaren emaitzak ikusita, euskalkiak aztertzen jarraitzeko gogoia eta beharra sentitzen dugu, gure ikergaia leku eta garai guztietako hizkeren azterketara zabaltzeko nahia. Horregatik, euskal sintaxiaren historia ezagutzeko aukera emango diguten baliabideak garatzea da orain ikusmiran dugun proiektua: euskararen corpus historikoaren arakatzaille sintaktikoa, hain zuzen. Izan ere, urte edota leku desberdinetan hainbat autorek zenbait euskalkitan idatzitako testu-bilduma handi bat sintaxi-mailan arakatzeko baliabideak izateak ikerketarako ate ugari zabaltzen dizkigu, euskararen sintaxiak (euskara idatzian oinarriturikoak) mendeetan zehar izan duen bilakaera ezagutzeko aukera.

Gure ikerketa-lana aurrera eramateko, alor linguistikoa eta informatikoa uztartzen ditugu, biak beharrezkoak eta bateragarriak direlako ustean. Hain zuzen ere, BASYQUE aplikazioaren datu-basean gordetzen den informazioa funtsezkoa da hizkeren arteko bariazio sintaktikoa aztertzeko, baina era berean, informazio hori baliatzen dugu lantzen ari garen aldaeren prozesamendua bideratzeko ere. Eta horrelako atlas linguistikoetan HAPren alorreko baliabideak integratzea berrikuntza ez ezik, aurrerapauso garrantzitsua ere bada, horrek erraztu egiten baitu batetik, aldakortasun sintaktikoaren inguruko datuak eskuratzeko eta aztertzeko aukera, eta, bestetik, bildutako adibideen tratamendu automatikoa bideratzeko lana.

## 2. BASYQUE aplikazioa

BASYQUE hizkuntza bereko hizkeren artean dagoen aldakortasun sintaktikoa aztertzeko xedearekin garatu dugun aplikazioa da.

Gure ikerketa Iparraldeko hizkeretara mugatu da orain artean. Iparraldeko hizkeren arteko bariazio sintaktikoa oso nabarmena da fenomeno linguistiko jakin batzuetan; adibidez, *lanari lotu zaio* esaten dute batzuek eta beste batzuek, aldiz, *lanari lotu da* (datibo-komunztadura egin gabe) esan dezakete; leku batzuetan *Xan artzaina zen* diote eta beste batzuetan *Xan artzain zen* (determinatzailea isilpean gordeaz); eremu dialektal batzuetan *Auzapezak sosa ebatsi omen du* esango dute, eta beste batzuetan, berriz, *Auzapezak sosa ebatsi du omen* esango dute (intonazio-hausturarik gabe, *omen* partikularen posizioa aldatuz). Aldakortasunari dagozkion horrelako adibideak jasotzeko eta kontsultatzeko aukera eskaintzen digu, hain zuzen ere, BASYQUE aplikazioak. Era berean, eta sarreran aipatu dugun moduan,

---

18. <<http://ixa2.si.ehu.es/atlas2/index.php?lang=eu>>.

19. Azterketa soziolinguistikoak egiteko, beraz, aproposa da aplikazio hau.

gure datu-basean jaso ditugun adibide horiek izan dira euskal hizkeren tratamendu automatikoa bideratzen hasteko lanaren abiapuntua.

Gure aztergaia Iparraldeko hizkerak diren arren, aplikazioa baliagarria da gainerako hizkeren inguruko azterketak ere bideratu ahal izateko.

Atal honetan, BASYQUE aplikazioa garatzeko jarraitu dugun metodologia aurkeztuko dugu, aplikazioak bilaketak egiteko eskaintzen dizkigun aukerak azalduko ditugu eta lehen urrats honetan atera ahal izan ditugun ondorio batzuk laburbilduko ditugu.

## 2.1. Metodologia

Aplikazioa garatzeko metodologiari dagokionez, hiru urrats nagusi aipa daitezke: aplikazioaren diseinua prestatzea, datuak biltzea eta bildutako datuak etiketatzea.

Diseinuari dagokionez, esan beharra dugu BASYQUE aplikazioan hiru informazio-iturritatik bildutako adibideak gordetzeko eta kontsultatzeko aukera prestatu dugula, horretarako hiru atal bereiziz: *Galdeketak* atala, *Literatur corpusak* atala eta *Bestelako iturriak* atala. *Galdeketak* atalean, galdeketa zehatzen bidez eskuratutako adibideak gordetzen ditugu; *Literatur corpusak* atalean, testu literarioetan aurkitutakoak; eta, azkenik, *Bestelako iturriak* atalean, beste lan edo proiektu batzuetatik eskuratutako adibideak.

Beraz, hiru iturri desberdinetatik jasotzen ditugu adibideak. Horregatik, hain zuzen, hiru atalak bereizita daude, bateko zein besteko datuak nahastu gabe, eta datuak kontsultatzen hasi baino lehen, erabiltzaileak informazio-iturria zehaztu behar du aplikazioaren orri nagusian. Datuei dagokienez, fenomeno linguistiko berdinak izan ditugu aztergai hiru kasuetan eta bildutako informazioa kudeatzeko, kontsultatzeko eta prozesatzeko aukerak ere oso antzekoak dira hiru ataletan. Baina datuen jatorria eta ezaugarriak desberdinak izanik, datuak biltzeko modua desberdina da atal bakoitzean.

### 2.1.1. Galdeketak atala

Orain artean, galdeketen atalari eman diogu lehentasuna, corpus zabaletan baino errazagoa baita guk prestatutako galdeketen bidez aztergai ditugun fenomeno linguistikoei dagozkien adibideak eskuratzea. Baina literatur corpusetan nahiz bestelako iturri batzuetan aurkitutako adibideak biltzeko atalak izatea ere garrantzitsua iruditzen zaigu, horrek bariazio sintaktikoa erakusten duten adibide kopuru handiagoa eskuratzeko aukera ematen baitigu, eta adibide guztiak izan daitezke adierazgarriak aldakortasun sintaktikoaren azterketarako. Azken finean, ongi hornitutako datu-base bat prestatzea komeni da ondorio linguistiko garbiak atera ahal izateko eta bariazioaren inguruko hipotesiak egiaztatu ahal izateko.

Horrelako azterketetan datu fidagarriak eta baliagarriak eskuratu ahal izateko, garrantzitsua da baliatuko diren metodoak ongi aukeratzea. «Syntactic dialect atlas projects, in particular, have to deal with valid methods for obtaining reliable and sufficient syntactic data from dialects across a large geographical area» (Cornips eta



Poletto, 2005: 940). Hala ere, ez dago metodo egoki edo baliagarri bakar bat, hau da, metodo guztiek dituzte beren abantailak eta desabantailak. «The advantages and disadvantages of both naturalistic methods behind corpus-data and elicitation techniques have been recurrently discussed. It may easily be acknowledged that none of them is exempt of problems» (Carrilho, 2010: 58). Hori horrela izanik, gure ondorio nagusietako bat izan da datuak edota adibideak hainbat informazio-iturritatik biltzea aberasgarria dela, eta datu-bilketarako teknika edo metodo bat baino gehiago baliatzea bakar bat erabiltzea baino egokiagoa dela, bakoitzak dituen alde onak eta txarrak kontuan hartuz.

*Galdeketak* atala osatzeko, aztertuko genituen fenomeno linguistikoak hautatu genituen lehenik; ondoren, fenomeno linguistiko horiek aztertzeko aukera emango ziguten galdeketak prestatu genituen; galdeketa horiek zein herritan eta zein datu-emailerri egin erabaki genuen gero; elkarrizketak burutu eta azken urratsean, grabatutako erantzunen transkripzioa, datu-baseratzea eta etiketatzea egin genituen.

#### 2.1.1.a. Fenomeno linguistikoaren hautaketa

Iparraldeko hizkeren artean aldakortasun sintaktikoa erakusten duten **fenomeno linguistikoak** azterketa bibliografiko sakon batean oinarrituta hautatu genituen. Hizkuntza-tipologiarako esanguratsuak izan daitezkeen aztergaiak lehenetsi genituen eta hauexek dira orain artean, landu ditugun fenomenoak<sup>20</sup>:

- datibo-komunztadura
  - *Ez **dio neho**ri deus eman*
  - *Marik ez **du deus**ik **eman neho**ri*
- postposizio-egiturak
  - *Etxea **mendiari parez da***
  - *Etxea **mendiaren parean da***
- determinatzailearen erabilera
  - *Peio **gezurti bat da***
  - *Peio **gezurtia da***
- *omen* partikularen posizioa
  - *Bere osaba preso **omen da***
  - ***Omen**, bere osaba presondegian da*
  - *Bere osaba presondegian da **omen***
- *ote* partikularen posizioa
  - *Zer gertatu **ote** zaio?*
  - *Zer **ote** gertatu zaio?*
  - *Zer gertatu zaio **ote**?*

---

20. *Literatur corpusak eta Bestelako iturriak* ataletan ere fenomeno linguistiko hauek izan ditugu aztergai.

- zenbatzaile eta aditzaren arteko komunztadura
  - *Ikasle **anitz** ikusi **ditut** aurten eskolan*
  - *Ikasle **anitz** ikusi **dut** aurten eskolan*

### 2.1.1.b. Galdeketen osaketa

Datu-basea prestatu eta fenomeno linguistikoak hautatu ostean, galdeketen atala osatzeari eman genion lehentasuna. Horretarako, **6 galdeketa prestatu** genituen, fenomeno linguistiko bakoitzeko galdeketa bat, eta galdeketak bost galdera mota desberdinekin osatu genituen:

- Itzulpen-ariketak (kasu honetan frantsesetik tokiko euskarara):
  - Il s'est appliqué au travail.*
  - Pierre s'est appuyé contre le mur.*
  - Pierre est berger.*
- Hutsuneak betetzeko ariketak:
  - Ikusi zuelarik ez zuen agurtu. Hori Piarresi ainitz gaitzitu \_\_\_\_\_.*
  - Egunero Piarresi jarraitzen \_\_\_\_\_ .*
  - Piarresi emaztea eritu \_\_\_\_\_.*
- Onarpen-testak:
  - Posible ote da zure hizkeran?*
    - a. Oihanari barna joan da.*
    - b. Oihanean barna joan da.*
    - c. Oihana barna joan da.*
    - d. Oihan barna joan da.*
- Hobespen-testak:
  - Bi aukera hauetatik zein da hurbilenik zure hizkeratik?*
  - Dena beretzat atxiki du.*
    - a. Ez dio nehoi deus igorri*
    - b. Ez du nehoi deus igorri.*
  - Gaur haurrak ongi portatu dira.*
    - a. Ez dute elgarri deus aurtiki/bota*
    - b. Ez diote elgarri deus aurtiki/bota*
- «Erabiltzen duzu hitz hau? Etsenplu bat eman» motako ariketak:
  - Oldartu hitza erabiltzen ote duzu? Etsenplu bat eman.*
  - Gaitzitu hitza erabiltzen duzu? Hala bada, eman etsenplu bat.*

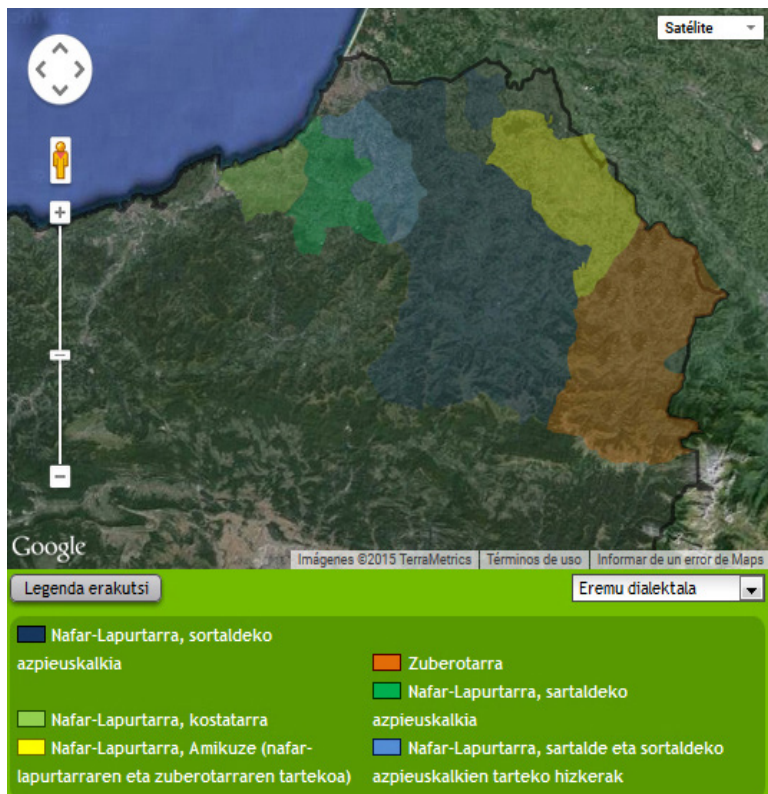
Horrelako ikerketa batean emaitza fidagarriak eta baliagarriak eskuratu nahi badira, ezinbestekoa da datu-emaileei egingo zaizkien galdeketak ongi diseinatzea. Guk batik bat itzulpen-ariketak baliatu ditugu (% 71,94). Izan ere, itzulpen-ariketek hiztun bakoitzaren hizkera modu nahiko naturalean erabiltzeko aukera eskaintzen dute, eta guri berezko hizkera natural hori interesatzen zaigu. Hala ere, ez dago egokiena den galdera mota bakar bat, galdera mota guztiek dituzte beren alde onak eta txarrak. Galdera motaren egokitasuna landu nahi den fenomeno linguistikoaren arabera ere bada, hau da, fenomeno batzuk lantzeko egokiagoak dira galdera mota batzuk, baina beste hainbat fenomeno lantzeko, aldiz, bestelako galderak baliatzea komeni izaten da. Horregatik ez gara itzulpen-ariketak egitera bakarrik mugatu eta onarpen- zein hobespen-testak (% 16,18 eta % 8,63, hurrenez hurren), hutsuneak betetzeko ariketak (% 2,15) eta «*Erabiltzen duzu hitz hau?*» motakoak (% 1,07) ere baliatu ditugu. Sei galdeketen bidez, 278 galdera egin zaizkio guztira datu-emaile bakoitzari (1. taula).

Galdeketa / Galdera mota	Itzulpenak	Hutsuneak bete	Onarpen-testa	Hobespen-testa	Erabiltzen?	Guztira
Datibo-komun.	129	6		7	3	45
Postposizioak	10		20			30
Determinatzailea	53					53
<i>Omen</i> partikula	3		13			16
<i>Ote</i> partikula	4		12			16
Zenbatzaileak	1			17		18
	<b>200</b> (% 71,94)	<b>6</b> (% 2,15)	<b>45</b> (%16,18)	<b>24</b> (%8,63)	<b>3</b> (%1,07)	<b>278</b>

**1. taula. Galdeketa bakoitzean egindako galdera mota desberdinen kopuruak.**

### 2.1.1.c. Herrien hautaketa

Ondoren, **galdeketak egiteko herriak** aukeratu genituen. Eskualde desberdinetako herriak hautatu genituen Iparraldeko hizkeren ikuspegi zabal eta osatu bat jasotzeko helburuarekin. Zehazki, Iparraldeko 33 herritan egin ditugu galdeketak, eta hautatutako herriak Koldo Zuazok (Zuazo, 2008) Iparralderako definitutako sei eremu dialektaletan kokatzen dira: sartaldeko azpieuskalkia, sortaldeko azpieuskalkia, sartalde eta sortaldeko azpieuskalkien tarteko hizkerak, kostatarra (erdialdekoaren eta nafar-lapurarraren tarteko hizkera), Amikuzeko hizkera (nafar-lapurarraren eta zuberotarraren tarteko hizkera) eta zuberotarra (3. irudia). Galdeketak hainbat herritan egin ditugun arren, herri bereko hainbat datu-emaile ere elkarrizketatu ditugu, horrek herrian bertan aldakortasunik gertatzen ote den aztertzeko aukera ematen digularik.



### 3. irudia. Iparraldeko eremu dialektalak BASYQUEn.

#### 2.1.1.d. Datu-emaielen aukeraketa

Hautatutako herrietan galdeketak erantzungo dituzten **datu-emaiileak bilatu genituen** ondoren. Lau adin-tarteko datu-emaiileak elkarrizketatu ditugu: 35 urtez beherakoak, 36-50 urte bitartekoak, 50-70 urte bitartekoak eta 70 urtez gorakoak. Datu-emaielen adin-tartea zehazteak hizkeren bilakaera ere aztertzeke aukera ematen digu. Gazteen artean, euskaraz ikasi duten eta ikasi ez duten hiztunak bilatzen ditugu, eta gazteei egindako elkarrizketek aukera ematen digute ikusteko i) tokian tokiko nahiz familia-giroko hizkerak zenbateraino dirauten bizirik gazteen artean, ii) eskolatzeko hizkeren erabileran edo erabileran ez zuten izan dezakeen eragina eta iii) frantsesa bezalako hizkuntza menderatzaile batekin harreman zuzenean bizitzeak zein eragin izan dezakeen gazteen hizkeran. Momentuz, elkarrizketatu ditugun datu-emaiile gehienak 50 urtetik gorakoak dira, eta kasu guztietan, datu-emaielen anonimotasuna gordetzen dugu. Orain artean, 54 datu-emaiile elkarrizketatu ditugu guztira (2. taula):

Eremu dialektala	Datu-emaileen kopurua
NL, kostatarra	8
NL, sartaldeko azpieuskalkia	10
NL, sartalde-sortaldekoaren artekoa	5
NL, sortaldeko azpieuskalkia	22
NL, Amikuze	3
Zuberotarra	6
	54

2. taula. Eremu dialektal bakoitzean elkarrizketatu dugun datu-emaileen kopurua.

### 2.1.1.e. Elkarrizketak egitea

**Galdeketak egiten eta elkarrizketak grabatzen** Iparraldeko hizkerak ezagutzen dituen bertako pertsona bat ibili da. Garrantzitsua da elkarrizketagileak euskalkien ezagutza izatea, baita galdeketak egiten esperientzia izatea ere, momentu jakin batzuetan datu-emaileei behar duten laguntza eskaini ahal izateko. Elkarrizketak osorik grabatzen ditugu. Ondoren, erantzun bakoitza aparteko audio-fitxategi batean gordetzen da, eta audio-fitxategi bakoitza dagokion erantzunari lotzen zaio, transkribatutako erantzun bakoitza entzun ahal izateko.

### 2.1.1.f. Erantzunen transkripzioa eta etiketatzea

Datu-emaileen **erantzunak transkribatu** eta datu-base batean gordetzen ditugu. Erantzun bakoitzak hiru transkripzio ditu: fonetikatik hurbil den transliterazioa (hau da, erantzuna bere horretan), erantzunari dagokion idazkera formala eta idazkera formalari dagokion testu normalizatua:

Fonetikatik hurbil den transliterazioa: *Deneri arriyak botatu diozkate.*

Idazkera formala: *Deneri harriak botatu diozkate.*

Testu normalizatua *Denei harriak bota dizkiete.*

Hiru transkripzio horiek gordetzeko erabakia hartu genuen, horrek aukera ematen digulako, besteak beste, corpus paraleloak sortzeko, ezinbestekoak direnak aztergai ditugun hizkeren tratamendu automatikoa bideratu ahal izateko. Hain zuzen ere, corpus paralelo hori izan da Iparraldeko hizkeren azterketa eta prozesamendua bideratzen hasteko gure abiapuntua (Etxeberría *et al.*, 2015; Etxeberría *et al.*, 2014, eta Uria *et al.*, 2011). Bestalde, testu normalizatua ezinbestekoa zaigu, baita ere, adibideak automatikoki glosatu ahal izateko, hau da, esaldiko elementu bakoitzari dagokion informazio linguistikoa gehitzeko:

[*Denei*]QUANTuniv-DET-plur-DAT      [*harriak*]-N-DET-plur-ABS      [*bota*]-V  
[*dizkiete*]AUX3P(abs)3P(erg)3P(dat).

Glosak automatikoki esleitzen dira UPV/EHUko IXA taldean garatutako bitarteko informatikoak baliatuz (forma dialektalak analizatzen dituen analizatzailerik oraindik ez dagoenez, beharrezkoa genuen testu normalizatu hori).

**Transkribatutako erantzunak etiketatu** egiten ditugu, hau da, bildutako adibide bakoitzari propietate linguistiko jakin batzuk esleitzen dizkiogu. Propietate linguistiko horiek informazio sintaktikoari buruzkoak dira, eta horiei esker jakin daiteke adibide bakoitza zein motatako bariazio sintaktikori dagokion. Fenomeno linguistiko bakoitzari propietate jakin batzuk dagozkio. Esaterako, datiboaren eta aditz-laguntzailearen arteko komunztadura aztertzeke, ondoko propietate hauek zehazten dira: datibo-argumentua badagoen edo ez, argumentu mota zein den (izen berezia, izenordaina, ezezko polaritadedun elementua, datibo bihurkaria, datibo elkarkaria...) eta datibo-argumentuaren eta aditzaren artean komunztadurarik badagoen ala ez:

*Lanari lotu **zaio**.*

Propietate linguistikoak: Dative BAI  
Dative argument: DP definite  
Agreement BAI

*Arropa **horri** botoi bat josten ahal **zenuke**?*

Propietate linguistikoak: Dative BAI  
Dative argument: demonstrative  
Agreement EZ

Postposizioen adibideei dagokienez, berriz, esleitzen diren propietateek postposizioa sinplea edo konplexua den adierazten dute. Egitura konplexuen kasuan, postposizioaren lehen osagaiaren deklinabide-atzizkia eta postposizio-elementua zehazten dira; egitura sinpleetan, berriz, deklinabide-atzizkia:

*Eni **buruz** itzuli da.*

Propietate linguistikoak: POSTP\_complex  
POSTPdative  
postp-BURUZ

*Oihanean **ibili** dira.*

Propietate linguistikoak: POSTP\_simple  
Inessive

Determinatzailearen galdeketa bidez, datu-emaileek erantzunetan determinatzaile erabili duten ala ez aztertu nahi dugu, eta erabiltzen duten kasuetan, determinatzaile zehaztua, zehaztugabea edota partitiboa erabili duten:

*Pierre **artzain** zen.*

Propietate linguistikoa: DET\_in\_PRED EZ (artikularik ez predikatuan)

*Pierre **artzaina** zen.*

Propietate linguistikoa: DETsing\_in\_PRED BAI (artikulu zehaztu singularra predikatuan)

*Pierre **artzain bat** zen.*

Propietate linguistikoa: DETindef-sing\_in\_PRED BAI (artikulu zehaztugabe singularra predikatuan)

Ebidentzialitateari dagokionez, *omen* eta *ote* partikulak esaldi barruan zein posizioan agertzen diren (Aditza + omen + Laguntzailea, Laguntzailea + omen+ Aditza...) zehazten dugu propietate linguistikoen bidez:

*Bere osaba preso **omen** da.*

Propietate linguistikoa: PRT-omen AUX

***Omen**, egun horretan etorri ziren.*

Propietate linguistikoa: Parenthetical omen

*Gaitzitu **ote** dugu?*

Propietate linguistikoak: V PRT-ote AUX

*Zer **ote** zaio gertatu?*

Propietate linguistikoak: PRT-ote AUX V

Azkenik, zenbatzaileei dagokien galdeketaren bidez bildutako erantzunetan, propietate linguistiko hauek esleitzen ditugu: zenbatzaile mota (*anitz* ala *guti*), zenbatzailea zein argumentutan agertzen den (subjektuan, objektuan ala zehar-objektuan) eta argumentu horren eta aditzaren arteko komunztadura singularra ala plurala den:

*Ikasle **anitz** ikusi **ditut** aurten eskolan.*

Propietate linguistikoak: QUANTmany

QUANTmany\_in\_ARGobj

Agreement-PLUR

*Rugbyko jokalaria **guti** da euskalduna.*

Propietate linguistikoak: QUANTfew

QUANTfew\_in\_ARGsubj

Agreement-SING

Propietate linguistiko hauek oso baliagarriak dira bilaketak egiteko garaian. Izan ere, propietate linguistikoen bidez, ikertzaile bakoitzak interesatzen zaizkion datuak eskura ditzake oso modu errazean: datibo-argumentuarekin komunztadurarik ez duten adibideak, ezezko polaritadedun elementua duten eta aditzarekin komunztatzen ez diren datibo-argumentuak, *guti* zenbatzailearen eta aditzaren artean komunztadura plurala erakusten duten adibideak, eta abar eta abar.

### 2.1.2. Literatur corpusak atala

*Literatur corpusak* atalean, *Galdeketak* atalean landutako fenomeno linguistiko berdinak izan ditugu aztergai. Eta lehen fase honetan, bost autoreren lan hauek aztertu ditugu: Lartzabalen *Lan Guztiak*, Etcheberryren *Frantziako Erregina*, Salaberryren *Ene sinestea*, Thikoiperen *Hotz bero* eta *Jin bezala* eta Etxamendiren *Idazlanen bilduma-1,2,3,4*. Corpus horretan aurkitutako 3.420 adibide gorde eta etiketatu ditugu orain artean (3. taula). Testuak irakurri ahala edota bilaketa zehatzak eginez detektatu ditugun adibideak dira.

<b>Autorea</b>	<b>Datiboa</b>	<b>Postp.</b>	<b>Determ.</b>	<b>Omen</b>	<b>Ote</b>	<b>Zenb.</b>	<b>Nom.</b>	<b>Guztira</b>
Lartzabal	1.009	26	48	66	90	2	67	<b>1.308</b>
Etcheberry	579	8	27		4		31	<b>649</b>
Salaberry	652	98	154	17	4		73	<b>998</b>
Thikoipe	5	1	4	21	31	2	8	<b>72</b>
Etxamendi	194	36	83	35	19	4	22	<b>393</b>
	<b>2.439</b>	<b>169</b>	<b>316</b>	<b>139</b>	<b>148</b>	<b>8</b>	<b>201</b>	<b>3.420</b>

**3. taula. Landutako obra bakoitzean aztergai ditugun fenomeno linguistiko bakoitzeko aurkitutako adibide kopurua.**

### 2.1.3. Bestelako iturriak atala

*Bestelako iturriak* atalean, berriz, beste lan edo proiektu batzuetatik eskuratutako adibideak jasotzen ditugu, betiere, hautatuak ditugun fenomeno linguistikoei lotutako adibideak. Orain artean, bi eratako datu-iturburuak baliatu ditugu, *Euskal Herriko Ahotsak* eta *Norantz* proiektuetako datu-baseetan erakusgai direnak, hain zuzen. Bi proiektu horiek helburu desberdinak dituzte: *Euskal Herriko Ahotsak* taldearen xedea herriko ahozko ondarea eta herri-hizkerak bildu eta hedatzea da. Horretarako, gai konkretu baten inguruko testigantzak grabatzen dituzte; jardun libreak dira. *Norantz* proiektuan, aldiz, galdeketak egiten dira, gaur egungo Iparraldeko euskaldunen gramatika-oinarriak, joera handienetan, zein bidetatik doazen ikusteko helburuarekin. Datuak biltzeko eta lantzeko jarraitu duten metodologia, beraz, ezberdina da kasu bietan: batean gai baten inguruko jardun libreak biltzen dira, eta, bestean, galdera zehatzei dagozkien erantzunak.

*Euskal Herriko Ahotsak* proiektuko adibiderik ez dugu sartu oraindik BASYQUE aplikazioan, Iparraldeko hiztunen testigantza ugari bildu dituzten arren, elkarrizketa horiek ez daudelako oraindik transkribatuta. Transkripzio-lan hori egina denean, testigantza horietan aldakortasun sintaktikoari lotutako adibideak baldin badaude, esaldi horiek etiketatu egingo ditugu adibide bakoitzari dagozkion propietate linguistikoak gehituz (galdeketen bidez jasotzen ditugun adibideekin egiten dugun bezalaxe).

*Norantz* proiektuko datu-basean 59 hiztunengandik jasotako erantzunak gorde ziren. Denetara, 746 galderaz osaturiko galdeketa zabala prestatu zen eta horien artean gure ikerketan aztergai izan ditugun fenomeno linguistikoak ere landu ziren. Guk aztergai ditugun fenomeno linguistikoei dagozkien erantzunak identifikatu eta adibide horietako batzuk gure datu-baseko *Bestelako iturriak* atalera ekarri genituen, 1.581 adibide guztira (4. taula):



Fenomeno linguistikoa	Lapurdi	Nafarroa Beherea	Guztira
Datibo-komuntadura	304	317	621
Determinatzailearen erabilera	379	313	692
<i>Omen</i> partikularen posizioa	3	4	7
<i>Ote</i> partikularen posizioa	20	25	45
Zenbtz.-aditz. komuntadura	121	95	216
	827	754	1.581

4. taula. *Norantz* proiektuan bildu eta geureganatu ditugun adibideen kopurua.

## 2.2. Aplikazioak informazioa kontsultatzeko eskaintzen dituen aukerak

Behin datuak bilduta (aipatutako hiru ataletan), BASYQUE aplikazioak hizkeren arteko aldakortasun sintaktikoaren inguruan askotariko bilaketak egiteko aukera eskaintzen digu. *Galdeketak* atalean, adibidez, informazio-iturri desberdinen, galdeketen, galdeketetako atalen, datu-emaileen jaiotze-urtearen edota adin-tarteen, herrien, eremu dialektalen edota propietate linguistikoen araberrako bilaketak egin ditzakegu, besteak beste. Bilaketa-menuan agertzen diren aukera horietako bakoitza hautatzean, bakoitzari dagozkion atributuak bistaratzen dira; adibidez, *Herria* aukeratuz gero, galdeketak egin diren herrien zerrenda ikusten da; edo *Propietate linguistikoa* hautatuz, erantzunei dagozkien propietateak bistaratzen dira. Erabiltzaileak nahi adina konbinazio edo bilaketa mota egin ditzake, interesatzen zaizkion erantzunak jaso ahal izateko. Horrela, askotariko datuak eskura ditzake, esaterako:

- herri edo eremu dialektal jakin batean (edo batzuetan) bildu diren erantzunak,
- herri edo eremu dialektal batean (edo batzuetan) galdeketa jakin batekin (edo batzuekin) jaso diren erantzunak,
- galdeketa batekin (edo batzuekin) adin-tarte jakin bat duten datu-emaileek emandako erantzunak,
- galdeketa batekin (edo batzuekin) eta herri edo eremu dialektal konkretu batean (edo batzuetan) adin-tarte jakin bateko datu-emaileek emandako erantzunak,
- propietate linguistiko jakin bat edo batzuk dituzten eta aditz-klase konkretu bati dagozkion erantzunak,
- etab<sup>21</sup>.

21. Kontuan izan behar da, baldintza batzuk aukeratzen direnean, beste batzuk menutik desagertu egiten direla. Adibidez, galdeketa bat aukeratzen bada, galdeketa horri dagozkion propietate linguistikokoak bakarrik ikusiko dira menuan, eta ez erantzun guztiei esleitutakoak; edo herri konkretu bat aukeratzen bada, herri horretan bildu diren erantzunei dagozkien ezaugarriak agertuko dira menuan, eta herri horri ez dagozkion informazioa ezkutatu egingo da. Horrela, elkarrekin erlazionaturik dauden baldintzak bakarrik aukeratzen direnez, emaitzarik gabeko erantzunak ekiditen ditugu.

Adibide gisa, beheko irudiak datibo-komunztadurarik gabeko erantzunak eskuratzeko egin den bilaketaren emaitza erakusten du. Hain zuzen, datiboa eta komunztadura-eza duten 3.514 adibide bildu ditugu, orain artean, gure datu-basean (4. irudia):

The screenshot shows the Basyque website interface. At the top, there is a green header with the logo 'Basyque' and the text 'Euskal sintaxiaren datu-basea'. Below the header is a navigation menu with buttons for 'Hasiera', 'Informazioa', 'Kudeaketa', 'Kontaktua', and 'Laguntza'. A dropdown menu is set to 'Euskara'. Below the navigation menu are three tabs: 'Galdeketak', 'Bestelako iturriak', and 'Literatur corpusak'. The main content area features a map of the Basque region with a data popup. The popup contains the following information:

Erantzun kopurua	168
Herria	Beskoitze / Briscous
Eremu dialektala	Nafar-Lapurterra, sortaldeko azpieuskalkia
Eskualdea	Lapurdi Ekialdea
Probintzia	Lapurdi / Labourd
Estaturia	Frantzia / France

Below the map, there are search filters for 'Orria' (1), 'Erantzunak orriko' (5), and 'Erantzunak' (3514). There are also buttons for 'Esportazio aurreratua', 'PDF', 'XML', and 'TXT'. On the right side, there is a 'Bilaketa aukerak' section with a dropdown menu for 'Propietate linguistikoa', a search input field, and a dropdown menu for '>ABS/ERG'. There are also buttons for 'Galdeketa: 01 Dative-agreement', 'Propietate Linguistikoa: Dative - Bai', 'Propietate Linguistikoa: Agreement - Ez', 'Bilatu', and 'Aukerak'.

#### 4. irudia. Datibo-komunztadurarik ez duten adibideen kokapena, eremu dialektalen mapan.

Bilaketa bakoitzean eskuraturako erantzunei dagozkien kokalekuak mapan islatzen dira. Puntu bakoitzaren gainean klikatuz, herri horri dagokion informazioa bistaraten da taulatxo batean. Maparen azpiko taulan erantzun kopurua zehazten da. Gainera, eskuraturako adibideak eta horiei dagokien informazioa *pdf*, *xml* edota *txt* formatuko fitxategietara esporta daitezke; horretarako, *Esportazio aurreratua* atalean erabiltzaileak berak aukera dezake zein datu esportatu nahi dituen.

Halaber, taulatxo horren behealdean, erantzun bakoitzari dagokion informazio zehatua zehazten da, hiru leihotan. Lehenengo leihoa (5. irudia) erantzunari dagokiona da eta hauek dira bertan kontsulta daitezkeen datuak:

- datu-emaileari egin zaion galdera (beheko adibidean, frantsesetik euskalkira itzultzeko esaldia),
- galderaren ingelesezko itzulpena,
- datu-emailearen erantzuna,
- erantzunei dagozkien idazkera formala eta normalizatua,
- idazkera normalizatuari automatikoki jarritako glosak,
- iruzkinak, datu-emailearenak nahiz -biltzailearenak,
- erantzunari dagozkion propietate linguistikoak,
- erantzuna entzuteko audioa.

##### 5. irudia. Erantzunari dagokion informazioa.

Bigarren leihoan (6. irudia) erantzun bakoitzari dagokion galdeketaren inguruko informazioa zehazten da:

- galdeketa bera,
- galdeketako atala,
- galdera mota,
- nork bildu duen erantzuna eta noiz,
- zein meta-kategoriari dagokion,
- gako-hitza, baldin eta galderari gako-hitzen bat gehitu bazaio.

<b>Galdeketa</b>	01 Dative-agreement	<b>Meta-kategoriak</b>	Motion
<b>Atala</b>	1.1.2 THROW class / BOTA-JAURTI klasea	<b>Gako-hitzak</b>	
<b>Galdera mota</b>	Itzulpena		
<b>Datu-biltzailea</b>	Xan Aire		
<b>Erantzunaren urtea</b>	2012		

6. irudia. Galdeketari dagokion informazioa.

Azkenik, hirugarren leihoan (7. irudia), datu-emailearen datuak kontsulta daitezke:

- sexua,
- jaiotze-urtea,
- euskara estandarra menderatzen duen ala ez,
- bizilekua.

<b>Sexua</b>	Gizonezkoa	<b>Estatua</b>	Frantzia / France
<b>Jaiotze-urtea</b>	1940	<b>Probintzia</b>	Lapuri / Labourd
<b>Alfabetatua</b>	Ez	<b>Eskualdea</b>	Lapuri Garaia
		<b>Eremu dialektala</b>	Nafar-Lapurterra, sartaldeko azpieuskalkia
		<b>Herria</b>	Sara

7. irudia. Datu-emaileari dagokion informazioa.

Azkenik, BASYQUE aplikazioaren bidez egiten diren bilaketan **erantzunak** interfazean dagoen **mapan bistaratzen** dira. Mapak atlas sintaktikoen ezaugarri nagusiak dira eta bilaketa bakoitzari dagozkion erantzunak kokatzeko balio dute. Aplikazio honetan, erabiltzaile bakoitzak aukera du gehien gustatzen zaion mapa mota ikusteko (mapa arrunta, satellite bidezko argazkia edota bien arteko konbinazioa), eta mapan zehar mugitzeko nahiz zoom-a erabiltzeko aukera ere badago, Google Maps-ek eskaintzen dituen teknologiei esker.

Halaber, erantzunak bi mapa desberdinetan ikus daitezke: probintzien mugak dituen mapan edota eremu dialektalen mapan. Hain zuzen ere, maparen behealdean aukera dezake erabiltzaileak bistaratu nahi duen mapa. Mapa mota aukeratzean, *Legenda erakutsi* botoia klikatuz zabaltzen den taulan, kolore bakoitzari dagokion eremu dialektala adierazten da. Eta etorkizun laburrean mapak hobetzeko modua aztertzen hasi gara, bilaketa bakoitzeko adibide kopuru gehien eta gutxien zein lekutan dagoen argiago erakusteko.

Esan bezala, gure azterketa Iparraldeko hizkeretara mugatzen da, baina BASYQUE aplikazioan Euskal Herri osoko mapa prest dago edozein euskalkiren inguruko azterketak egin eta emaitzak mapan islatzeko (8. irudia).



8. irudia. Zuazoren (2008) eremu dialektalen mapa, BASYQUEn.

Hemen *Galdeketak* ataleko adibideak erakutsi ditugu. *Literatur corpusak* eta *Bestelako iturriak* ataletan datuak sartzeko, ez dugu galdeketarik prestatu beharrik, ez dugu datu-emailerik aurkitu beharrik eta ez dugu hiztunei elkarrizketarik egin beharrik. Galdeketen atalean datuak datu-emaleei egindako elkarrizketetatik eskuratzen ditugun moduan, *Literatur corpusak* atalean sartzen ditugun adibideak testu idatzietan aurkitzen ditugu, eta, beraz, bilaketen emaitzak ez dira mapa gainean kokatzen. *Bestelako iturriak* atalean sartzen ditugun adibideak, berriz, beste proiektu edota datu-baseetatik geureganatzen ditugu. Kasu horretan, emaitzak bistaratzen dira mapan, proiektu horietan leku jakinetako datu-emaiak elkarrizketatu izan direlako. Baina, oro har, datuak biltzeko, kudeatzeko, kontsultatzeko eta prozesatzeko aukerak berdinak dira hiru ataletan.

### 2.3. Datuak eta ondorioak

BASYQUE aplikazioaren bidez, Iparraldeko hizkeren artean dagoen bariazio sintaktikoa aztertzeko ezinbesteko oinarria izango den datu-base sendo bat osatzen ari gara. Datu-base horrek euskararen sintaxian egitura aldetik oinarrizkoak diren forma sintaktikoen baitan aurkitzen diren aldaerak jasotzen ditu eta hipotesi gramatikalen garapenerako ezinbesteko datu konparatiboak eskaintzen dizkigu.

*Galdeketak* atalari dagokionez, Iparraldeko 33 herritan 54 datu-emailerik egindako 6 galdeketen bidez 13.270 adibide jaso eta landu ditugu guztira, orain artean (5. taula):

	Datiboa	Postp.	Det.	Omen	Ote	Zenbatz.	Guztira
<b>Kostatarra</b>	1.129	146	422	68	75	116	1.956
<b>Sartaldekoa</b>	1.443	192	530	89	101	174	2.529
<b>Sarralde-sortald.</b>	725	95	265	58	49	84	1.276
<b>Sortaldekoa</b>	3.165	441	1.163	214	226	378	5.587
<b>Amikuze</b>	444	55	159	33	33	60	784
<b>Zuberotarra</b>	661	94	265	20	42	56	1.138
	<b>7.567</b>	<b>1.023</b>	<b>2.804</b>	<b>482</b>	<b>526</b>	<b>868</b>	<b>13.270</b>

5. taula. Galdeketa eta eremu dialektal bakoitzeko erantzun kopurua.

Galdeketek duten abantailetakoa bat da posible egiten dutela aipaturiko hautuak zein baldintza gramatikaletan egiten diren aztertzea. Hala, bildutako adibide horiei esker, hainbat ondorio linguistiko ateratzeko eta zenbait hipotesi betetzen direla ziurtatzeko aukera izan dugu. Hauexek batzuk:

Datibo-komunztadurari dagokionez, eta orain arte bildutako datuen arabera, ikusi dugu ohikoagoa dela, eremu dialektal guztietan, komunztadurarik ez egitea aditzaren eta datibo-argumentuaren artean (*Jeanek ez du nehorri igorri libururik* motako adibideak). Ikusi dugu, halaber, kostaldeko azpieuskalkian datibo-komunztadura

egitearen (*Galdetu **dio** ehun euro bere **amari***) eta ez egitearen (*Galdetu **ditu** ehun euro bere **amari***) arteko aldea ez dela beste azpieuskalkietan bezain nabarmena. Badirudi, gainera, datibo-komunztaduraren erabilera igotzen ari dela eta datibo-komunztadurarik ezarena, berriz, gutxitzen; hau da, datu-emaile helduenen artean gazteagoen artean baino ohikoagoa da komunztadurarik ez egitea (elkarrizketatu diren lau datu-emaile helduenen artean, komunztadurarik gabeko 489 adibide bildu ditugu eta komunztadura duten 184 adibide; lau datu-emaile gazteenen artean, berriz, komunztadurarik gabeko 291 adibide eta komunztadura duten 201 adibide). Datibo-argumentua isilduta dagoenean, berriz, beti egin ohi da komunztadura aditzarekin (*Ezpainetarat hurbildu **diote** baso bat ur, Iduritzen zait bota **dakotela harri bat***), ez da komunztadura-ezarik gertatzen.

Determinatzailearen erabilerrari dagokionez, argumentua objektua denean, frantsesezko artikulu zehaztugabe plurala (*J'ai mangé **des** pommes, J'ai vu **des** vaches*) itzultzean, gure datu-emaileek gehiago erabili dute artikulu zehaztu plurala (*Jan ditut **sagarrak**, Ikusi ditut **behiak***), artikulu plural zehaztugabea baino (*Jan ditut **sagar batzuk**, Ikusi ditut **behi batzuk***), eremu dialektal guztietan. Determinatzailearik gabeko objektuak (***Sagar** jan dut, Manexek **ur** edan dik*) Zuberoan bakarrik erabiltzen dira, eta gutxi.

Predikatuaren kasuan, frantsesezko jatorrizko esaldiko artikulu zehaztugabe plurala (*Pierre et Jean sont **des** enfants, Les chiens sont **des** animaux*) itzultzean, artikulu zehaztu plurala (*Pierre eta Jean **haurrak** dira, Xakurrak **alimaliak** dira...*) gehiago erabili dute datu-emaileek artikulu zehaztugabe plurala baino (*Pierre eta Joanes **haur batzu** dira, **Zakurrak** **alimalia batzu** dira*), eremu dialektal guztietan. Eta oro har, predikatuan, artikulu zehaztu singularra (*Peio **gezurtia** da, Pintto **xakurra** da*) erabili ohi da gehien eremu dialektal guztietan, Zuberoan izan ezik; Zuberoan artikulu zehaztugabea erabiltzen da gehiago (*Peio **gezurter bat** duk, Pintto **txakur bat** duk*); artikulu singularra, aldiz, oso gutxi erabiltzen dute; artikulurik gabeko predikatuak erabiltzen dituzte gehien zuberotarrek (*Jean **artzain** zen*); eta artikulurik gabeko predikatuak batez ere Zuberoan erabiltzen dira.

OMEN eta OTE partikulei dagokienez, OMEN partikularekin egitura ohikoena eremu dialektal guztietan 'V omen AUX' da (*Egun horretan **etorri omen** ziren*). Eta OTE partikularen posizioari dagokionez, berriz, kostaldean, sartaldean, sartalde-sortaldean eta sortaldean egitura ohikoena 'V ote AUX' (*Zer **gertatu ote** zaio?*) eta 'ote AUX V' (*Zer **ote** zaio **gertatu**?*) dira; Amikuzen eta Zuberoan, berriz, ohikoena 'ote AUX V' da, eta ondoren 'V ote AUX'. Ordena horiek, ezberrik gabe, laguntzaileak Iparraldeko hizkeretan aditz lexikoaren aurreratzerik gabe agertzeko duten aukerari lotuak dira. Datu-basean aurkitzen diren beste hurrenkera batzuk (hala nola 'AUX V omen', 'AUX V ote', 'V AUX omen' eta 'V AUX ote') partikula ebidentzialari berari zor zaion posizio-aldakortasunarekin lotuak dira.

Azken galdeketari dagokionez, zenbatzaileen eta aditzaren artean komunztadurarik ez egitea (*Gizon horrek auzo **anitzi/gutiri** ekartzen **du** laguntza, **Bi** **bertsolariri** **aditu** **dut***) ez da batere ohikoa, baina topatu ditugu horrelako adibide batzuk sartalde eta sortaldeko azpieuskalkien tarteko hizkeretan.

Hauexek *Galdetegiak* atalean orain artean bildu dugun informazioari esker atera ahal izan ditugun ondorio batzuk. Datu-emaile gazteagoen erantzunak biltzean, interesgarria izango da ikustea egitura sintaktiko hauen erabilera nola aldatzen den adinaren arabera.

Gainontzeko bi ataletan ez dugu oraindik ondorio zehatzik atera. Erantzun gehiago bildu nahi ditugu hori egiteko.

### 3. Euskarazko corpus historikoaren arakatazile sintaktikoa

Euskal hizkeren arteko aldakortasun sintaktikoa aztertzeke garatu dugun BASYQUE aplikazioa ikerketarako tresna baliagarria dela ikusita, hasierako lan hori euskarazko corpus historikoa maila sintaktikoan arakatzeko gai izango den bitarteko bat garatzeko asmoarekin zabaldu dugu.

Beste hizkuntza batzuetan, jada, prestatu dituzte horrelako baliabideak, hau da, testu zaharretan maila lexikoan, morfologikoan nahiz sintaktikoan araketak egiteko aukera eskaintzen duten baliabideak. Horren adibide dira, esaterako, ingelesaren corpus historikoa (*Penn Corpora of Historical English*<sup>22</sup>), portugesearen corpus historikoa (*Tycho Brahe Corpus, a parsed corpus of historical Portuguese*<sup>23</sup>), frantsesaren corpus historikoa (*Modéliser le changement: les voies du français, a parsed corpus of historical French*<sup>24</sup>), islandieraren corpus historikoa (*IcePaHC, Icelandic Parsed Historical Corpus*<sup>25</sup>) edota irlandiararen corpus historikoa (*POMIC, The Parsed Old and Middle Irish Corpus*<sup>26</sup>), guztiak maila sintaktikoan etiketatuak. Hizkuntzari buruzko azterketak egiteko eta hizkuntzen bilakaera ezagutzeko baliabide garrantzitsuak dira guztiak.

Guk ere horrelako corpus bat osatu eta prestatu nahi dugu. Izan ere, euskarazko corpus historikoa sintaktikoki arakatzeko baliabideak prestatzeak aukera emango digu euskal gramatikaren ezaugarri nagusien inguruan azterketa diakroniko sistematiakoak egiteko, egitura gramatikal jakin batzuen bilakaera ezagutzeko eta gaur egungo sintaxi diakronikoaren eta hizkuntzalaritza teorikoaren alorretan egin diren aurkikuntza teorikoekin alderatzeko. Hain zuzen ere, sintaxi-mailan etiketatutako corpus historikoak egitura sintaktikoak mendez mende nola aldatzen joan diren erakutsiko digu.

Horretarako finkatu dugun lan-metodologiak ondoko urrats nagusi hauek barne hartzen ditu: corpusaren osaketa eta prestaketa, corpusaren etiketatzea eta kontsultak egiteko interfazearen garapena.

---

22. <<https://www.ling.upenn.edu/hist-corpora/>>.

23. <<http://www.tycho.iel.unicamp.br/~tycho/corpus/en/>>.

24. <[http://www.voies.uottawa.ca/corpus\\_pg\\_en.html](http://www.voies.uottawa.ca/corpus_pg_en.html)>.

25. <[http://www.linguist.is/icelandic\\_treebank/Icelandic\\_Parsed\\_Historical\\_Corpus\\_%28IcePaHC%29](http://www.linguist.is/icelandic_treebank/Icelandic_Parsed_Historical_Corpus_%28IcePaHC%29)>.

26. <[https://www.dias.ie/index.php?option=com\\_content&view=article&id=6586&Itemid=224&lang=en](https://www.dias.ie/index.php?option=com_content&view=article&id=6586&Itemid=224&lang=en)>.



### 3.1. Corpusa

Lehenik eta behin euskarazko corpus historiko adierazgarri bat osatu beharra dugu, garai eta leku desberdinetako euskalkiak ordezkaturako izango dituen corpusa, hain zuzen. Corpus hori osatzeko, Klasikoen Gordailua (KG)<sup>27</sup> webgunean eskuragarri dauden testu digitalizatuak baliatzeko asmoa dugu<sup>28</sup>. Izan ere, testu horietan egon daitezkeen erroreen detekzioa eta zuzenketa egin beharko dugula jakin arren, testuak egungo grafian digitalizatuta izateak lan handia aurrezten ditu.

Klasikoen Gordailua XX. mendea bitarteko euskarazko testu klasiko ia guztien bilgunea da. 12 milioi hitzetik gorako corpus hori euskalkien bilakaeraren ikuspegi zabala eskainiko digun oinarri sendoa da. Literatur genero desberdinak biltzen dira KGn: narratiba, poesia, koplak eta bertsoak. Guk prosan idatzitako testuak bereizi egingo ditugu gainontzekoetatik, sintaxia aztergai izanik, egitura sintaktikoak desberdinak izan daitezkeelako batean eta bestean. Eta testuak garai, hizkera eta literatur generoa kontuan hartuz sailkatuko ditugu.

Corpusaren tamainari dagokionez, ahalik eta testu gehien aztertzen joatea da gure helburua, corpus etiketatu sendo bat beharrezkoa baita emaitza esanguratsuak lortu ahal izateko. Landuko dugun corpusaren tamaina, ordea, jasotzen ditugun diru-laguntzen arabera izango da. Ondorioz, mende bakoitzeko autore baten edo biren testuekin hasi nahi dugu lanean. Momentu honetan aztergai izango ditugun lanen aukeraketa-fasean gaude.

### 3.2. Corpusaren etiketatzea

Garai eta leku desberdinetako testuekin lan egitean, aberastasun dialektal handiko corpus bat osatzen dugu, baina euskalki eta aldakien ugaritasunak konplexutasuna gehitzen dio lanari. Hala, bi etiketatze-fase bereizten ditugu:

- corpusa normalizatzea helburu duen etiketatzea
- corpusean aztergai ditugun egitura sintaktikoak etiketatzea

Lehen faseari dagokionez, testu zaharren normalizazio-prozesuan aurki ditzakegun arazoak zeintzuk diren aurreikusteko, azterketa pilotu bat egin dugu Axularren *Gero* (1643) eta Jose Antonio Mogelen *Peru Abarka* (1802) lanekin. Literatur obra horiek eskuz anotatzeak lan handia eta denbora asko eskatuko lukeenez, testu zati batzuk etiketatu ditugu lehenik (aipatu lan bakoitzaren % 15), ez-estandarrik diren hitzei dagokien baliokide estandarra esleituz<sup>29</sup>. Etiketatze-lana *Brat* aplikazioarekin egiten dugu (9. irudia).

---

27. <<http://klasikoak.armiarma.eus/>>.

28. Corpus horrek dituen hainbat muga ezagutzeko, ikus Bilbao eta Gómez (2014).

29. Etiketatze-lan honen inguruko xehetasun guztiak aurkeztu berri den Etxeberria (2016) tesi-lanean aipatzen dira.



### 9. irudia. Etiketaturako hitz batzuen adibideak, *Brat* tresnaren interfazean.

Ondoren, eskuz etiketatutako corpus zatia oinarri hartuta, gainontzekoa automatikoki etiketatu da ikasketa automatikoa eta metodo estatistikoak baliatuz. Hala, IXA taldean euskarazko testu ez-estandarrek (zaharrak nahiz dialektalak) normalizatzeko sistema bat osatu da (Ettxeberria, 2016; Ettxeberria *et al.*, 2015; Ettxeberria *et al.*, 2014; eta Uria *et al.*, 2011), normalizazioaren inguruan ikerketak egiteko aukerak zabaldu dizkiguna.

Behin corpusa normalizatuta, testu horietan aztergai ditugun fenomeno linguistikoak etiketatuko ditugu. Aukeratuta ditugu corpusean landu nahi ditugun egitura sintaktikoak. Izan ere, XVI. mendetik aurrera idatzitako testuetan oinarrituta egin diren azterketa teorikoei esker, urteetan zehar aldaketak jasan dituzten forma gramatikal ugari ezagutzen ditugu. Eta ezagutza horretan oinarrituta, hasiera batean behintzat, gure lana honako egitura sintaktiko hauetan zentratuko da: denbora eta modua (Mounole, 2011), determinatzaileak (Manterola, 2015), aditz perifrastikoen egiturak (Mounole, 2011), numero morfologikoaren sorrera eta bilakaera (Ettxeberria, 2014), kasu morfologikoaren eta funtzio gramatikalen parekatzea (Fernández eta Ortiz de Urbina, 2010), forma pronominal zehaztugabeen sorrera (Ettxepare, 2001), aditz laguntzaile jokatuen estatusa eta komuntzadura-ezaugarriak (Gómez eta Sainz, 1995), menderagailuen bilakaera adverbio-sintagmen eremuan, erlatibozko perpausak eta egitura korrelatiboak (Rebuschi, 2000, 2009; Rebuschi eta Liptak, 2009) eta ebidentzialak (Ettxepare, 2010a eta 2010b; Ettxepare eta Uria, 2016), postposizio-egituren garapena (Aurnague, 2001), eta, azkenik, galdegiaren eta aditzaren arteko hurrentasun-hertsidura eta hurrenkerari dagozkion beste hertsidura batzuk (Ettxepare, argitaratze bidean).

Oraindik, baina, ez diogu ekin aztergai izango ditugun egitura sintaktiko horiei dagozkien adibideak corpusean etiketatzeari. Horretarako, egitura horietako batzuk testuen normalizaziorako egiten dugun etiketatzearekin batera anotatzea aurreikusita

dugu, eta fase horretan detektatzen ez diren egiturak corpusean aurkitzeko, berriz, IXA taldean garatutako analizatzaile morfosintaktikoa (Aduriz *et al.*, 2000; Alegria *et al.*, 1996) eta *Treebanka* (zuhaitz sintaktikoen corpora) (Aranzabe, 2008) egokituko ditugu. Izan ere, testu zaharrenek (XVI. mendekoek), batik bat, egungo euskaratik gehien aldentzen diren hizkeretan idatziak direnez, guretzat arkaismoak diren eta gure analizatzaileek ezagutzen ez dituzten forma ugari dituzte, eta, ondorioz, ezinbestekoa izango da tresna batzuk egokitzea.

### 3.3. Kotsultak egiteko interfazea

Etiketatzela amaitzean, eta beste hizkuntzetan egin duten antzera, BASYQUE moduko aplikazio bat garatuko dugu, corpus historikoaren arakatzaille sintaktikoa, aukera emango diguna garai eta leku desberdinetako obra jakinetan agertzen diren egitura sintaktikoak bilatu eta, bilaketa horietan oinarrituta, ondorioak ateratzeko nahiz gure hipotesiak betetzen ote diren ikusteko.

## 4. Ondorioak

Euskara hizkuntza ez-indoeuroparra eta Europako sprachbund-etik aski urrun dagoen hizkuntza izanik (*Standard Average European* delakoa; ikus Haspelmath, 2001; Heine eta Kuteva, 2006; Auwera, van der, 2011), interes handia izan ohi dute haren inguruan egiten diren ikerketek. Orain artean, azterketa asko egin dira euskararen jatorriaren, sorreraren edota antzinatasunaren inguruan. Gure ikerketa, aldiz, hizkuntzaren bilakaera historikoan zentratzen da, hau da, gure hizkuntzak eta hizkerak izan duten bilakaera aztertu nahi dugu, batez ere sintaxiari/gramatikari erreparatuz. Izan ere, gaur egun, oraindik, tradiziozko mintzaira ugari dira gure herrian eta hizkera horiek ia aztertu gabe daude sintaxiaren ikuspuntutik. Ondorioz, Hizkuntzaren Azterketa eta Prozesamenduaren alorrean garatzen ditugun tresnek eskaintzen dizkiguten abantailak eta bitartekoak baliatuz, izaera dialektala duten corpusak maila sintaktikoan aztertu eta automatikoki tratatzeko baliabideak garatu nahi ditugu, baliabide horiek gure hizkuntza-komunitaterako ekarpen interesgarriak egin ditzaketela uste dugulako.

Abiapuntu gisa, BASYQUE aplikazioa garatu dugu, hizkeren arteko aldakortasun sintaktikoaren inguruko adibideak eta adibideei dagokien informazioa biltzeko, kudeatzeko eta kotsultatzeke aukera anitz eskaintzen dituena. BASYQUE aplikazioaren bidez, eta hiru informazio-iturri baliatuz, jaso diren adibideek euskarri sendo bat osatzen dute Iparraldeko hizkeren arteko bariazio sintaktikoaren azterketarako.

BASYQUE aplikazioa tresna baliagarria da dialektologiaren lan-eremuan barrendu eta ekarpen berriak egiteko. Alde batetik, datu ugari eta informazio zehatza gorde daitekeelako bertan, eta, bestetik, askotariko bilaketak egiteko aukera eskaintzen duelako, erabiltzaile bakoitzaren interesen arabera. Hala, bildutako datuetan oinarrituta, aldakortasunaren inguruko hipotesiak betetzen diren egiaztatzeke edota ondorio linguistikoak edota dialektologikoak ateratzeko baliabide aproposa da.

Aplikazio hori, momentuz, Iparraldeko hizkeren azterketara mugatu da. Alde batetik, Iparraldeko testugintzaren izaera dialektala lurraldeko beste hizkuntza-ereduetan baino markatuagoa delako, eta bestetik, gaur egun ez dagoelako Ipar Euskal Herriko hizkeren inguruko informazioa corpusen ikuspegitik modu eskuragarrian biltzen duen datu-base edo plataformarik. Hala ere, BASYQUE prest dago edozein hizkeraren azterketa egin ahal izateko.

BASYQUE datu-basearen hedapen gisa, euskararen corpus historikoaren arakatzailerik bat sortzeko beharra sumatu genuen, hau da, euskal sintaxiaren bilakaera diakronikoa aztertzeko eta euskararen sintaxia urte eta leku jakin batzuetan nolakoa izan zen jakiteko aukera emango digun arakatzailerik sintaktikoa. Horretarako prestaketa-lanak bidean dira oraintxe.

Guk dakigula behintzat, ez dago estandarra ez den euskara sintaxi-mailan aztertzeko/arakatzeko ez corpusik ez eta baliabiderik ere eskuragarri. Horregatik uste dugu artikulu honetan aurkeztutako bi baliabideak ekarpen interesgarriak direla gure hizkuntza-komunitaterako, bai hizkuntzalaritzaren ikuspuntutik (euskal hizkeren azterketa sintaktiko-diakronikoak egiteko, garai bateko eta oraingo egitura sintaktikoak konparatzeko, garai jakinetan erabiltzen ziren forma gramatikal ohikoenak ezagutzeko, etab.), baita HAPren ikuspuntutik ere (erronka berria baita orain artean euskara estandarrean eta estandarrerako egin den lan handia hizkeren azterketarako egokitzea, hau da, estandarrak ez diren egitura sintaktikoak aztertzeko gai diren baliabideak garatzea).

Beraz, esku artean ditugun bi proiektu hauekin euskararen sintaxi diakronikoa bitarteko informatikoen laguntzaz aztertzeko oinarriak finkatzen hasi gara eta alor honetan lanean jarraituko dugu. BASYQUE aplikazioari dagokionez, datu-bilketarekin segituko dugu (orain arte osatu duguna baino corpus zabalago eta sendago bat eskuratu arte) eta horrekin batera, helburu dugun euskarazko corpus historikoaren arakatzailerik sintaktikoa garatzeko xedea dugu, sintaxi diakronikoa aztertzeko oinarriak finkatzearekin batera.

## 5. Erreferentziak

- Aduriz I.; Agirre E.; Aldezabal I.; Alegria I.; Arregi X.; Arriola J.; Artola X.; Gojenola K.; Sarasola K. eta Urkia M. (2000): «A Word-grammar based morphological analyzer for agglutinative languages», in *The Proceedings of the International Conference on Computational Linguistics. COLING 2000* (ISBN: 1-55860-717-X), 1. bol., 1-7, Saarbrücken (Alemania).
- Alegria, I.; Etxebarria, I.; Huldén, M. eta Maritxalar, M. (2010): «Porting Basque Morphological Grammars to foma, an Open-Source Tool», *Finite-State Methods and Natural Language Processing*, 6062/2010 bol., 105-113.
- Alegria I.; Artola X.; Sarasola K. eta Urkia M. (1996): «Automatic morphological analysis of Basque», Oxford University Press (ed.) (ISSN 0268-1145), *Literary & Linguistic Computing*, 11. bol., 4. zk., 193-203.
- Aranzabe, M. (2008): *Dependentzia-ereduan oinarritutako baliabide sintaktikoak: zuhaitzbankua eta gramatika konputazionala*, doktore-tesia (ISBN: 978-84-8438-320-8).

- Aurnague. (2001): *Entités et relations dans les descriptions spatiales: L'espace et son expression en basque et en français*, Mémoire d'habilitation, Université de Toulouse-Le Mirail.
- Auwers, van der J. (2011): «Standard Average European», in B. Kortmann eta J. van der Auwers (ed.), *The Languages and Linguistics of Europe*, 1. bol., Mouton de Gruyter, Berlin, 291-306.
- Bilbao, G. eta Gómez, R. (2014): «Textos antiguos vascos en Internet», *Janus, Anexo 1 (2014)* [on line], 111-121. Interneten eskuragarri: <<http://www.janusdigital.es/anexos/contribucion.htm?id=11>>.
- Carrilho E. (2010): «Tools for dialect syntax: the case of CORDIAL-SIN (an annotated corpus of Portuguese dialects)», in Gotzon Aurrekoetxea eta Jose Luis Ormaetxea (ed.), *Tools for Linguistic Variation*, Universidad del País Vasco, Bilbo, 57-70.
- Cornips, L. eta C. Poletto (2005): «On standardising syntactic elicitation techniques (part I)», *Lingua* 115, 7, 939-957.
- Etxeberria I.; Alegria I. eta Uria L. (2015): «Induction of Phonology and Morphology for the Normalization of Historical Texts», in *Proceedings of the 22nd International Conference on Historical Linguistics*, Naples, 2015eko uztailearen 27-31.
- Etxeberria I.; Alegria I.; Hulden M. eta Uria L. (2014): «Learning to map variation-standard forms using a limited parallel corpus and the standard morphology», *Revista Procesamiento del Lenguaje Natural*, 52 (ISSN edizio inprimatua: 1135-5948) (ISSN edizio digitala: 1989-7553), 13-20.
- Etxeberria, U. (2014): «Basque nominals: from a system with BNs to a system without», in Aguilar-Guevara, Le Bruyn & Zwarts (eds), *Weak Referentiality*, John Benjamins.
- Etxepare, R. (2001): «NZ-bait izenordain zehaztugabeen jatorriaz [Sur l'origine des pronoms indéfinis du type NZ-bait en Basque]», *Lapurdum VI* (ISBN: 2841271560), 93-117.
- , (2010a): «Omen bariazioan», in B. Fernandez, P. Albizu eta R. Etxepare (ed.), *Euskaren sintaxia, Supplement de la Revue International Journal of Basque Language and Linguistics* (ISBN: 978-84-9860-397-2), Université du Pays Basque, Bilbo, 85-112.
- , (2010b): «From hearsay evidentials to samesaying relations», *Lingua* 120 (ISSN: 00243841), 604-627.
- , (argitaratze bidean): «Laguntzailearen izaera eta hurrenkera hertsidurak mendebaldeko testu zaharretan», *Lapurdum XVIII*.
- Etxepare, R. eta Uria, L. (2016): «Microsyntactic variation in the Basque hearsay evidential», in Fernandez eta Ortiz de Urbina (ed.), *Microparameters in the grammar of Basque*, John Benjamins (ISBN: 9789027208309), 265-288.
- Euskaltzaindia. (2004): *Adierazpena: Euskalkien erabileraz irakaskuntzan, komunikabideetan eta administrazioan*, 137. araua.
- Fernández, B. eta Ortiz de Urbina, J. (2010): *Datiboa Hiztegia [El dativo en el léxico]*, UPV/EHU (ISBN: 978-84-9860-433-7), Bilbo.
- Gómez, R. eta Sainz, K. (1995): «On the origin of the finite forms of the Basque verb», in Jose Ignacio Hualde, Joseba A. Lakarra & Robert L. Trask (ed.), *Towards a history of the Basque language*, John Benjamins, Amsterdam/Philadelphia, 235-274.
- Haspelmath, M. (2001): «The European linguistic area: Standard Average European», in M. Haspelmath, E. König eta W. Oesterreicher (ed.), *Language Typology and Language Universals. An International Handbook*, Mouton de Gruyter, Berlin.
- Heine, B. eta Kuteva, T. (2006): *The Changing Languages of Europe*, Oxford University Press, Oxford.
- King, A. (1983): «Batua bai, euskalkiak ere bai!», *Argia*, 1983-II-13.
- Lakarra, J.A. (1996): *Refranes y Sentencias (1596): ikerketak eta edizioa*, Euskaltzaindia,

- Bilbo.
- Larrigan, L.M. (1993): «Euskara batua eta euskalkiak. Aintzat hartu beharreko praktikak», *Jakin*, **76**, 11-39.
- Maia, J. (2001): «Euskara Batuaren eta euskalkien arteko harremanen bilakaera irakaskuntzaren ikuspuntutik», in EUSKALGINTZA XXI. MENDEARI BURUZ. Euskaltzaindiaren nazioarteko XV. Biltzarra, *IKER-19*, Bilbo, 2001eko iraila, 119-128.
- Manterola, J. (2015): *Towards a history of Basque morphology: articles and demonstratives*, UPV/EHU doktorego-tesia.
- Mounole, C. (2011): *Le verbe basque ancien: étude philologique et diachronique*, UBM-UPV/EHU doktorego-tesia.
- Rebuschi, G. (2009): «Basque correlatives and their kin in the history of northern Basque», in Aniko Liptak (ed.), *Correlatives cross-linguistically* (ISBN: 978 90 272 892 1), John Benjamins, Amsterdam, 81-130.
- Rebuschi, G. (2000): «A propos d'une construction non-standard du basque», *Lapurdum V - Etudes basques*, UMR 5478 : CNRS, Université Michel de Montaigne Bordeaux III, Département interuniversitaire d'Etudes basques de Bayonne, 237-282 <<https://artxiker.ccsd.cnrs.fr/artxibo-00000069/document>>.
- Rebuschi eta Liptak. (2009): «Les deux types de relatives libres détachées à gauche en basque du nord», in D. Apotheloz, B. Combettes & F.N Eveu (ed.), *Les linguistiques du détachement, Actes du colloque international de Nancy*, Peter Lang, Sciences pour la communication, Berna, 2006ko ekainaren 7-9a, 441-456.
- San Martin, J. (1998): *Euskararen ostarteak*, Erein, Donostia.
- Uria, L. eta Etxepare, R. (2012): «Hizkeren arteko aldakortasun sintaktikoa aztertzeke metodologiaren nondik norakoak: BASYQUE aplikazioa», *Lapurdum, Euskal ikerketen aldizkaria* (ISSN elektronikoa 1965-0655), **XVI**, 117-135.
- Uria L.; Hulden M.; Etxeberria I. eta Alegria I. (2011): «Recursos y métodos de sustitución léxica en las variantes dialectales en euskera», in *Proceedings of the Workshop on Iberian Cross-Language Natural Language Processing Tasks (ICL 2011)* (ISSN: 1135-5948), Huelva (Espainia), 70-76.
- Villasante, A.L. (1998): *Euskararen auziaz*, Estella-Lizarrar.
- Zuazo, K. (1998): «Euskalkiak, gaur», *Fontes linguae vasconum: Studia et documenta* (ISSN 0046-435X), **78**, 191-234.
- , (2000): *Euskararen sendabelarrak*, Alberdania, Irun.
- , (2005): *Euskara batua: ezina ekinez egina*, Elkar.
- , (2008): *Euskalkiak: euskararen dialektoak*, Elkar (ISBN: 978-84-9783-626-5).



