

# Hizkuntza-aldagaien aldakortasuna: lehen hurbilketa

Gotzon Aurrekoetxea<sup>1</sup>, Juan Abasolo<sup>2</sup>, Esteve Clua<sup>3</sup>, Iker Usobiaga<sup>4</sup>,  
Miquel Salicrú<sup>5</sup>

<sup>1</sup>UPV/EHUko irakasle ohia, <sup>2</sup>UPV/EHU, <sup>3</sup>UPF, <sup>4</sup>Estatistikan masterduna, <sup>5</sup>UB

Geolinguistikan orain arteko ikerketa gehienek helburua hizkuntzaren bariazioak geografian duen proiektzioa ikertzea izan da. Ikerketa honetan, aldiz, hizkuntza-ezaugarriek marrazten dituzten isoglosen artean patroik geografikoki den ikertzen da. Hizkuntzalaritzan maiz entzun den «hitz bakoitzak bere historia du» esaldia egia den frogatzeko metodologia berri baten lehen urratsak eman dira bertan. Horretarako gramatika-ezaugarriak erabili dira; hain zuzen ere, aditz jokatuari dagozkion ezaugarriak. Bide berria garatzean dialektologia kuantitatiboan ezagunak diren *cluster* analisia eta *Multidimensional Scaling* (MDS) erabili dira: *cluster* analisiaren bidez aztertzen diren aldagaien arteko balizko patroik geolinguistikoak zehazteko eta MDS analisia aldagaien hurbiltasun-urruntasunak ikertzeko. Gramatika-ezaugarrietan ezaugarri batzuek patroik berak erakusten dituzte eta aplikatu den metodologia aurreikuspen interesgarriak eskaintzen dituela ondorioztatzen da.

GAKO-HITZAK: Hizkuntza-bariazioa · *Cluster* analisia · Multidimensional Scaling · Euskara · Aditza.

## Variability of linguistic variables: First approach

Most of the research in Geolinguistics has been dedicated to analyzing the realization of linguistic variation in space. This research, however, proposes the investigation of the divergence between the isoglosses that trace linguistic variables. This is the first step to make sure the well-known saying that «every word has its history» is true. For this purpose, instead of lexical characteristics, grammatical characteristics have been used. As for as the used techniques, well known techniques in quantitative dialectology such as cluster analysis and multidimensional scaling-MDS have been used. Among the most outstanding conclusions, it should be noted that certain characteristics have been found that have same or similar geolinguistic patterns.

KEY WORDS: Linguistic variation · Cluster Analysis · Multidimensional Scaling · Basque language · Verb.

<https://doi.org/10.26876/uztaro.118.2021.5>

*Jasotze-data:* 2020-09-08 *Onartze-data:* 2020-11-08

## 1. Sarrera

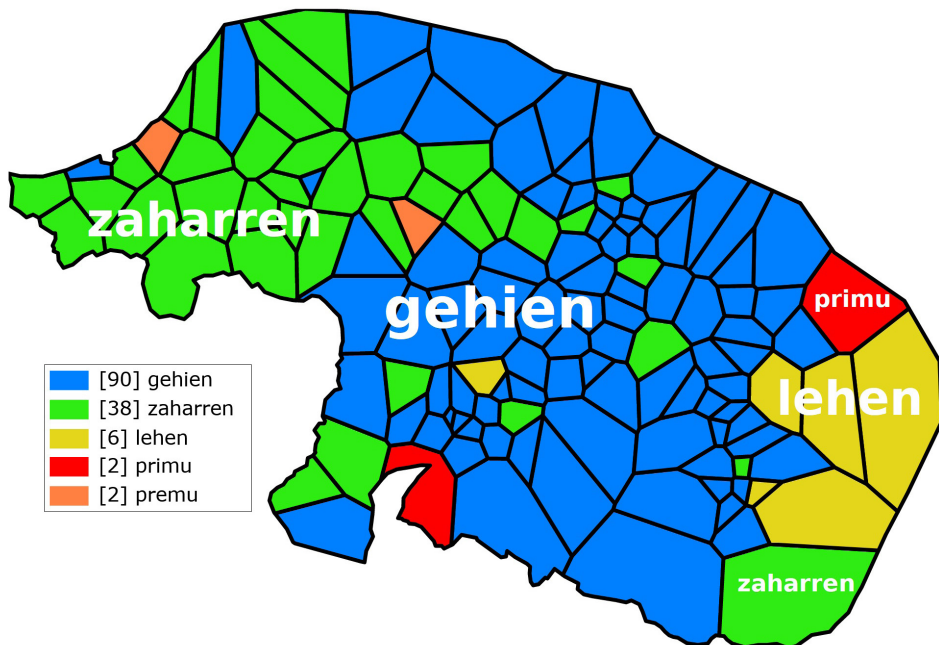
Dialektologian eta hizkuntza bariazio-ikerketetan hainbat lerro landu dira historian zehar, bai bariazio geolinguistikoa, bai bariazio soziolinguistiko edo diastratikoa eta baita bariazio estilistiko edo diafasikoa ere; ahaztu gabe hizkuntzalaritzaren sasoi batean izaera zentrala izan duen bariazio diakronikoa. Kasu guztietan hizkuntzaren bariazioan faktore estralinguistikoen eragina aztertu izan da orain artean: denboraren, geografiaren, gizartearen edo hiztunen xedearen eragina, hain zuzen ere. Bariazio geolinguistikoari erreparatzen bazaio, ikertzen den lurraldearen araberrako hizkuntza bariazioa da ikertu dena; zein diren eremu dialektal bat osatzen duten herriak, zein hizkera diren bi eremuren arteko muga osatzen dutenak edo zein diren hizkera hurbilenak edo urrunenak, adibidez. Bariazio soziolinguistikoaren bezainbestean bariazioa sortzen edo eragiten duten faktore sozialak izaten dira aztergai, zein faktore sozialek zein aldaketa sortzen duen, alegia. Faktore horien artean, ezaguna denez, adina, generoa, arraza, ikasketa, posizio soziala, irabazpideak, etab. dira ikertuenak. Faktore horiek egoera bakoitzean modu desberdinean eragin lezaketenez, beharrezkotzat jotzen da faktore bakoitzaren portaera testuinguru desberdinetan ikertzea. Bariazio estilistikoari dagokionez, solaskide mota eta komunikazio-testuinguru desberdinetan hiztunen jokamolde eta asmoak dira ikergai.

Bariazio geolinguistikoaren ikerketa espazioaren araberrako hizkuntza-egituraketaren analisisan oinarritu da gaurdaino. Eskuarteko diziplinak izendapen desberdinak ezagutu ditu, orokorra den dialektologiaz gain; hala nola hizkuntza-geografia, geolinguistika, geografia dialektala eta hizkuntzalaritza erregionala (*area/linguistics*), besteak beste. Denetan agian geolinguistika da adiera zorrotzago edo laxoagoan erabil daitekeen terminoa; izan ere, historian zehar bi adieratan agertzen da literatura zientifikoan. Hemen geolinguistika edo bariazio geolinguistikoaren deitzen da geografian gaindi agertzen den hizkuntzaren bariazioa izendatzeko edo «Geolinguistics, as conceived by Trudgill and Chambers (1980), is primarily concerned with the relationships between language and geography: the study of language in its geographical context» (Hernández Campoy eta Conde-Silvestre, 2005: 105) edo «the geographical dispersions of linguistic elements» (Chambers, 1982: 1) izendatzeko. Bere sasoiaren, hala ere, beste adiera zehatzago bat ere izan du, Chambers/Trudgill-ek (1980)<sup>1</sup> proposatzen duten bidetik. Hortaz, ez hizkuntza guztietan, ezta ikertzaile guztiengan edota sasoi guztietan ez da berdina eta zehaztasun berez erabiltzen. Dena den, gaur egun ia salbuespenik gabe aurrekoen sinonimo gisa erabiltzen da geolinguistika (Fukushima eta Heap, 2008)<sup>2</sup>.

1. Ezaguna da *Dialectology* liburuak edizio eta berrargitalpen ugari izan dituela. Deigarria da lehen edizioako 12. kapituluaren «Towards Geolinguistics» izenburua hurrengo argitalpenetan (1998kotik aurrerakoetan) aldatu egin zutela eta «Cohesion in Dialectology» ezarri. Eta kapituluaren izenburua aldatzeaz gain, gaiak berak ere aldaketa handia ezagutu du eta jada ez da agertzen *Geolinguistics* hitza kapituluaren zehar; ezta hizkuntza-geografia, hiri-dialektologia eta giza geografia korronteak diziplina bakarrean iker litezkeela eta diziplina hori geolinguistika izenda litekeela ere. Halere, liburu horren 2 edizioen tartean gaia behin baino gehiagotan aztertzen dute, bai Trudgillek (1993, 1992...) eta baita Chambersek (1982) ere, besteak beste.

2. Nazioarteko erabilera, gainera, badira horrela izendatzen diren elkarte eta aldizkariak; hala nola 1965ean sortutako *The American Society of Geolinguistics* ([www.geolinguistics.org](http://www.geolinguistics.org)) eta berak argitaratzen duen *Geolinguistics*; Staffordshire Unibertsitateko *Working Papers in Geolinguistics* aldizkaria; *The International Society for Dialectology and Geolinguistics* (<http://www.geo-linguistics.org/>) eta bere

Dialektologia tradizionalen eremu dialektalak edo geografikoak zehazten dira, zeinetan biltzen diren hizkuntza-kidetasun edo antzeko hizkuntza-ezaugarriak dituzten hizkerak (1. mapa). Mapa hau adibide bat da hizkuntza-geografian eraikitzen diren mapa analitikoak erakusteko; bertan ‘aîné / primogénito’ kontzeptuari euskaraz Iparraldean garai batean zegozkion hitzen eremuak agertzen dira, Bourciez corpusean oinarrituz (Aurrekoetxea; Videgain eta Iglesias, 2007: 13); bost eremutan banatu dira aipatu corpusean jaso diren hitzak: *gehien* erdigunean kokatuta eremu zabalena hartzen duena (90 herritan bildua), *zaharren* nagusiki eremuaren mendebaldean jaso (38 herritan bildua), *lehen* ekialdean agertzen dena, erdialdeko salbuespen bakarrarekin (6 herritan jaso), *primu* (bi herritan) eta bere aldaera den *premu* (beste bitan bildua) guztiz sakabanatuta.



**1. mapa. Bourciez corpuseko ‘aîné / primogénito’ kontzeptuari dagokion mapa berregina.**

Sintesi-mapak ere egiten dira dialektologia tradizionalen, batez ere muga dialektalak zehazte orduan, kontzeptu edo gramatika-ezaugarri baten baino gehiagoren isoglosak aztertuz; horietakoak dira, esate baterako, Allièresek (1994: 51) eraiki zituenak euskalkien arteko muga esanguratsuak kartografiatuz.

Dialektologia kuantitatiboan, datu kopuru handiak eta mapa sintetikoak erabiliz, hizkeren arteko hizkuntza-hurbiltasun gradua zehaztasun handiagoz lortzeko aukera dago gaur egun; are gehiago, zeinahi hizkeraren arteko hizkuntza-hurbiltasuna

aldizkaria den *Dialectologia et Geolinguistica*; *Asian Geolinguistic Society of Japan*; *Géolinguistique* aldizkaria (<https://journals.openedition.org/geolinguistique/>); edota «International Conference on Asian Geolinguistics» biltzarra.

neurtu ere egiten da. Azterketen emaitzak teknika kartografiko automatizatuen bidez erakusten dira. Orain arteko dialektologia kuantitatiboaren ibilbidean nagusiki hizkuntza-ezaugarriak eremu geografikoetan nola islatzen diren eta nolako eremuak sortzen dituzten ikertu da. Aldiz, mapa dialektalak eraikitzeke erabiltzen diren hizkuntza-ezaugarriek beren artean dituzten portaeren ikerketan aurrerapauso gutxi eman da; hots, hizkuntza-ezaugarri bakoitzaren bariazioaren eta hizkuntza-ezaugarrien arteko patroiak ez dira ikertu orain artean, salbuespen gutxi batzuk izan ezik: Pick & Rumpf-ek (2012: 201) adibidez, *variant-based dialectometry* deitzen duten ikuspuntua landuz, *classical dialectometry* edo *dialectometría estándar* (Polian eta Léonard, 2018) kontzeptuaren oposizio legez, herri bakoitzeko erantzun edo aldaera, *variant* haien terminologian, desberdinen kopuruaren araberrako azterketa proposatzen dute. Antzeko bidea ibiliz, Francok (2017) dibertsitate lexikala alor semantikoekin eta kontzeptuen zehaztugabetasunarekin lotu du.

Gure ekarria, hizkuntza-ezaugarriak ikertzen baditu ere, bestelakoa da: ezaguna da hitz edo ezaugarri bakoitzak portaera geografiko desberdina duela edo izan lezakeela, bederen; eta izatez, ia geolinguistikaren ibilbidean hitz edo hizkuntza-ezaugarri bakoitzak portaera geolinguistikoko desberdina duelako ustea nagusi izan da. Bide honetan oso ezaguna da «hitz bakoitzak bere historia du»ela dioen esaera<sup>3</sup>, zeina Iordan-ek (1970: 170) Gilliéron-i egotzi zion eta Jaberg-i (1908: 6; alegia, «In Wirklichkeit hat jedes Wort seine besondere Geschichte») Hinskens-ek (2018: 88)<sup>4</sup>. Egia esateko, hizkuntzalaritzaren historian errepikatzen den eztabaidan kokatzen du Hinskens-ek hau ere: neogramatikoen arauen aurka XX. mendearen hasieran jarri ziren dialektologoen posizioa adierazten du gordinki (2011: 445).

Garai hartako dialektologoek defendatzen zuten aforismoak gaur ere defendagarria izaten jarraitzen duen ikertu nahi da ekarpen honetan. Baina lexikoan oinarritu ordez, gramatika-ezaugarrietan eginez eta egungo baliabide teknologikoak erabiliz; hots, ea gramatika-ezaugarrietan ere ezaugarri bakoitzak «bere historia duen» ala ezaugarri batzuek baduten nola edo hala berdinak edo antzekoak izan litezkeen patroia geografikoak aztertzea. Hitz bakoitzak bere historia duela esatean arau fonologikoen ez dutela hitzetan eragiten ulertzen dugu, hitzak *libre* direla arauetatik, edo eremu bateko arau fonologikoen ez dutela zertan eragin hitzetan. Betiere, esaera hori «historia» hitzak duen zentzu edo esanahi nagusitik aldentuz eta portaera geolinguistikoko propio bezala ulertuz.

Horretarako datu kopuru handiak erabiltzeko aukera profitatuz, dialektologia kuantitatiboaren lanabesak erabiltzeak ematen dituen erraztasunak baliatu ditugu hizkuntza-ezaugarrien portaera aztertzeko. Egun, edozein hizkuntza-ezaugarriren portaera edota ezaugarri desberdinen arteko aztertzeke aukera dago; horrela, jakin daiteke, esate baterako, isoglosa berdintsuak edo berdinenak edo desberdinenak dituztenak zein diren. Era berean, ezaugarrien artean sortzen diren ereduak ikertu daitezke beren arteko patroia argirik agertzen den zehazteke.

---

3. Esaera hau ez da lan honetan bere literaltasunean hartzen, zentzu geografikoan baizik; hots, hitz bakoitzak bere eremu propioa duela eta berak sortzen duen isoglosa izatera propioa duela eta gainerako hitzetatik berezia dela.

4. Gai hau sakonago aztertzeke ikus Malkiel 1967.

Lerro hauetan aurkezten den ikerketa aditz-morfologiako (eta zehazki aditz jokatuena) ezaugarrien artean dagoen bariazioari buruzko ikerketaren lehen hurbilketa da. Ikerlerro berria da, ez arakatua orain artean, behinik behin, gure hizkuntzaren inguru hurbilenean. Lehen hurbilketa honek aditz-ezaugarrien harremanen jokabidea haztatzen du, beren arteko joerak ikertuz eta sortzen dituzten isoglosak berdintsuak diren ala ez zehaztuz. Horretarako *cluster* edo multzokatze-analisia izena duena erabili da.

Irakurleak ez du eremu dialektal jakinetako ezaugarriak zein diren ikasiko, baizik eta zein hizkuntza-ezaugarriak duten soslai berdintsua; hots, isoglosa berak eta eremu dialektal berak eraten dituzten ezaugarriak: bariazio bera sortzen duten ezaugarriak, alegia. Geolinguistikan orain artean ikertu denaren ifrentzua da ikerketa hau, hain zuzen ere. Oro har, geolinguistikan, baina bereziki dialektologia kuantitatiboan, gaur egunera arte, hizkuntza-ezaugarri berdinak dituzten hizkerak edo herriak zein diren ikertu da. Lan honetan, herri beretan erabiltzen diren ezaugarriak zein diren ikertzea proposatzen da. Zein diren eremu berdin edo berdintsutik doazen ezaugarriak edo distribuzio geografikoaren patroia berak sortzen dituztenak. Fokua ez da geografian kokatzen, hizkuntza-ezaugarrietan baizik.

Gaur, teknologiaren laguntzarekin, zehaztasun handiagoz bilatu genezake «hitz bakoitzak bere historia» duen, arau fonologiko bakoitzak bere historia duen, gramatika-ezaugarri bakoitzak bere historia duen ala, ostera, batzuk eta besteak modu batean edo bestean elkartuta dauden, bere osotasunean ez bada ere. Hain zuzen ere, aipatu teknika erabiliz, *cluster* edo multzokatze-analisia, hitz zein arau fonologiko desberdinen jokabidea azter daiteke. Oso garrantzitsutzat jotzen da hizkuntza-ezaugarri bakoitzak bere historia duen ziurtatzea eta ondorioz ez den patroia geolinguistikorik ala ezaugarrien artean badiren patroia batzuk, ezaugarri batzuk batera eta beste batzuk bestera joateko. Esate baterako, aditz-aldagai batzuek talde homogeen bat osatzen duten eta jarraibide geografiko berdin edo antzekoak dituzten. Zenbait gramatika-ezaugarrien artean partekatutako patroiak daudela egiaztatuko balitz hizkuntzalaritzak urrats bat gehiago emango luke hizkuntzaren bariazioaren ikerketan, gutxienez mende bat iraun duen uste mugiezina gaindituz.

Beraz, ikerketa prozedura metodologikoetan kokatzen da. Erakutsi nahi da ea hemen aurkezten den metodologia baliagarria den esku artean dugun gaiari irtenbide berriak proposatzeko. Ez da, ondorioz, gaia bere sakontasunean aztertzea eta behin betiko emaitzak aurkeztea lanaren helburua. Zehaztasunen artean, azkenik, ikerketa gaur egungo datuetan oinarrituz burutu dela esan behar da; badakigu jakin, historian zehar ezaugarriak aldatuz doazela, diakroniak ongi erakutsi duen legez, baina ekarpen honek egungo hizkuntza-egoeraren analisia izan nahi du, ez besterik. Bigarren atalean ikerketarako metodologia aztertzen da, hirugarrenean emaitzak analizatzen dira, laugarren analisi honen baliagarritasunaz hitz egiten da eta azkenengoan ondorioak ateratzen dira.

## 2. Metodologia

Datu-basea eraikitzeko erabilitako irizpideak hauek dira: datuen oinarria *Euskararen Herri Hizkeren Atlasa-EHHA*ko VI. liburukiko adizki jokatuak dira (Euskaltzaindia, 2014). Sortu den datu-baseak 136 aldagai<sup>5</sup> biltzen ditu, zeinetatik 102 aditz laguntzaileak diren eta 34 aditz trinkoak. Aditz laguntzaileen artean argumentu bakarreko eta bi argumentuko *izan* aditzari dagozkionak, bi eta hiru argumentuko *edun* aditzari, eta *\*edin* eta *\*ezan* aditzei dagozkienak bildu dira. Trinkoetan *etorri*, *joan*, *ibili*, *egon*, *eraman*, *jakin*, *ekarri* eta *eduki* aditzei dagozkien adizkiak bildu dira. Irakurlea erraz konturatuko denez, den-denak aditz jokatuak dira.

Bi bide, behintzat, badira adizkien azterketa egiteko: adizkiak bere horretan erabiltzea edota adizkietan agertzen diren marka gramatikalak (erroa, pertsona-markak...) erabiltzea. Gure kasuan lehen bidea aukeratu dugu, baztertu gabe etorkizunean bigarren aukera ere baliatzea. Hautu honek ez du baztertzen bigarrena, baina ekarpen honetan indarrak teknika kuantitatiboetan ezartzea izan da apustua. Teknika horiek baliagarritasuna erakusten badute, etorkizunean gaia berrartzeko beharra ekarriko du, nahitaez.

Oinarritzko material horrekin datu-basea sortzean zenbait faktore izan dira kontuan. Bi hauek dira garrantzitsuenak:

a) Datu-basean hutsune gutxi izan dadin, ez dira jaso 10 herritan baino gehiagotan erantzunik ez duten aldagaiak. Adizki baten datuetan 10 herritan baino gehiagotan ez bada jaso aditz jokaturik, adizki hori ez da datu-basean ezarri. Hutsuneez eragiketa kuantitatiboetan sortzen duten distortsioa handia izan ohi da; eragin hori gutxitzeko hartu da irizpide hau.

b) Datuek etiketatze-prozesu bat jasan dute; hots, ez dira analisisian datu hutsak erabili; ez dira informatzaileek ekoiztiko formak bere horretan erabili. Informatzaileengandik jasotako formek hizkuntza-orrazketa bat jasan dute, zeinaren bitartez ñabardura fonetikoak ezabatu diren, eta morfemetan oinarrituz etiketa bakoitza lortu da. Etiketatze-prozedura izatez, hizkuntza-orrazketa da. Eraldaketa honetan adizkietan jokatzen duten morfemei (erroa, pertsona-markak, pluralgileak...) eman zaie garrantzia. Beraz, informatzaileengandik jasotako forma gramatikalek jasan duten hizkuntza-orrazketaren ondoren sortu diren formei deitzen zaie hemen etiketa. Aditz-aldagai bakoitzak etiketa kopuru desberdina izan lezake, erroen, pluralgileen, morfemen posizioaren... ugaritasunaren arabera.

Datuetan nabarmentzen diren ezaugarrien artean aldaera bat baino gehiago dituzten aldagaiak gailentzen dira: aldagai batzuetan herri guztietan erantzun edo aldaera bakarra dute eta ondorioz erantzun unibertsalak (monolexematikoak) dira; eta, hortaz, etiketa bakarra izango dute; beste batzuek, ordea, herri bakarreko aldaerak erakusten dituzte. Hala ere, gehientsuenek aldaera bat baino gehiago dituzte eta usu aldagai bakoitzean 10 aldaera desberdin bildu dira.

---

5. Ikerketa honetan *aldagai* eta *ezaugarri* sinonimotzat jo dira. Aldagai edo ezaugarri bakoitzak aldaera bat edo gehiago izan ditzake, hala nola \*edun [+orain, nik-hura] aldagaiak *dut*, *det*, *dot* aldaerak. UZTARO 118, 95-113

Hizkuntza-aldagaien analisirako teknika kuantitatiboak erabili dira: *cluster* analisia eta *Multidimensional Scaling* deituak. *Cluster* analisiak aztertzen diren aldagai guztien *cluster* hierarkikoa egiten du eta aldagai guztiak hierarkikoki egituratuta dendrograma batean ezartzen ditu, zeinetan patroï geografiko berak dituzten aldagaiak elkarrekin agertzen diren eta patroï desberdinak dituztenak hierarkiaren beste muturrean kokatzen diren. *Cluster* edo multzokatze-analisia dialektologia kuantitatiboaren lanabes indartsua da eta orain artean hizkerak sailkatzeko tresna ezin hobea bihurtu da. Funtsean bere prozedura honakoa da: hizkuntza-distantzia jakin baten arabera hizkeren arteko distantzietan oinarrituz, horiek berdintasunen edo antzekotasunen arabera multzokatzen ditu. Lehen multzoan hizkuntza-ezaugarri gehien duten hizkerak biltzen dira, ondoren bigarren hizkuntza-ezaugarri gehien dutenak eta horrela hizkera guztiak hierarkia batean ezarri arte. Teknika honen azken emaitza dendrograma izena duen irudian agertzen da (ikus 1. irudia). Dendrograman linguistikoki berdinenak diren (komunean ezaugarri gehien duten) hizkerak elkarren ondoan eta linguistikoki desberdinenak urrunen kokatzen dira. Hizkeren sailkapenak egiteko azken 30 urteotan erabili den teknika da. Hizkerak sailkatzeko erabili den teknika hau kontzeptualki ez dago arazorik hizkuntza-ezaugarriak sailkatzeko ere erabiltzeko. Hori eginez gero, hizkera beretan erabiltzen diren hizkuntza-ezaugarriak bilduko liriateke lehen eragiketetan eta hizkera komunik ez duten hizkuntza-ezaugarriak azkenik. Hizkereri aplikatuz gero, hizkuntza-ezaugarri berdinak edo antzekoak dituztenak biltzen dira multzoetan, eta, hizkuntza-ezaugarri aplikatzen zaienean, hizkera beretan erabiltzen direnak bilduko dira multzoetan. Hau da, gure ezagutzan, hizkuntza-ezaugarri *cluster* analisia ezarri zaien lehen aldia.

Lehen irudiaren behealdean aztergai diren hizkuntza-aldagaiak ikusten dira (ikerketan honetan 136 aldagai aztertzen dira, gorago aipatu bezala). Aldagai bakoitzari zuhaitz hierarkikoaren adar edo marra (bertikal) bat dagokio. Dendrogramaren eskalaren 0 mailan (eskala irudiaren ezkerrean kokatzen da) 136 marra bertikal agertzen dira. Lehen eragiketa estatistikoak antzekotasun handiena duten bi aldagai biltzen ditu multzo edo *cluster* batean. Hurrengo urratsean berriro eragiketa bera egiten du; hots, hurrengo profil geografiko hurbilen edo antzekoenak sortzen dituzten aldagaiak biltzen ditu. Eta horrela jarraitzen du prozedurak, harik eta aldagai guztiak multzo bakarrean bildu arte.

Irakurlea erraz ohartuko denez, aldaera bakarra duten aldagaiak sortzen duten banaketa geografikoa sinplea da, hain zuzen ere aldaera bakar horrek espazio osoa hartzen duelako. Aldiz, aldaera bat baino gehiago dituzten aldagaien kasuan sortzen dituzten banaketa geografikoak oso desberdinak izan daitezke, aldaera bakoitza eremu desberdinetan izan daitekeelako ezaguna eta erabilia, eta horregatik beragatik bakarrik hizkuntza-eremu desberdinak konfiguratu.

Aldagaien arteko distantzia neurtzeko RIV distantzia-unitatea erabili da (Relative Index Value) (Goebel, 1992: 436-438). Distantzia-unitate hori nominal edo kategorikoa da eta datu lematizatu edo etiketatutak neurtzeko sortua da. Datu lematizatu edo etiketatuetan aldaeren arteko desberdintasunak morfologikoak edota etimologikoak izan ohi dira (Aurrekoetxea, Gaminde, Ormaetxea eta Videgain, 2019: 23-25). Desberdintasunak zenbatzean aldaeren arteko desberdintasunak 0 edo 1 izan ohi



dira (eta ehunekoetan denean % 0 - % 100). Saillkapen hierarkikoari dagokionez, *Ward* algoritmoa erabili da. Algoritmo hori aukeratzea ez da kasualitatea izan. Hain zuzen ere, algoritmo hori beste batzuekin alderatuz (*Complete Linkage* edo *Furthest neighbour*, *Single Linkage* edo *nearest neighbour*, *Average Linkage*, *Weighted Pair Group Method using Arithmetic averages-wpgma*, *Weighted Pair Group Method using Centroids-wpgmc*, *K-means...*) aseptikoentzat hartzen da eta saillkapen egokienak lortzen ditu (Kuiper eta Fisher, 1975; Heeringa eta Nerbonne, 2001).

Zenbat eta beherago bildu aldagaiak, orduan eta patroï argiagoak egongo dira. Zaila da ikerketaren fase honetan patroï geografikoak badirela edo ez direla erabakitzea, hain zuzen ere oraindik ez delako hasi baino egin gai honi buruzko ikerketa. Une honetan esan daitekeena honakoa da: dendrogramaren behealdean biltzen diren aldagaiak ikertzen jarraitu behar da, dendrogramako multzo horietako aldagaiak mapetan ezarri eta bakoitzak marrazten duen konfigurazioa ikertu.

Bigarren teknika kuantitatibo bat ere jasan dute oinarrizko datuek, *cluster* analisisiaz gain: *Multidimensional Scaling* (aurrerantzean MDS) deitu analisia, hain zuzen ere. *Cluster* hierarkikoek hizkeren arteko muga argiak marrazten dituzten arren, beti ez da errealitatea hala izaten; batzuetan muga argiak izan ordez *continuum* bat izaten da, zeinen bidez eremu dialektal batetik besterako joana leuna izaten den, eta ezaugarriak poliki-poliki aldatuz joaten dira. Muga mota bi horiek modu argiago batean ikusteko MDS teknika oso egokia da. Teknika horrek hizkuntza-berdintasun eta -desberdintasuna hurbiltasun-urruntasun planoan ezartzen du: hurbiltasunak mugarik eza edo *continuum*a adierazten duen bitartean, urruntasunak muga dagoela erakusten du (ikus 8. irudia).

### 3. Datuen analisia

Datuen analisia oinarrizko datu-baseari gorago aipatu den hizkuntza-distantzia aplikatuz lortutako distantzia-matrizean oinarritzen da. Ondoren, distantzia-matrizea *cluster* analisiaren abiapuntu bihurtzen da, zeinen bidez aldagaien arteko similaritateak zehazten diren eta beren artean hierarkiarik den begiratzen den. Azkenik, MDS teknika erabiliz, analisi ez-hierarkikoa ezartzen zaio distantzia-matrizeari, hizkuntza-aldagaion arteko harremanak espazioan ikusarazteko.

#### 3.1. Hizkuntza-aldagaien arteko distantziak

Distantzien matrizeak hizkuntza-aldagaien arteko desberdintasunei buruzko informazioa eskaintzen du (1. taula). Erakusten den taula osoaren zatia baino ez da: bertan lehen 10 aldagaiak agertzen dira lehen zutabea eta horiez gain beste bi lehen lerroan. Ikusten denez, taula honek sarrera bikoitza du, aldagaiak zutabea eta lerroan agertzen baitira. Zutabea agertzen diren aldagaiak hauek dira: *naiz*, *haiz*, *da*, *gara*, *zara*, *zarete*, *dira*, *nintzen*, *hintzen* eta *zen*. Horiez gain, *ginen* eta *zineten* aldagaiak agertzen dira lerroan. Ikus daitekeen bezala, gramatika-aldagaien similaritate-matrizean aldagaiak elkarren artean konparatzen dira, bakoitzak duen patroï geolinguistikoa araberan. Taulan agertzen diren aldagaien arteko desberdintasun edo distantziak ehunekoetan agertzen dira. Taula hau ohikoagoak



diren hirien arteko distantzien taula bezala ulertu behar da; kasu honetan hirien arteko distantziak neurtu ordez hizkuntza-ezaugarrien arteko distantziak neurtzen dira.

Erakusten den matrize zati honetan aldagaien arteko desberdintasunak % 15,08tik % 50,3ra artean kokatzen dira; hurbilen dauden aldagaiak *zara* eta *gara* dira (bien arteko desberdintasuna % 15,08 da); ondoren *zen* eta *naiz* aldagaiak dira (% 16,09). Eta distantzia handiena *da* eta *zarete* aldagaien artean jasotzen da (% 50,3). Hau da, *zara* eta *gara* aldagaiak dira beren aldaerek patroï geolinguistiko antzekoenak dituztenak; aldiz, *da* eta *zarete* aldagaiak dira desberdinenak dituztenak. Lehen kasuari erreparatzen bazaio, alde batetik *zara* aldagaiaren ondoko aldaerak agertzen dira datu-basean *zara*, *zera* eta *zira* (bokalarteko -r- galdu zein ahoskatua izan), eta *gara* aldagaiaren aldaerak, berriz, *gara*, *gera* eta *gira* (hemen ere -r- bokalartekoa ahoskatu zein ahoskatugabea izan). Dialektologian zertxobait ibili denak ere badaki *zira* aldaeraren eremuan arruntena *gira* aldaera aurkitzea dela eta antzera gertatzen dela gainerako aldaerekin. Bigarren kasuan, *da* eta *zarete* aldagaiena, oso aldaera kopuru desberdina duten aldagaiekin aurkitzen gara: lehena unibertsala den (herri guztietan aldaera bera erabiltzen den) bitartean, bigarren aldagaiaren aldaera kopurua oso handia da: *zarete*, *zerate*, *zare*, *zaree*, *zarezte*, *zirade*, *zaizte*, *zirezte*... Erraz ondoriozta daiteke oso patroï geolinguistiko desberdina dutela.

### 1. taula. Hizkuntza-aldagaien distantzien matrizearen zatia.

|         | naiz  | haiz  | da    | gara  | zara  | zarete | dira  | nintzen | hintzen | zen   | ginen | zineten |
|---------|-------|-------|-------|-------|-------|--------|-------|---------|---------|-------|-------|---------|
| naiz    | 0     | 18,41 | 32,11 | 22,46 | 20,28 | 20,98  | 22,56 | 22,36   | 22,55   | 16,09 | 19,21 | 16,39   |
| haiz    | 18,41 | 0     | 47,07 | 28,39 | 21,95 | 18,03  | 30,16 | 33,31   | 27,74   | 24,56 | 30,34 | 22,64   |
| da      | 32,11 | 47,07 | 0     | 34,43 | 43,53 | 50,3   | 29,96 | 16,63   | 31,53   | 30,56 | 23,71 | 37,57   |
| gara    | 22,46 | 28,39 | 34,43 | 0     | 15,08 | 26,66  | 17,3  | 23,41   | 24,19   | 26,31 | 24,51 | 21,9    |
| zara    | 20,28 | 21,95 | 43,53 | 15,08 | 0     | 19,47  | 22,45 | 30,89   | 26,68   | 23,7  | 28,83 | 20,95   |
| zarete  | 20,98 | 18,03 | 50,3  | 26,66 | 19,47 | 0      | 29,3  | 36,29   | 29,36   | 24,98 | 31,94 | 19      |
| dira    | 22,56 | 30,16 | 29,96 | 17,3  | 22,45 | 29,3   | 0     | 18,18   | 20      | 26,24 | 21,05 | 20,41   |
| nintzen | 22,36 | 33,31 | 16,63 | 23,41 | 30,89 | 36,29  | 18,18 | 0       | 16,87   | 25,3  | 16,47 | 23,81   |
| hintzen | 22,55 | 27,74 | 31,53 | 24,19 | 26,68 | 29,36  | 20    | 16,87   | 0       | 24,71 | 21,22 | 21,57   |
| zen     | 16,09 | 24,56 | 30,56 | 26,31 | 23,7  | 24,98  | 26,24 | 25,3    | 24,71   | 0     | 19,9  | 17,95   |

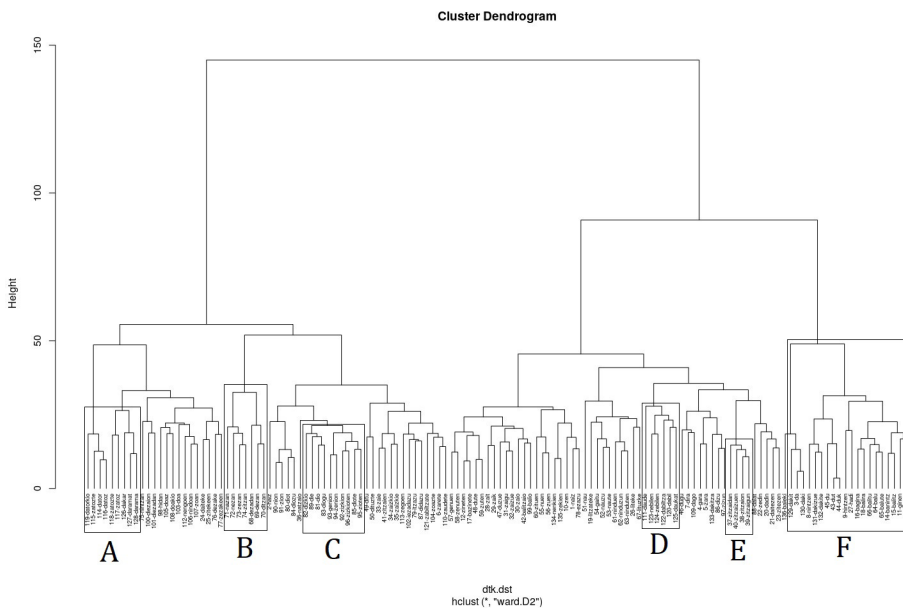
Gehiago sakonduz, *naiz* eta *haiz* aldagaien arteko desberdintasuna (edo distantzia) % 18,41 dela esateak zer adierazten duen azalduko da ondoko lerrootan. Aintzat hartu behar da hizkuntza-ezaugarri edo aldagai bakoitzak portaera geolinguistiko desberdina izan lezakeela. Hizkuntza-aldagai batek aldaera bakarra duenean, aldaera hori unibertsala izaten da (gorago erakutsi den bezala *da* aldagaiaren kasuan). Hizkuntza-aldai hori herri guztietan agertzen da; horregatik esaten da *unibertsala* dela. Aldiz, hizkuntza-aldagaiek aldaera bat baino gehiago dituztenean sortzen dituzten eremu geografikoak berdinak izan litezke (aldaerak herri beretan erabiltzen direnean) edo desberdinak, aldagai bakoitzaren aldaerek ez dituztenean patroï geolinguistiko berak eta ez direnean herri beretan erabiltzen. Aldagai biren artean, *naiz* eta *haiz* adibidez, dagoen desberdintasuna % 18,41 dela esaten denean, honakoa esan nahi da: bi aldagai horiek dituzten aldaeren artean ehuneko horretan bat datozela aldaerok dituzten patroï geolinguistikoak.

Kuriositatea asetzeko baino ez bada ere, aztertzen diren aditz-ezaugarrien taula osoa kontuan hartuz, zentzu honetan aldagai hurbilenak *duk* eta *dut* dira. Horien aldaerek herri gehienetan bat egiten dute; hau da, patroï geolinguistiko antzekoena dute. Hurrengo hurbilenak *daki* eta *da* dira; hirugarren lekuan *nion* eta *zion* aldagaiak aurkitzen dira. Azkenik, laugarren posizioan *zineten* eta *bazinete* aldagaiak aurkitzen ditugu.

Aldagaien portaera geolinguistikoa desberdina denean, hau da, aldaeren patroï geolinguistikoak desberdinak direnean, desberdintasun hori neur daiteke, ikus daitekeen bezala. Are gehiago, neurtu behar da, hain zuzen ere, gaur hori matematikoki egingai delako. Horretarako abiapuntua honako hau da: desberdintasun oro zehaztasun handienaz neurtua izan daitekeela.

### 3.2. Cluster analysis

Datuen analisia egiteko, metodologian esan bezala, *cluster* hierarkikoa erabili da. Ezaguna denez, analisi honen emaitza dendrograma izeneko grafiko batean gauzatzen da (1. irudia).



**1. irudia. Aditz-morfologiari dagokion aldagaien dendrograma: RIV distantzia, Ward algoritmoa.**

Aztertutako hizkuntza-aldagaien dendrogramaren lehenengo azalpen orokorrean esan behar da, batetik, irudiaren behealdean agertzen direla aztertzen diren aldagaiak; bestetik, ezkerrean eskala bat agertzen dela (0-tik 150-erainoko eskala). Eskala horrek aldagaien arteko desberdintasuna markatzen du; hots, aldagaiak zenbat eta beherago bildu, hainbat eta berdintsuagoak izango direla

esan nahi du. Guztira 136 aldagai aztertzen direnez, irudiaren behealdean beste horrenbeste lerro bertikal agertzen dira, aldagai bakoitzari bat dagokio eta. Aldagaiak ordezkatzeko dituzten lerro bertikalak bilduz joango dira dendrograma guztiz osatu arte; zenbat eta lasterrago bildu (zenbat eta eskalan maila apalagoan bildu), orduan eta berdintsuagoak izango dira aldagaien patroio geografikoak.

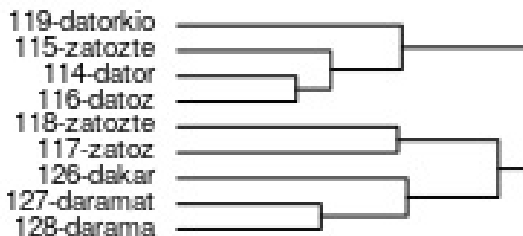
Dendrograma bere osotasunean ikusten bada, aldagaiak bi multzo handitan bananduta daudela ikusten da: ezkerreko multzoan 60 aldagai bildu dira eta eskuinekoan 76. Ikus daiteke ere ezkerreko multzoa trinkoagoa dela; hots, eskuinekoa baino eskalaren maila apalagoan osatzen dela.

Azterketa sakonagoa egiteko maila apalagoako *cluster* batzuk baino ez dira kontuan hartuko; guztira sei, A-tik F-ra irudikatu diren karratuak.

### 3.2.1. Aldagaien A taldea

Irudiaren ezkerrean agertzen den A taldea nahiko txikia da; 9 aditz-aldagaiak osatzen dute, denak aditz trinkoak: 119-datorkio, 115-zatozte, 114-dator, 116-datoz, 118-zatozte, 117-zatoz, 126-dakar, 127-daramat eta 128-darama. Lehenengo sei aldagaiak *etorri* aditzari dagozkio, zarpigarrena *ekarri* aditzaren aldagaia da eta azken biak *eramanenak* (ikus 2. irudia). Aldagaien irakurgarritasunean irabazteko dendrograma osotik A taldea banandu eta taldearen irudia etzanik ezarri da.

Izatez talde hau bi multzok osatzen dute: lehen 4 aldagaiak (119-datorkio, 115-zatozte, 114-dator eta 116-datoz) osatzen dute bata eta azken bostek (118-zatozte, 117-zatoz, 126-dakar, 127-daramat eta 128-darama) bestea. Gainera, bigarren multzo horrek hurbilago ditu beste multzo batzuetan biltzen diren aditz-aldagai batzuk, dendrogramak adierazten duen bezala. Hala ere, bi multzoak batera aurkeztu nahi izan dira nabarmentzeko *clusterrak* zergatik ezarri dituen batzuk besteen ondoan. Izan ere, lehen multzoko aldagaiak harreman estuena bigarren multzokoekin dute.



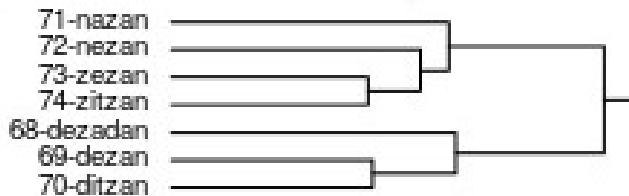
2. irudia. A taldeko aldagaiak.

Aldagai hauek hainbat aldaera dituzte, batzuek gehiago eta beste batzuek gutxiago; 119-datorkio aldagaiak, adibidez, gutxienez aldaera hauek ditu: *datorko*, *yatorko*, *datorkio*, *etorriko jako*, *etortzen zaio*, *ematen dio*, *eldu zaio* eta *eldu dakio*; 118-zatozte-k, beste batzuen artean, *satose*, *satosie*, *zatoz*, *zatozte*, *eldu zaitze*, *zatozai*, *eldu zeate*, *eldu zai* eta *eldu zarate* aldaerak. Aldaera kopuruak desberdinak

izateak ez du esan nahi eredu edo patroi geolinguistiko desberdina izan behar dutenik, baina bai aldaera desberdinek geografian duten eremuarekin. Adibidez, 118-zatozte aldagaiaren *sato*se aldaerak antzeko eredu geolinguistikoa konpartitu dezake 127-daramat aldagaiaren *dar*oat aldaerarekin. Hau da, nahiz eta aldagai batzuek aldaera kopuru desberdina izan, bere aldaerek inguru edo eremu geografiko bera izan dezakete. Egia da, batzuetan aldaerek ez dutela eremu geografiko jakin bat osatzen herri bakar batean baino ez direlako jaso. Herri bakarreko aldaera horiek ez dute parte hartzen eremu geografikoak osatzean.

### 3.2.2. Aldagaien B taldea

B taldea 7 aditz-aldagaik osatzen dute: 71-nazan, 72-nezan, 73-zezan, 74-zitzan, 68-dezadan, 69-dezan eta 70-ditzan. Guztiak *\*ezan* aditz laguntzailearen NNk saileko formak (3. irudia). Ikerketa honetan aztertzen diren bi argumentudun *\*ezan* aditzaren aldagaiak guztira 12 dira: aipatu horiez gain, beste 5 adizki (74-nintzen, 76-dezake, 77-zezakeen, 78-ezazu eta 79-itzazu) ikertu dira. Azken 5 horiek dendrograman sakabanatuta daude.

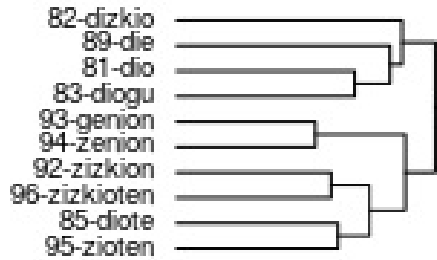


3. irudia. B taldeko aldagaiak.

*\*Ezan* aditzak osatutako talde honek argi uzten du aldagai gramatikal batzuek, behintzat, patroia geografiko bera osatzen dutela; eta, beraz, kasu honetan behintzat, askotan aipatutako «hitz bakoitzak bere historia du» esana ez da betetzen, bere osotasunean, behintzat.

### 3.2.3. C taldea

C taldea 10 aldagaik osatzen dute, denak *\*edun* aditz laguntzailearen NNN saileko indikatiboko formak: 82-dizkio, 89-die, 81-dio, 83-diogu, 93-genion, 94-zenion, 92-zizkion, 96-zizkioten, 85-diote eta 95-zioten (ikus 4. irudia). Badituzte partekatzen dituzten beste ezaugarri batzuk ere: batetik, adizki guztiak zehar-objektu singularrekoak dira, bat izan ezik (die). Eta bestetik, denak batek izan ezik, objektu zuzena singularrean dute.

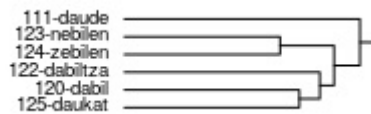


#### 4. irudia. C taldeko aldagaiak.

Talde honek, B taldeak legez, frogatzen du aldagai gramatikal batzuk elkarri oso lotuta daudela eta antzekotasun handiko eremu geografikoak osatzen dituztela.

##### 3.2.4. D taldea

D taldea txikia da eta sei aldagaik osatzen dute, denak aditz trinkoak: 111-daude, 123-nenbilen, 124-zebilen, 122-dabiltza, 120-dabil eta 125-daukat. *Ibili* aditzaren lau aldagai: *nenbilen*, *zebilen*, *dabiltza* eta *dabil*; *egon* aditzaren bat, *daude*; eta beste bat *eduki* aditzarena, *daukat* (ikus 5. irudia).

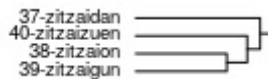


#### 5. irudia. D taldeko aldagaiak.

Talde honen ezaugarri nabarmenena honakoa da: aditz trinkoz bakarrik osatutako taldea dela. Beraietatik bat izan ezik denak intrantsitiboak eta gehienak *ibili* aditzaren adizkiak dira.

##### 3.2.5. E taldea

E taldea oso txikia da, lau aldagaik baino ez dute osatzen: 37-zitzaidan, 40-zitzaizuen, 38-zitzaion eta 39-zitzaigun. Denak dira *izan* aditzaren lehenaldiko NN saileko formak (6. irudia).

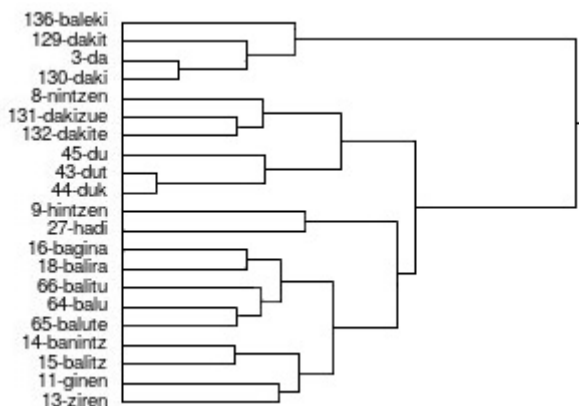


#### 6. irudia. E taldeko aldagaiak.

*Izan* NN saileko iraganeko 5 aldagai aukeratu dira lan honetarako eta lau bildu dira talde honetan; bildu ez den bakarra *zitzaien* da eta urrun ageri da dendrograman; hain zuzen ere *zaie* aldagaiaren ondoan, horrekin baitu antzekotasun handiena marrazten dituzten eremuei dagokienez. Beraz, taldearen ezaugarria *izan* NN saileko iraganeko adizkiak izatea da.

### 3.2.6. F taldea

Taldeen azterketarekin amaitzeko aldakortasun handiena aurkezten duen F taldea dugu. Talde hau oso aproposa da lan honen nondik norakoak ulertzeko. 21 aditz-aldagai osatzen dute: 136-baleki, 129-dakit, 3-da, 130-daki, 8-nintzen, 131-dakizue, 132-dakite, 45-du, 43-dut, 44-duk, 9-hintzen, 27-hadi, 16-bagina, 18-balira, 66-balitu, 64-balu, 65-balute, 14-banintz, 15-balitz, 11-ginen eta 13-ziren. Aldagai horien artean *jakin* aditzaren adizki trinkoen bost aldagai aurkitzen dira (*baleki*, *dakit*, *daki*, *dakizue* eta *dakite*), N saileko *izan* aditzaren 10 aldagai: *da*, *dira*, *ginen*, *ziren*, *nintzen*, *hintzen*, *bagina*, *balira*, *banintz* eta *balitz*; \**edun* aditz laguntzailearen NNk saileko 5 aldagai: *du*, *duk*, *balitu*, *balu* eta *balute*; eta \**edin* aditz laguntzailearen aldagai bakarra: *hadi*. Datu-basean dauden *jakin* aditzaren 8 aldagaietatik, 5 talde honetan daude. Gainera, F taldean ez dago \**ezan* aditzaren aldagairik, ezta \**edun* aditzaren NNN sailekorik (7. irudia).



7. irudia. F taldeko aldagaiak.

Zer dute komunean aditz-aldagai hauek? Arestian aipatu denez, batzen dituen ez da aldaeren kopurua. Aditz desberdinen aldaerek eremu geografiko bera konpartitzea da interesgarriena, eta ez aldagai hauek aldaera kopuru bera ez izatea.

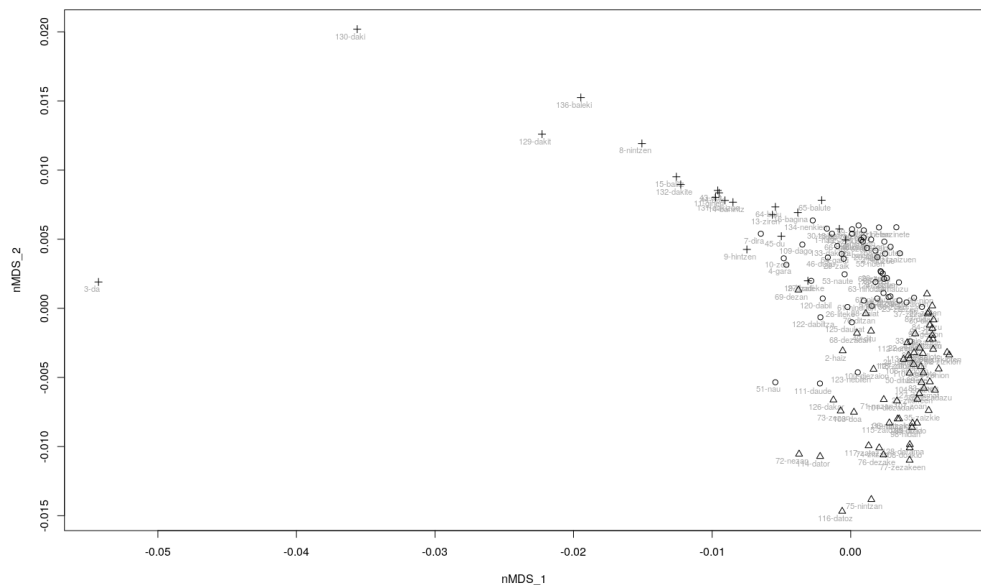
Talde honen barnean ere azpitalde batzuk atzeman daitezke, hala nola azken lau ezaugarri edo aldagaiek osatzen dutena, *banintz*, *balitz*, *ginen* eta *ziren* aldagaiek, hain zuzen ere.

### 3.3. MDS azterketa

*Cluster* deterministak etorkizun handiko emaitzak erakusten ditu hizkuntza-aldagaien, kasu honetan aditz-aldagaien arteko ikerketan. Teknika honek hizkerak modu zatikatuan aurkezten ditu, muga zurruneekin eta beren artean diren desberdintasunaren tamaina ez da erraz antzematen. MDS analisiak, aldiz, aztergai diren objektuak plano batean ezartzen ditu, horretarako objektuak dimentsio gutxitara ekarriz. Hemen aurkezten den irudian (ikus 8. irudia) bi dimentsiotan aurkezten dira.

8. irudian bertan hiru hizkuntza-aldagai talde ikusten dira (aldagaien talde bakoitza kolore batez bistaritzen da). Talderik trinkoena aldagai berdeek osatzen dute; gorriek ere trinkotasun handia erakusten dute, hiruzpalau aldagai izan ezik. Talde urdinean aldagai gehienak batuta agertzen diren arren, aldakortasuna handiagoa erakusten dute, eta bi aldagai (3-da eta 130-daki) nabarmen aldentzen dira.

Orlegiz idatzitako aldagaiek hauen hiru multzo nagusietako bat osatzen dute, hain zuzen ere dendrogramaren F multzoa. Guztira 21 aldagaik osatzen dute multzo hau. Multzo txikia da, gainerako biek in erkatuz gero, baina beste biek baino aldakortasun handiagoa erakusten du MDSk eskaintzen duen irudiaren arabera. Ezaugarri hori, aldakortasun-maila, ez du erakusten *cluster* analisiak, baina garden erakusten du 8. irudiak. Gorriz idatzitako aldagaiek osatzen dute erdiguneko multzoa. Irudiak ondo erakusten ez duen arren, taldea 55 aldagaik osatzen dute eta bertan kokatzen dira, besteak beste, gorago aztertu diren D eta E taldeetako aldagaiek. Aski talde trinkoa da, nahiz aldagai batzuk (51-nau, 111-daude, 123-nebilen ta 22-zedin) aldentuta agertzen diren. Orlegiz idatzitako multzoa dendrogramaren ezkerrean agertzen den multzo berbera da. Multzo hori 60 aldagaik osatzen dute eta bertan kokatzen dira gorago aztertu diren A, B eta C taldeak.



### 8. irudia. MDS ez-metrikoren irudia.

Irudi honek ez du balio aldagai guztien azterketa zehatza egiteko; alabaina, aldentzen diren aldagaien azterketa sakonago bat egitea ahalbidetzen du. Lehenengoak, 3-da *izan* aditzaren indikatiboko orainaldiko hirugarren pertsona singularrak, forma linguistiko bakarra aurkezten du: *da*. Aldentzen den bigarrena 130-daki aldagaia da, *jakin* aditzaren indikatiboko orainaldiko hirugarren pertsona



singularra, eta aldaera bakarra erakusten du: *daki*. Bi aldagai horiek unibertsalak dira eta aldaera bakarra dute. Horiekin batera, aldentzen diren beste hiru aldagai daude: 129-dakit, 136-baleki eta 8-nintzen. Hiru horiek beren artean aldentuta dauden arren, bereiz agertzen dira gainerako aldagaietatik.

#### 4. Eztabaida

Lortutako emaitzak adore-emaileak dira, ezen egiaztatzen baita linguistikoki esanguratsuak diren aditz-aldagai batzuek eremu geografiko bera hartzen dutela. Ezelango ondorio sendo eta orokorrik ateratzeko goiz den arren, unea da erabili den metodoa egokia den eztabaidatzeko eta lehenengo eta behin-behineko emaitzak aztertu eta interpretatzeko.

Lan honetan erabilitako metodo estatistikoen eztabaida bi alderditan zentratu behar da gure ustez: alderdi metodologiko-kontzeptualean eta lortutako emaitzen interpretazioan. Metodologiari dagokionez, eta hau da lan honetako ekarpen nagusia, badira bi puntu kontuan hartu behar direnak: erabilitako distantzia-unitatea eta aditz-aldagaiak sailkatzeko erabilitako metodo estatistikoak. Distantzia-unitateaz den bezainbestean, ikerketa honetan erabilitako distantzia RIV izan da. Distantzia-unitate hori ondo egokitzen zaie lan honetan erabilitako aditz-aldagaien ezaugarriari. Metodologia atalean aipatu denez, ikerketan erabilitako aditz-aldagaiak alde aurretik etiketatuta izan dira. Etiketatze horretan bildu diren aditz-aldakerak osatzen dituzten morfemak baino ez dira kontuan hartu; hortaz, kontuan hartu diren aditz-aldakerak, adizkietako morfemen erabileragatik edo erabilerarik ezagatik bereizten dira euren artean. Horregatik, aldagaiak lematizatu edo etiketatuta ditugunean, egokiena distantzia-unitate nominal edo kategoriko bat erabiltzea da eta ez *string* erako distantzia bat, *Levenshtein* modukoa, adibidez.

Erabilitako metodo estatistikoei dagokienez, alde batetik esan behar da *cluster* analisisa edo sailkapen hierarkikoa egitea egokitzat jotzen dela, metodo horrek aldagai bakoitzaren jokabide geografikoaren arabera hierarkia sortzen duelako eta eremu berdintsuak osatzen dituzten aldagaien multzo edo *clusterrak* sortzen dituelako; hau da, patroi dialektal bera edo antzekoa duten aldagaiak biltzen ditu. *Cluster* teknika orain artean dialektologia kuantitatiboan hizkerak sailkatzeko erabili den teknika izan arren, lan honen ekarpena eta berrikuntza teknika bera hizkuntza-aldagaiak sailkatzeko erabili izana da. Alderdi metodologiko-kontzeptualetik begiratuta paradigma-aldaketa baino ez da: hizkerak sailkatzetik hizkuntza-aldagaiak sailkatzera doan paradigma-aldaketa, hain zuzen ere. Aldaketa horrek, ostera, bide berriak irekitzen ditu aldagaien konfigurazio geografikoen ikerketan. Hau da, gure ustez, ekarpenaren garrantzia.

MDSk mugarik gabeko espazioa irudikatzen badu ere, argiki ikusarazten du aldagaien arteko hurbiltasun-urruntasuna espazioarekin jokatzu. *Cluster* analisisiak multzo beretan ezarri dituen ezaugarri edo aldakerak oso hurbil kokatuta erakusten ditu. Eta hori beste froga bat da aldagai horiek espazio geografiko berak edo antzekoak marrazten dituztela.

Lortutako emaitzak adore-emailetzat edo pozgarriztat hartu behar dira. Aldagai

gramatikalen aldakortasun-ereduen ezagutzan lehenengo urratsa da. Ikerketa-lerro honek arlo honetan aspaldidanik ezaguna den «hitz bakoitzak bere historia du» esaldiaren benetakotasuna aztertu nahi du; hau da, hitz bakoitzak bariazio-eredu propioa duela esaten denean, ea gauza bera defendatu behar den aldagai gramatikalen kasuan ere. Edo, ostera, gramatika-aldagaien artean jokabide geolinguistikoaren eredu berdinak dauden, lan honetan antzematen denez, eta metodo kuantitatiboak erabiliz eredu horiek ikusteko gai garen.

Linguistikoki adierazgarriak diren aditz-aldagaiak osatutako *clusterrak* aurkitu dira; hau da, bariazio-eredu berdinak dituzten aditz-aldagaiak osatutako multzoak. Deigarrienak B, C eta D taldeetako aditz-aldagaiak dira.

Ikerketa honek ez du aditz-ezaugarri bakoitzaren aldakortasun-faktorea kontuan hartu, ezta bere inplikazioa bariazio-ereduak eratzeko ere; hau da, aditz-aldagai jakin batek taldeen osaketan izan dezakeen aldaera kopuruak duen eragina sailkapenean. Adibidez, orainaldiko *izan* [hura-zuei] aditzaren aldagaiak, 27 aldaera dituenak, lotura handiagoa izan dezake aldaera asko dituzten beste batzuekin, gutxi edo bat ere ez duten beste batzuekin baino, hala nola orainaldiko *izan* aditzaren [hura] *da* aldagai bakarra duenarekin.

Prozesuetan edo erregela inplikazionaletan (Bickerton, 1971, 1973; Bailey, 1973) aldagaien artean egon daitezkeen inplikazioak ere ez dira aztertu. Horiek beste ikerketa-lan baterako utzi dira, nahiz eragin zuzena izan dezakeen. Gramatika-aldagaien aldakortasuna ikertzean ikuspuntu kuantitatiboaren ekarria zein izan litekeen aztertu nahi izan da, ez besterik.

Lortutako emaitzak itxaropentsuak diren arren, beste maila linguistikoetan ere frogatu behar dira, izen-morfologian, syntaxian, edo lexikoan, alegia; interesgarria eta beharrezkoa izango litzateke ere beste hizkuntza batzuetako corpusetan eta beste segmentazio dialektal mota bat aztertzea.

## 5. Ondorioak

Jakinik, hemen aurkeztu den ikerketa hizkuntza-aldagaien artean patroio geografikorik den aztertzeko metodologia berri baten lehen urratsa dela, lortu diren ondorioak behin-behinekoak direla ondorioztatu behar da eta hala kontsideratu behar dira.

Lehenengo ondorio gisa frogatu da erabilitako corpusean behintzat, aditz-aldagai batzuek patroio geografiko berdinak edo antzekoak dituztela. Eta, ondorioz, ezin dela frogatu aditz-aldagai bakoitzak bere historia propioa duela; izan ere, aldagai batzuek elkarrekin harremanak dituzte, bariazio-ereduak eta antzeko edo berdinak diren gune geografikoak osatzen dituzte.

Maila *makroan*, dendrogramaren goiko taldeei erreparatuz, ikusten da multzo batzuek aditz-erro bera duten aldagaiak osatzen dituztela, arestian aipatutako B, C eta E taldeak, adibidez; eta maila *mikroan* (dendrogramako behealdeko multzoetan), aldiz, pertsona-morfema jakin batzuek dituzten ereduak agertzen direla (C taldean

agertzen direnak, adibidez).

Ondorio hauek behin-behinekotzat jobehar diren arren, aurreikuspen interesgarriak erakusten dituzte eta zirrikitu berriak irekitzen dituzte hizkuntza-ezaugarrien patroï geografikoen azterketan, izan dadin gramatika-aldagaiei dagokienez, zein lexikoari dagokionez. Abiatu baizik egin ez den bide honetan azterketa sakonagoak egitea beharrezko ikusten da, zeinetan gramatika-aldagai guztiak ikertzea proposatzen den, ondoren gramatika-sailak elkarrekin alderatzeko eta ikusi ahal izateko hizkuntza-sail bakoitzak eredu edo patroï desberdinak aurkezten dituen.

Halaber, ikerketa beste maila batean kokatuz, komenigarria litzateke ezaugarri morfologiko independenteekin (aditz-ezaugarriei dagokienez, esate baterako, edo pertsona-markei, aditz-erroei, pluralgileei...) azterketa berriak egitea lortutako emaitzekin alderatzeko.

Azkenik, aztertzeo gelditzen da ea beste analisi-teknika batzuk erabili beharko liratekeen, hemen baliatu diren *cluster* eta MDS analisisiez gain. Horretan, analisi kuantitatiboan esparrua zabala da eta lagungarri izan liteke hizkuntza-aldagaien patroï geografikoen analisisian.

### Erreferentziak

- Allières, Jacques (1994): «Euskal Herriko Atlas Etnolinguistikoaz», in Rikardo Gómez eta Joseba Andoni Lakarra (ed.), *Euskal Dialektologiako kongresua (Donostia, 1991ko iraila)*, Julio Urkixo Euskal Filologi Mintegiaren Aldizkariaren Gehigarriak, XXVIII, 45-58.
- Aurrekoetxea, Gotzon; Gaminde, Iñaki; Ormaetxea, Jose Luis eta Videgain, Xarles (2019): *Euskalkien sailkapen berria*, UPV/EHU, Leioa.
- Aurrekoetxea, Gotzon; Iglesias, Aitor eta Videgain, Xarles (2007): *Bourciez Bildumako Euskal Atlas (BBEA-1): 1. Lexikoa. [El atlas lingüístico Bourciez: 1. Léxico]*, *Anuario del Seminario Julio Urquijo* 38:2 (2004) [ed. 2007] <<http://www.ehu.es/ojs/index.php/ASJU/article/view/10245/9443>> (Kontsulta: 2019-XII-14).
- Bailey, Charles-James (1973): *Variation and linguistic theory*, Center for Applied Linguistics, Arlington, Va.
- Bickerton, Derek (1971): «Inherent variability and variable rules», *Foundations of Language*, 7, 457-492, <<https://www.jstor.org/stable/25000558>> (Kontsulta: 2019-X-12).
- , (1973): «The structure of polylectal grammar», in Roger Shuy (ed.), *Proceedings of the 23<sup>rd</sup> Annual Round Table*, Georgetown University Press, Washington DC, 17-42.
- Chambers, Jack K. (1982): «Geolinguistics of a Variable Rule», *Discussion Papers in Geolinguistics*, 5, 1-17.
- Euskaltzaindia (2014): *Euskararen Herri Hizkeren Atlas - EHHA*, VI. bol., Euskaltzaindia, Bilbo.
- Franco, Karlie (2017): *Concept features and lexical diversity: A dialectological case study on the relationship between meaning and variation*, doktorego-tesia, University of Leuven & Radboud University, Leuven & Nijmegen.
- Fukushima, Chitsuko eta Heap, David (2008): «A report on the international conference: geolinguistics around the world», *Dialectologia* 1, 135-156, <<http://www.publicacions.ub.edu/revistes/dialectologia1/documentos/369.pdf>> (Kontsulta: 2010-I-30).
- Goebel, Hans (1992): «Problèmes et méthodes de la dialectométrie actuelle (avec application à

- l' AIS)», in Gotzon Aurrekoetxea & Xarles Videgain (ed.), *Nazioarteko dialektologia biltzarra*, Euskaltzaindia, Bilbo, 429-475.
- Heeringa, Wilbert eta Nerbonne, John (2001): «Dialect Areas and Dialect Continua», in David Sankoff, William Labov eta Anthony Kroch (ed.), *Language Variation and Change*, New York, Cambridge University Press, vol. 13, 375-400.
- Hernández Campoy, Juan Manuel eta Conde-Silvestre, Juan Camilo (2005): «Sociolinguistic and Geolinguistic Approaches to the Historical Diffusion of Linguistic Innovations: Incipient Standardisation in Late Middle English», *International Journal of English Studies*, 5 (1), 101-134.
- Hinskens, Frans (2011): «Lexicon, phonology and phonetics. Or: rule-based and usage-based approaches to phonological variation», in Peter Siemund (ed.), *Linguistic universals and language variation*, Mouton de Gruyter, Berlin, 425-466.
- , (2018): «Dialectology and Formal Linguistic Theory: The Blind Man and the Lame», in Charles Boberg, John Nerbonne eta Dominic Watt (ed.), *The Handbook of Dialectology*, John Wiley & Sons, Oxford, 88-105.
- Jordan, Iorgu (1970): *An Introduction to Romance Linguistics*, University of California Press, Berkeley & Los Angeles [DOI: <https://doi.org/10.1017/S0022226700003042>].
- Kuiper, F.K. eta Fisher, L. (1975): «A Monte Carlo Comparison of six Clustering Procedures», *Biometrics*, 31, 777-783.
- Malkiel, Jacob (1967): «Each word has a history of its own», *Glossa: A Journal of Linguistics*, 1:2 (1967), 137-149. Berrargitaratua in *From Particular to General Linguistics: Selected Essays 1965-1978*, John Benjamins, Amsterdam/Philadelphia, 1983, 217-226.
- Pickl, S. eta Rumpf, J. (2012): «Dialectometric concepts of space: Towards a variant-based dialectometry», in Sandra Hansen, Christian Schwarz, Philipp Stoeckle & Tobias Streck (ed.), *Dialectological and Folk Dialectological Concepts of Space*, Mouton de Gruyter, Berlin, 199-214.
- Polian, G. eta Léonard, J.-L. (2017): «Herramientas cuantitativas en el proyecto Alto (Atlas Lingüístico del Tselal occidental) y evaluación de la zona dialectal sur», in L. Orozco & A. Guerrero Galván (ed.), *Estudios de variación geolingüística*, Secretaría de Cultura, Instituto Nacional de Antropología Histórica, Ciudad de México, <[http://jll.smallcodes.com/commons/layout/partUploaderView.jspCM=FILEVIEW&FILE\\_TO\\_UPLOAD=WF\\_10954\\_%5BC+55%5D\\_2013\\_Tselal\\_Herramientas\\_cuantitativas\\_pro.pdf%2B%2B%2Bapplications%5Cwebwork%5Csite\\_jll%5Clocal%5Cdocument%5C010954.%5BC+55%5D\\_2013\\_Tselal\\_Herramientas\\_cuantitativas\\_pro.pdf&TREATASATTACH=no](http://jll.smallcodes.com/commons/layout/partUploaderView.jspCM=FILEVIEW&FILE_TO_UPLOAD=WF_10954_%5BC+55%5D_2013_Tselal_Herramientas_cuantitativas_pro.pdf%2B%2B%2Bapplications%5Cwebwork%5Csite_jll%5Clocal%5Cdocument%5C010954.%5BC+55%5D_2013_Tselal_Herramientas_cuantitativas_pro.pdf&TREATASATTACH=no)> (Kontsulta: 2019-X-20).
- <<http://colfa.utsa.edu/conferences/2010/nwav39/workshops/geolx.pdf>>.
- <<http://geolinguistics.org/>>.

