

# IPARRALDEKO HIZKEREN SAILKAPENA (1): LEXIKOA

Gotzon Aurrekoetxea<sup>2</sup>

EHU

## Abstract

*Using the data from Bourciez collection, this article shows the relationship between the dialect varieties of the Northern part of the Basque Country and their classification. At the beginning of the work, there are texts gathered by A. Bourciez in the end of 19th century; the work continues with the creation of the date-base and the lemmatisation of the words of the texts. The process of varieties' classification starts with the accuracy of linguistic distance between dialect varieties, the search of the focus of dialects and their classification. In this article, the author shows a new methodology to classify Basque dialect varieties. This is the first time that the VDM computer program is used in the Basque dialectology.*

## 1. Sarrera<sup>1</sup>

Aspaldi honetan euskal dialektologian hizkeren sailkapena pil-pilean jarri den eginkizuna bihurtu da. 100 urte abantzu geldirik eta lo egon ondoren arrapiztu, eta indarrez gainera, egin dela ezin da ukatu. Zoriontzeko eta pozarren egoteko abagunean gaude, euskal dialektologiak hartu duen indarra eta erakusten duen grina, beharbada, gainerako diziplina batzuetan baino biziagoa delako.

Ikerketa eta sailkatze prozeduretan ere aberastasun handia erakusten ari da, dialektologian oro har, ezagutzen diren bi joerek hemen ere euren islada erakusten dutelako: ohiko dialektologiaren lanabes eta prozedurak batetik, eta azken 30 urteotan bortizki garatzen diharduten prozedura eta metodologia berriak, bestetik.

Hizkeren azterketetarako jarraibideetan legez, hizkerak sailkatzekoetan ere joera berdintsu edo berak erakusten ari da euskal dialektologia: batetik hizkerak sistema gisa hartzen eta lantzen dituztenak aurkituko genituzke, bestetik hizkerak elkarren arteko loturarik gabeko datu multzo gisa ulertzen dutenak.

Ez da Euskal dialektologian molde bakarrean lantzen eta ikertzen gaur egun. Prozedura eta metodologian anitza eta aberatsa da gure lurra. Badira molde klasi-

---

<sup>1</sup> Ikerketa hau EHUK babestutako 00033.130-H-15286/2003 ikerketa-proiektuaren barruan garatu da.

<sup>2</sup> Ikerketa honen oinarrian dagoen datu-basea G. Aurrekoetxea, Xarles Videgain eta Aitor Iglesiasek sortua da (Videgain eta Aurrekoetxeak egin dute datuen aukeraketa; Aurrekoetxeak eta Iglesiasen gain izan da datu-basearen osatzea).

koak gurago dituztenak, badira euren burua molde berriagoetan gusturago sentitzen dutenak. Nola haiek hala hauek badute alorra.

## 2. Abiapuntua

Ikerketa honen abiapuntua E. Bourciezen «Recueil des Idiomes de la Région Gascogne»eko euskal testuak dira. Testu hauen transkribatze, ikertze eta argitaratzea 2002an Euskal Herriko Unibertsitateak babestutako 00033.130H-14888/2002 ikerketa proiektuaren bidez lortu ahal izan da.

Edouard Bourciez-ek bilduma hau egiteko «seme xahutzailearen parabola» erabili zuen eta herrietako errientei bidali, ministerioko bidea erabiliz, herriko hizkerara itzul zezaten eskatuz (edo hobe esan aginduz). Gaskoi herrian zehar bildutako 4.444 bertsioen artean aurkitzen dira Ipar Euskal Herrian eta euskaraz jasotako 150 bertsio.<sup>3</sup> Testuok, eskuizkribuak, Bordeleko Michel de la Montaigne Unibertsitateko bibliotekan gordetzen dira.<sup>4</sup>

150 euskal testu horiek Ipar Euskal Herrian euskararen eremu osoa hartzen dute, Hendaiaitik hasi eta Eskuilaraino. Ez dira, haatik, herri guztietan testuak bildu ahal izan: batzuetan euskal testuaren ordez gaskoi testua bildu da (hala nola Baiona [1988-1991 testuak], Angelu [1197], Basusarri [1199]; Biarritz [1200], Ahurti [1239], Labetze-Bizkai [1289], Oserain [1296], Montori [1356]); beste batzuetan ez da inolako testurik bildu (hala nola Arrosan, Isturitzen, Behorlegin, Gamarte, Ithorrots-Olaibi eta Berrogaine-Lahüntze). Beste batzuetan, azkenik, baten ordez bi testu desberdin bildu dira: desberdintasunak izan daitezke edo ortografikoak (euskararen edo frantses ortografia erabiliz) edo bestelakoak. Orotara 138 herritako testuak bildu dira bilduman.

Bourciez-en bilduma hau oso garrantzitsua da euskal dialektologiarako bezainbat euskal filologiarako. Izan ere, ez da maiz aurkitzen XIX. mendean bildurik ia herri guztiak biltzen dituen testurik. Idatzia izanagatik ere, ahozko hizkeraren ordezkotza duen testua dela esan eta defenditu behar dugu, testu batean baino gehiagotan itzultzaileak herriko jendearen aitzinean irakurri eta haren onespena jasotakoa dela ziurtatzen duelako.

Honelako testuen garrantziaz ohartu gura dugu, berez edozein hizkuntzatan duten garrantzia areagotu egiten delako literatura klasiko eta dokumentazio zaharrean hain urria den euskararen kasuan. Izan ere, dialektologian edo hizkuntzalaritzan kanpo-lana egiten hasi zirenetik (Georg Wenker-ek Alemanian 1876an hasi zuen postazko inkesta ezaguna hasi zuenetik, alegia) molde desberdinetan bildu izan dira datuak: postazko testuen bidez dela, postazko galdesorten bidez dela, zuzenezko testuen bilketaz eta gaur indarrean dagoen zuzenezko galdesorten bidezkoa; alde bat uzten dugu, datu-bilketa modu bakoitzak dituen mota desberdinak.

Eskuarteko corpusa osatzen duen hau ere bere garairi zordun da eta sasoiko bilketa eta ikerketa-moldeak erabili ziren. Bilketarako moldeak asko aurreratu eta ho-

<sup>3</sup> Xehetasunerako ikus Aurrekoetxea & Videgain (2004).

<sup>4</sup> Bihoakio hemendik beste behin ere gure esker ona Paul-Henri Allieux-i bildumako euskal testuez jabetzeko hurbildu ginenean hain modu abegikorrean hartu izanagatik.

betu diren arren, aspaldiko moldeek ere datu interesgarriak biltzen zituzten. Horren froga Bourciez bilduma da. Zalantzarik gabe, ikertzailea bilketako unean leku horretan ez izateak, edo ikertzaileak datuen bilketaren kontrol zorrotza ez izateak, dakartzan arrisku eta kalteak jasan zituen Bourciez-ek ere. Testuetan zehar bidaiatuz eta testuen ondoren agertzen diren oharrak irakurriz biltzaileen artean joera bat baino gehiago izan zela esan daiteke: batzuk, gehienak zorionez, zeuden herriko hizkerari modu hertsienean atxikiz jardun zuten; beste banaka batzuk, aldiz, arinago jokatu, ez zuten, agian, zorrotzasun handiegiz euren herriko hizkera bildu. Honen adibide, Pagolako testua dugu. Geroago argiago eta luzeago mintzatuko garen arren, begi-bistakoa da herri honetako testuak ez duela herriko hizkera fideltasunez biltzen, inguruko herrietakoekin erkatuz argi gelditzen dena, bestalde.

### 3. Testutik datu-basera

Gorago esan da jatorrizko datuak herriz herriko testuak direla. Testuok makinatu ostein<sup>5</sup> ustiaketa sistematizatua egin da lehen-lehenik, baliabide automatizatuak erabiliz. Testuak hustu eta testu-hitz bakoitza dagokion testuaren kodearekin batera baseratu da.

Ibilbide luze eta arazotsua gertatu da eta arazo desberdinei irtenbide egokiak edo egokienak bilatuz egin da aurrera. Arazoen artean hitz mozketek izan dira denbora aldetik zeregin ugarien sortu dutenak. Izan ere, makinatutako testuak eskuizkribuak dituen ezaugarri berak gordetzen ditu eta honetatik ahal denik eta hurbilen den moldean utzi dira. Honek, ondorioz, zera esan nahi du: ez direla huts tipografikoak zuzendu, adibidez, edo ez direla eskuizkribuko «oker» moztuta agertzen diren hitzak lotu edo «oker» banandutakoak bereizi, ez direla erratak zuzendu... Aldiz, testutik datu-basea eraikitzean jatorrizko ezaugarriak biltzen dituen taula bere horretan utzi bada ere, beste batean «huts» horiek «konpondu» egin dira, hala nola, hitzen ez-ohiko banaketak.

Horrela sortu da 52.902 erregistro dituen datu-basea. Beraz, datu-base honetan eskuizkribuetan (eta transkribatutako testuetan) agertzen diren testu-hitz guztiak biltzen dira. 52.902 erregistro horietarik 12.499 item desberdin lortu dira. «Seme xahutzailearen parabola» luzeraz txikia izan arren datu-base txukuna, interesgarria eta adierazgarria lortu dela esan behar da.

### 4. Lematizazioak

Ezein corpus ikertu eta ustiatzeko gaur egun ezinbestekoa da lematizatzea. Lematizatu gabeko corpusen erabilera zailaz gain, nekeza eta astuna, gertatzen da. Ikerketa honen azken helburua Bourciez bildumako euskal testuetatik abiatuz Ipar Euskal Herriko hizkeren azterketa geolinguistikoa egitea denez, testuak baseratu ondorengo lehen zeregina lematizatzea izan da. Bi lematizazio aurreikusi dira: lematizazio ortografikoa eta lematizazio semantikoa. Lehena hitzen formari edo adareei dagokie, bigarrena kontzeptuei.

<sup>5</sup> Testuon argitalpenak izan duen ibilbidea zehatzagorako ikus Aurrekoetxea & Videgain (2004).

#### 4.1. Lematizazio ortografikoa

Lematizaziorako programa informatiko laguntzailea sortzea erabaki dugu.<sup>6</sup> Ezaguna da euskararen industrian badirela antzeko zereginetan erabiltzen diren tramankulu edo tresnak, baina gure beharretakoa deserosoak izateaz gain, ez direnez egokiak gertatu berria egitera behartuta aurkitu gara.



1. irudia

#### Lematizaziorako programa informatikoaren pantaila

Oso erabilera sinplea duen programa-informatiko honek ikertzaileari lematizatu gabeko testu-hitzen zerrenda ematen dio (ikus 1. irudia: pantailako ezkerreko «lematizatu gabekoak» zutabea). Ikertzailea zerrendatik testu-hitzak aukeratuz joango da eta aukeratu ahala dagokien lema ezarriko die «lema honekin lotu» laukitxoan. Ondoren «lema» jartzeko laukitxoaren azpian dagoen botoia joz eskuineko «Testu-hitza» eta «lema» buruan dituen zutabeen agertuko dira (testu-hitza zutabearen ezkerretara eta dagokion lema eskuinetara). Ezkerreko zutabeko testu-hitzen zerrenda agortu arte ekin behar du ikertzaileak.

Ondoren, badu aukera lema baten barnean sartu diren testu-hitzak berrikusteko ere, eta hala iritiz gero, lema batetik askatu eta beste batean kokatzeko. Horretarako pantailaren erdi-erdian agertzen den zutabeen «lema-zerrenda» agertzen da. Nahikoa da zerrenda horretan agertzen den lema baten gainean sakatzea eskuinekoan lema horri dagozkion testu-hitzen zerrenda ager dadin. Kasu honetan «adio» lema 10 aldaera bildu zaizkio: *adio*, *adio[r]ic*, *adioic*, *adioic*, *adioik*, *adioric*, *adiorik*, *adiórik*, *adios* eta *adioyk*. Horrela, aldaera lexikalak eta gramatikalak lema berean bildu dira.

Zerrenda horretan agertzen ez bada aukera du «iragazi» botoiari dagokion laukitxoan nahi den lema idatzi eta botoiari eragitea.

<sup>6</sup> Programa honen egilea Iñaki Kareaga, Euskaltzaindiako informatika-zerbitzuen arduraduna da. Gure esker ona adierazten diogu lerro hauen bidez.

«Lematizazio ortografikoa» deitu izan den eragintzaren bidez lortu da corpuseko hiztegia (ikus Aurrekoetxea & Videgain 2004). Hiztegi honetan lema sarreraren ondoan lemaren maiztasunaz gain, lemaren aldaera guztiak biltzen dira, aldaera bakoi-tza zein testutan agertzen den azaltzen delarik.

Izatez, hiztegi hau sortzean, lexikografian hiztegi sarrerak aukeratzeko erabil-tzen den prozedura eta metodologia bera erabili da, hitzen aldaeren artean auke-ratzeko orduan. Tresna honen bidez aldaera desberdinak bildu eta euren artean bat hautatu edo bat sortuz joan dira aldaera guztiak lemekin lotzen. Modu hone-tan hasierako 12.499 itemetatik 1.119 lema desberdin edo hiztegi sarrera lortu dira.

#### 4.2. Lematizazio semantikoa

Horrela lortu da argitaratu den corpus osoaren hiztegia. Baina hiztegi hau for-men hiztegia da eta ez adierena. Horretarako bigarren lematizazio bat sortu da: «le-matizazio semantikoa» deitu izan dena. Kasu honetan «seme xahutzailearen para-bola»ren jatorrizko frantses bertsioetik abiatu gara. Testu horretan agertzen diren kontzeptuen bilketa egin da lehenik. Ondoren kontzeptu bakoitzari dagozkion eus-kal ordainak bildu zaizkio.

Lematizazio semantikoaren bidez testuan agertzen diren 151 kontzeptu desber-dinetan bildu dira lematizazio ortografikoan lortu diren lemak. Ez dira, haatik, kontzeptu guztiak landu, erabilpen geolinguistikorako argiak eta baliagarriak dire-nak baino ez dira landu.

Lematizazio hau sortzean agertu diren arazoak ere asko eta larriak izan dira, kasu batzuetan. Noiz edo behin kontzeptu baten pean sartzeko zailtasun handiak izan di-tugu eta eztabaidagarritzat, eta are kritikagarritzat ere, jo daitezkeen erabakiak ere hartu direla onartu behar dugu. Esaten ari garenaren adibide argi bat ondokoa da: ‘tomber’ kontzepturako ondoko lemak agertzen dira testuetan: *erori, jarri, etzatu, pa-ratu, iraili* eta *aurtiki*. Honelako beste bat ‘traiter’ kontzeptuari dagokiona da. Lema horietan bildu dira egitura sintaktiko konkrituak erabiliz itzulitako pasarteak, itzul-tzaileak jatorrizko testuaren gainean izan duen sentiberatasun maila desberdinak, tes-tuarekiko fideltasun maila desberdinak... Hala ere, hauek bezalako eztabaidagarri izan litekeen kasu bat edo beste gorabehera, oso emaitza argi eta erabilgarriak lortu dira.

Lematizazio honen emaitzak erabili dira kontzeptukako mapak eraikitzeke eta hizkeren arteko sailkapena egiteko. Hau da mapa guztien eta azterketa sailkatzaile guztien abiapuntua. Hori dela eta, azpimarratu gura dugu urrats honen garrantzia, ondorengo ikerketa guztia bere gainean eraikiko delako. Eta garrantzitsua dela diogu, arazo biren aurrean jokatu behar izan delako: batetik, lexikografian erabil-tzen diren irizpideak erabili behar izan direlako; eta bestetik, kontzeptu bat itzultze-ko testuetan zehar agertzen diren hitz eta egiturez gain beste daturik ez delako izan. Halere, honek ez gaitu harritu behar; izan ere, behin baino gehiagotan aurkitu izan delako egoera berean bai dialektologoa, bai beste diziplinetako ikertzailea. Honekin esan gura izan dudana zera da: ez dela izan inoiz gertatu ez den edo arras gutxitan eman den egoera baten aurrean; oso ohikoa ez bada ere, ezaguna den egoera baten aurrean, alegia.

Ondorioz, datu-base desberdinak sortu dira:

- a) jatorrizko datuak biltzen dituen lematizazio ortografikoaren bidez eratu da. Datu-base honetan testuetako hitz eta unitate gramatikal guztiak biltzen dira.
- b) Datu-base lexikala «lematizazio semantikoa» deitu den prozedura jasandakoa, zeinen bidez testuen azterketa lexikalak egiteko prest utzi den.
- c) Datu-base gramatikala: ezaugarri morfologikoak, sintaktikoak eta fonologikoak bilduko dituen: oraindik eraiki gabea.

Testuen ustiaketa dialektologikoa datu-base hauen eraikitetik hasten da. Testuetan biltzen den informazio guztia eratu da eta ikertzailearen beharrei erantzuteko modu egokian bildu da.

Datu-base lexikala dugu ikerketa honetan aztergai. Testuetan biltzen diren kontzeptuen analisia egitean 227 kontzeptu desberdin bereizi genituen hasiera batean datu-basea eratzeko. Azterketa sakondu ahala konturatu gara horietako batzuk ez direla testu guztietan agertzen. Jatorrizko testu bera izan arren ez dira itzulpen guztietan jatorrizkoan agertzen diren kontzeptu guztiak agertzen. Beste batzuk kontzeptu nahasiak dira edo baliabide gramatikal desberdinez itzuli izan dira; hauek ere baztertu egin dira. Testu gutxitan baino agertzen ez ziren kontzeptuak ezabatuz eta kontzeptu nahasiak albo batera utziz, 151 kontzeptu gelditu dira.

Testuei dagokienez ere murrizketa bat egin da: bi testu desberdin bildu diren herrietan bakarra aukeratu da. Hori dela eta, bilduman agertzen diren 150 testuak 137tan utzi dira. Hauek dira bi testu dituzten herriak: Donibane-Lohizune, Uzta-ritze, Ezpeleta, Maule, Donapaleu, Arberatze-Zilhekoa, Behauze, Iholdi, Izura, Baigorri, Donibane-Garazi eta Atharratze. Erabaki hau datuen tratamendu geolinguistikorako erabili den programa informatikoak beharturik hartu da. Programa honek ez du herri berean eta kontzeptu berean erantzun bat baino gehiago ezar-tzea onartzen.

Datu-base lexikalean, beraz, 151 kontzeptu eta 137 herri agertzen dira. Honek guztira 20.838 item edo taxat, H. Goebl-en nomenklatura erabiliz, egiten ditu. Item edo taxat hauek izango dira lexikoari dagokion Ipar Euskal Herriko hizkeren sailkapenaren oinarri.

Garbi utzi gura dugu inor ez dadin desbideratu eta ondorio okerrak ez ditzan atera: 151 kontzeptu edo ezaugarri inoiz Euskal Herrian euskalkiak edo hizkerak sailkatzeko erabili den kopuru handiena da. Gainera, kasu honetan, Ipar Euskal Herriko hizkerak bakarrik sailkatzeko eta ez Euskal Herri osokoak. Bestalde, aztertzen den eremuko ia hizkera guztietako datuak daude. Eta hala ere, diogu, ikerketa honetan lehen aldiz argia ikusten duten Ipar Euskal Herriko hizkeren sailkapenaren mapak kontu handiz aztertu behar direla eta ez direla inondik inora ere behin-betikotzat jo behar: Bourciez-ek bildutako euskal testuetan agertzen diren ezaugarriak kontuan hartuz egindako hizkeren sailkapena da, eta hori baino ez. Esaten ari gara ondokoa da: ikerketa hasi besterik ez dela egin, ikerketa hau bideratu duen taldeak gaiarekin jarraitzeko erabaki irmoa duela, hemen lortu den sailkapena beste datu batzuekin erkatu behar dela, ikertzaileen eztabaida sustatu behar dela, eta metodologian eta datuetan iritzi bateratuak lortzen direnean eman daitezkeela, eta orduan bakarrik, emaitza fidagarriagoak.

Testuok, izatez, gugan baino lehenago piztu zuten interesa J. Allièresgan eta goangarria bezain ahantzia den «Petit Atlas linguistique basque français Bourciez 1, (20 cartes commentées)» 25 urte dituen lanean emaitza adierazgarriak argitaratu zituen. Lan honekin argi utzi zuen Jacques-ek testuok dialektologia egiteko baliagarriak ez eze interesgarriak ere bazirela. Haren bidea berribili dugu, ikasle zintzoen legez, baina oinetako eta zakuto erosoagoekin. Irakaslea beti gogoan, hark erakutsitako bidetik urrunago joateko esperantzaz hasitako bidea dugu, denborak esanen!

Guk proposatzen dugun ustiaketa geolinguistikoak ildo desberdinak lantzen ditu:

- ezaugarri batzuen ohiko ikusmoldeko ikerketak;
- ustiaketa geolinguistikoa Allières-ek hasi bidetik;
- ustiaketa dialektometrikoa, Hans Goebel-en moldeak jarraituz.

Artikulu honetan bigarren eta hirugarren bideak jorratzera abiatzen gara, ikerketa honekin testuon ustiaketa hasi baino ez baitugu egiten.

## 5. Datu-basetik similaritate eta distantzia matrizeetara

Gorago aipatu den datu-basea hizkuntza datuen basea da, hots, kualitatiboa. Era honetako datu-baseak deserosoak dira prozedura estatistiko-sailkatzaile automatizatuak egiteko. Askoz erosoagoak gertatzen dira datu kuantitatiboak dituzten baseak.

Ikerketa honetan hizkuntza jatorrizko datu-basea, hau da, datu linguistikoak dituen datu-basea, datu numerikoez osatuta dagoen datu-base batean eraldatu dugu. Horretarako egin diren urratsak ondokoak izan dira:

- a) Kontzeptu bakoitzeko taxat edo lema bakoitzari zenbaki bat esleitu zaio.
- b) Taxat edo lema jakin bat agertzen den herrian dagokion zenbakiarekin ordeztu da.

Prozedura estatistiko-sailkatzaileetarako H. Goebel-ek Salzburgoko Unibertsitatean garatu duen VDM programa erabili da. Erabaki honek aurrekariak batetik, eta zergatiak ere baditu, besterik.

Aurrekarien artean bat aipatuko dut: nire ikerketa ildo nagusietako bat hizkeren sailkapen eta egituraketa geolinguistikoaren kezka da. Egituraketa geolinguistikoaren ikerketa aurrera eramateko hasiera-hasieratik baztertu nuen eskuz eta buruz aritzea, jada 1970ean Jean Séguy-k, Okzitaniako Tolosako irakasle ezagun eta dialektometria hitzaren sortzaile eta diziplinaren aitzindariak, esanak aintzat hartuz. Diziplina honen jarraitzaileek eta oro har hizkeren sailkapenean aritzen direnek prozedura estatistiko-automatizatuak erabiltzean oinarrituz egin da hautua. Horrela, duela 14 urte ekin nion ildo honi Nafarroako hizkeren egituraketa geolinguistikoa aztertzen hasi nintzenean, ohiko bizkaieraren hizkeren azterketari ekin nion gero eta harrezkero beste ikerketa batzuk<sup>7</sup> ere garatu izan ditut.

<sup>7</sup> Eskerrak eman behar dizkiet, beste behin, Xabier Isasi eta Eduardo Oleari sasoi hartan era bateko zein besteko lanabesak erabiliz lagundu nindutelako.

Konbentziturik nago hizkuntza geografiak aurrera jo behar bada eta bere metodologia modernizatu edo eguneratu behar bada, atomismotik irten behar duela. Hizkuntza atlas gehienak hizkeren museo hilak gertatu dira. Hizkeretan gorde den aberastasun ikaragarria bildu da (modu eredugarri eta zehatzean gehienetan gainera), baina lo datza. Hizkuntza geografian dihardugunok behartuta gaude hizkuntza aldagarritasunaren teoriarantz ematera, aspektu txiki bat bakarrik kontuan hartuta bada ere. Baina horretarako hizkuntza atlasetan biltzen den ikaragarriko hizkuntza materialaren analisi sakona egitera bideratu behar ditugu gure ikerketak.

Ikerketa horien norabide bat hizkuntza atlasetan biltzen den aberastasunen sintesiak egitea dela esango nuke. Dialektometriaren hasiera-hasierako lanetan, J. Séguy eta batez ere H. Goebel-enetan, ondo demostratuta gelditu da gainerako zientzietan (giza-zientzia nahiz bestelakoetan) prozedura sailkatzaile eta sintetizatzaileak erabiltzen dituztela. Hizkuntzalaritzan eta dialektologian ere zilegitasun osoa dugu gainerako zientzietan ondo esperimentatuta dauden tresnak erabiltzeko.

Nire ildo honetako orain arteko ikerketetan, nolabait esateko, tresna moldagaitzak erabili izan ditut eta hizkuntzalaritzaz kanpoko adituen laguntza ezinbestekoa izan dut. Tresna informatiko desberdinak eta ez hizkuntzalaritzari ezta hizkuntzalariei egokituak. Halakoetan, maiz, denbora gehiegi iraungitzen da tresna ezagutzen eta behar bezalako trebetasunez erabiltzeko doitzen. Bestalde, hizkuntzalaria ez da bere lanean autonomo bihurtzen eta inoren laguntzaren zain gelditzen da.

Ikerketa-proiektu honen helburuetan dialektologoak hizkeren sailkapen geolinguistikoa ikertzeko tresneria informatikoa lortzea egon da, behin betikoz, ikerketaren urrats guztien jabe izan eta modu autonomoan ikertzeko. Alde horretatik Hans Goebel-ek garatu duen VDM programa informatikoa oso indartsua eta erabilterreza da: batetik, tresna estatistiko-sailkatzaileen abantaila guztiak biltzen ditu, bestetik kartografia automatizatua egiteko prestatuta dago, azkenik hizkuntzalaritzari eta hizkuntzalariei egokitua agertzen da. Eta dena tresna bakar batean eskaintzen du.

Egia esan behar bada, tresna erabilterreza izan arren oinarriko kontzeptuak ondo menperatu behar direla etekin duina atera nahi bada behintzat, formazio ona behar da izan: estatistikaren oinarri sendoa behar du ikertzaileak programak erabiltzen dituen kontzeptu asko ulertzeko eta behar bezalako etekina lortu ahal izateko.

Hans Goebel-ek, zeinen metodologia eta programa informatikoa estuki jarraitzen dugun ikerketa honetan,<sup>8</sup> sarri aipatu eta azaldu du similaritate edo distantzia matrizeen funtsa (ikus besteak beste oso eskuragarria den Goebel 1992). Batetik beste-rako bidea «similaritate-indize» bat aplikatuz egiten da. Taxonomia berriak honelako indize ugari eskaintzen ditu Goebel-ek behin baino gehiagotan aipatu duen

---

<sup>8</sup> Lerro hauen bidez agirian utzi gura dugu H. Goebel-enganako gure esker ona, gurekiko izan duen abegikortasun, prestutasun eta eskuzabaltasunagatik. Eskerrak berari eta bere laguntzaile guztiei eta bereziki Sławomir Sobota-ri eroapen handiz lagundu digulako, mapa digitalizatze-lanetan, gure beharretara programa informatikoa egokitzen eta beharrezko azalpenak ematen. Salzburgo egonaldia, maila pertsonalean ezin-ahantztekoa izateaz gain, euskal ikasketetarako aurrerapausu handia suposatuko duela itxaropena dugu.



legez (ikus adibidez, Sneath & Sokal 1973; Chandon & Pinson 1981; Bailey 1994 edo Gordon 1999). Dialektologoak bere beharrei eta helburuei egokien datorkien indizea aukeratu behar du. Guk, Goebl-i jarraituz, «Identitatearen indize erlatiboa-IRI» (edo RIW alemanez) eta «Identitatearen indize ponderatua (pisu batekin)-IPI» (edo GIW alemanez) erabiliko ditugu. Izan ere, bi indize hauek aspaldidanik erabiltzen ditu Goebl-ek eta euren erabilgarritasun taxometrikoa guztiz frogatuta utzi du dagoneko.

Bi indizeen arteko desberdintasuna zertan datzan ikusteko Goebl-en hitzak gure egingo ditugu (2003: 69):

Alors que l'IRI<sub>jk</sub> est un indice isocratique (ou <adansonien>, baptisé d'après le naturaliste français Michel Adanson, 1727-1806) et traite ainsi tous les taxats de la matrice de données numériquement sur un pied d'égalité, l'IPI(1)<sub>jk</sub> confère, comme indice anisocratique, un poids numérique plus grand aux taxats géographiquement moins répandus tout en défavorisant les taxats de grande diffusion spatiale. L'action pondératrice de cet indice traduit en termes taxométriques une opinion très habituelle parmi les linguistes selon laquelle les traits linguistiques universellement répandus disposeraient d'une importance classificatoire moindre que les traits linguistiques moins fréquents auxquels on attribue souvent la prérogative d'être ou bien des archaïsmes ou bien des innovations dignes d'un intérêt majeur.

Hots, IRI indizeak taxat edo erantzun guztiak berdin tratatzen dituen neurrian, eragiketa estatistiko-sailkatzaileak egitean denei pisu edo garrantzi bera emanez, IPI indizeak hedadura geografiko urria duten taxatei hedadura handia dutenei baino garrantzi handiagoa ematen die. Beste hitz batzuetan, IPI indizeak bazterreko edo eremu txikiko ezaugarriei, gehienetan arkaismoak edo berrikuntzak, gainerako ezaugarriei baino pisu gehiago ematen die estatistikak aplikatzean eta hizkeren sailkapena egitean. Horrela detektatzen ditu sarri ostendurik gelditzen diren eta hizkeren ezaugarrietan, ikuspuntu batetik behintzat, garrantzitsuak izan daitezkeen ezaugarriak. Bi indizeak erabiliko ditugu guk ere.

Similaritate indizearekin batera distantzia (disimilaritate edo desberdintasun indizea ere erabiltzen da) indizea IRD<sub>jk</sub> (distantziaren indize erlatiboa) eta IPD(1)<sub>jk</sub> (distantziaren indize ponderatua) ere erabil daitezke ondoko formula aplikatuz: IRD<sub>jk</sub> = 100 - IRI<sub>jk</sub>, indize erlatiboa denean, edo IPD(1)<sub>jk</sub> = 100 - IPI(1)<sub>jk</sub> indize ponderatua denean. Era honetan hizkeren arteko hizkuntz distantziak izango dira neurtuko direnak. Bi indize hauek, similaritatea eta distantzia, hizkeraren funtzio komunikatiboa eta funtzio bereizlea modu matematiko edo zehatz batean neurtzeko indizeak dira. Dialektologia «teorikoan» maiz erabili izan diren kontzeptuak izanik, garrantzitsua da eremu edo hizkuntza bateko hizkeren, bi funtzio hauek go-goan izanik, azterketa zehatzak egin ahal izatea.

## 6. Multzokatze algoritmoak

Multzokatze algoritmoa similaritate matrizeak erakusten duen hizkeren arteko desberdintasuna mailakatzeko edo multzokatzeko tramankulua da. Zertarako behar diren similaritate matrizeak mailakatu eta multzokatzea? Bada, hizkuntza distan-

tziak nola edo hala mapan islatzeko, mapan ikustarazteko moduan ipintzeko. Hizkuntza similaritate edo distantzia desberdinak banan-banan mapa batean ikustaraztea ezinezkoa denez, datu horiek multzokatu behar dira eta orduan bakarrik egiten da posible mapa batean erakustea.








Hori da egiten dutena multzokatze, edo gura bada tartekatze, algoritmoek. Hizkeren arteko similaritateak multzotan ezarri.

Estatistikan horretarako algoritmo bat baino gehiago ezagutzen da. Ikerketa honetan bi erabiliko dira: MINMWMAX eta MEDMW algoritmoak, hain zuzen ere. Ezagunenak izateaz gain, dialektologian erabilienak dira eta gure helburuetarako ere egoki-egokiak.

### 6.1. MINMWMAX algoritmoa (6 multzotan)

Algoritmo honen azalpen argia ematen du H. Goebel-ek (1981: 365 eta h.; 1987a: 79 eta h.) eta hemen laburbilduko dugu. Abiapuntua hizkera bat edo herri bat izaten da. Eta herri horretatik abiatuz gainerako herriekiko hizkuntza harremanak, hots, distantziak neurtzen dira, jokoan datu-baseko ezaugarri guztiak kontuan izanik:

- Maiztasunen distribuzioen media aritmetikoa kalkulatzeko da lehenik.
- Media aritmetikotik maiztasun txikienera dagoen tartea 3 zenbakiarekin zatitu; lortzen den zenbakia izango da multzo bakoitzaren luzera. Balore txikienetik abiatuz eta emaitza hori gehituz lortuko dira media aritmetikotik peko 3 multzoak.
- Prozesu bera jarraituko da media aritmetikotik goranzko tartean ere: maiztasun handiena eta media aritmetikoaren arteko distantzia 3 zenbakiarekin zatitu eta lortzen den zenbakia mediari gehituz joanez ateratzen dira mediaz goitiko 3 tarreak.

	[1] 61.19-tik 64.99-ra (n = 16)
	[2] 68.79-ra (n = 25)
	[3] 72.59-ra (n = 28)
	[4] 76.80-ra (n = 33)
	[5] 81.02-ra (n = 22)
	[6] 85.23-ra (n = 13)
	hizkeren gehiketa: 137

2. irudia

MINMWMAX algoritmoaren adibide bat








Kasu honetan media aritmetikoa 72,59 da eta maiztasun txikiena 61,19. Bien arteko desberdintasuna (11,4) 3 zenbakiarekin zatituz lortzen den zenbakia (3,8) erabiliko dugu 3 multzoak egiteko:  $61,19 + 3,8 = 64,99$  (lehen multzoa) +  $3,8 = 68,79$  (bigarren multzoa) +  $3,8 = 72,59$  (hirugarren multzoa).

Media aritmetikotik gorantz ere antzera jokatuko da: mediatik maiztasun handienerako tartea lortu lehenik (12,64), 3 zenbakiarekin zatitu (4,21) eta ondoren media aritmetikoari gehitu zenbaki hau eta lortzen dena (76,80) izango da 4. multzoaren muga, honi berriro gehitu 4,21 eta lortzen den zenbakia (81,02) izango da 5. multzoaren muga, eta, azkenik, 6. multzoa zenbaki honetatik maiztasun handienera doan tartearak beteko du.

## 6.2. MEDMW algoritmoa (6 multzotan)

Algoritmo hau ez da aurrekoa bezain «neutroa»; algoritmo honek datuak bortsatu egiten ditu: erdiko baloreak urritu eta muturrekoak indartu egiten ditu (Goebel 1987: 81-82). Algoritmo honen prozedura ondokoa da:

- a) Maiztasunen distribuzioko baloreak txikienetik handienera ordenatzen dira.
- b) Media aritmetikotik hurbilen den balorea hautatzen da.
- c) Txikienetik mediara hurbilen den balorerako tartean kokatzen direnak multzo batean biltzen dira.
- d) Gauza bera egiten da media aritmetikotik handienera kokatzen direnekin.
- e) Bi multzo horietatik bakoitza hiru multzotan zatitu behar da, sortzen den multzo bakoitzak maiztasun distribuzio kopuru bera edo bertsua duelarik.

	[1] 61.19-tik 66.44-ra (n = 23)
	[2] 69.80-ra (n = 23)
	[3] 72.41-ra (n = 22)
	[4] 75.17-ra (n = 24)
	[5] 77.85-ra (n = 23)
	[6] 85.23-ra (n = 22)
	hizkeren gehiketa: 137

3. irudia  
MedMw algoritmoaren irudia

Minmymax algoritmoan erabili diren datuak erabili dira algoritmo hau azaltzeko ere. Ikus daitekeenez algoritmo bietan balio txikien eta handienak berak dira; desberdintasuna balore horiek multzokatzean datza. Eta kasu honetan irizpide nagusiena hizkerak kopuru berdintsuetan multzokatzea da: 22 hizkera biltzen dira 3. eta 6. multzoetan, 23 hizkera 1., 2. eta 5. multzoetan, eta 24 hizkera 4. multzoan.

## 7. Histogramak

VDM programa informatikoaren ikusarazpen tresneriaren artean (mapa eta algoritmoez gain) baloreen multzo bakoitzaren pisua agertzen duten histogramak dira.

Mapa koropleten osagarri dira histogramak eta maiztasunaren distribuzioaren ezaugarri estatistikoaren ikusarazpena argi erakusteko balio dute. Histogramaren eraikitzea algoritmo zehatz bati dagokio.

Histograma hauetan multzokatze algoritmoan egin diren multzo desberdinen pisua aztertzen da, batetik: multzo batek zenbat eta hizkera gehiago bildu orduan eta histograman garaiera handiagoa izango du. Irakurlea ohartu behar da multzokatze algoritmoa multzo bakoitza bi zutabetan bananduta agertzen dela histograman (eta alde horretatik informazio gehiago eskaintzen duela).

Bestetik, mediana aritmetiko eta histogramako maiztasunen distribuziotik kalkulatuak desbiazio-estandarra (fr. *écart-type*, ing. *Standard deviation*, al. *Standardabweichung*) kurba normal teorikoaren proiektzioak hainbat zehaztasun estatistiko kontrolatzeko aukera ematen du: kurba orekatua bada (eskuinean eta ezkerrean hizkera kopuru bera bada) kurba simetrikoa dela esaten da. Honek abiapuntuak hizkeraren ezaugarri nagusi bat agirian uzten du: hizkera horrek hizkuntza harreman maila orekatua du; edo beste era batera esanda, hizkera horrek hizkuntza harreman onak dituen bezainbeste hizkera dituela hizkuntza harreman txarrekin. Aldiz, histogramako kurba eskuinetara asimetrikoa bada, hizkera horrek hizkuntza harreman txarrekin hizkera gehiago izango ditu onekin baino; eta alderantziz ezkerretara asimetrikoa bada. Azkenik, histogramaren zutabe bakoitzak biltzen duen hizkera kopuruaren zenbakia ere agertzen du (ikus adibidez, 5. mapa). Mapa honetan 1209 kodea duen hizkerak hizkuntza harreman onak izango ditu 70 hizkerekin (13rekin bikainak, 21rekin oso onak eta 36rekin onak) eta harreman txarrak 67 hizkerekin (18rekin ikaragarri txarrak, 26rekin oso txarrak eta 23rekin txarrak).

## 8. Kartografia eta mapa-zola

Testuetan agertzen diren hizkuntza datuak analisi estatistiko-sailkatzaileak jasan ondoren mapetan isurtzeko Delaunay-ren triangulazioaren gainean eraikitako Voronoi-ren poligonazioa erabili da. Poligonazio hau ez da berria hizkuntza geografian, H. Goebl-ek azken 30 urteotan ildo honi buruzko ikerketa guztietan erabili du eta haren zehaztasunak ere behin baino gehiagotan agertu ditu. Berak maiz aipatzen duen bezala, jada 1898 Carl Haag germanologoa gomendatzen omen zuen poligonazio mota hau hizkuntza geografiarako (zehaztasun gehiagotarako ikus Goebl 1981: 363-364; 1987b: 93-99; 2003: 72-73).

Poligono hauen bidez herri edo hizkera bati dagozkion ezaugarriak (hemen kolorearen bidez adierazten direnak) ezartzeko aukera ematen da.

Kasu honetan Bourciez bilduman euskarazko testua duten Ipar Euskal Herriko herri guztiak bildu dira mapan eta euri dagokien poligonoa sortu da, tresna automatikoak erabiliz Salzburg-eko Unibertsitateko Filologia Erromanikoaren Sailean. Poligonoak kodetzeko testuek duten kode bera erabili da; horrela, bietan erreferentzia bera izanik, irakurleak testutik maparako jauzia zailtasunik gabe egin lezake.

Irakurlea erraz ohart daitekeen legez, bilduman agertzen diren herriak baino ez dira agertzen mapan. Hots, euskal testuak jaso diren herriak baino ez dira azaltzen. Gainerako herriak (edo testu gaskoia bildu direnak edo testurik bildu ez diren herriak) ez dute mapan lekurik. Hori dela eta, beti izan behar da gogoan ikerketa honetan egiten diren azterketetan herri horietako datuak falta direla.

Ikerketa honetan agertzen ditugun mapak konbentzio estatistiko eta kartografiko zehatzak jarraitzen dituzte. Prozedura estatistikoetan lortu den aldagarritasun numerikoa mapetan ezarri ahal izateko uztarkuaren espektroaren logika erabiliz ordenatu dira koloreak. Ikertzaileak beti du eskura, hala guzaz gero, koloreak zein kolore kopurua (mapa sintetikoetan) aldatzeko aukera.

Bestalde, geolinguistikan ekidin ezin diren erantzun gabeko puntuak aurkituko dira mapetan zehar. Daturik gabeko puntu asko izateak emaitzak distorsionatzeko arriskua izaten duenez, ahal delarik horiek saihesten ahalegindu behar da. Eskuar-tean darabilgun corpusean daturik gabeko puntuek %1,14 baino ez dute egiten (20.838 datutatik 238 dira hutsik daudenak). Oso gutxi dira benetan eta eragin dezaketen distortsioa edo desbideraketak ez du ondorioz emaitzetan. Hala ere, hutsune bat behintzat duten mapa kopurua handi xamarra da (97 mapa). Datuetara joz, hutsune bat behintzat duten mapen hain kopuru handian herri bakar batek duen eragina oso altua dela agertzen da; izan ere, Hendaiaiko testua da kopuru hauen eragile handiena (83 kasutan hutsik agertzen da herri hau; %34,87tan, hain zuzen ere). Herri honi dagokion testuko datuak ezaba zitezkeen eta azterketatik kanpo utzi, eta agian hau izan zitezkeen erabaki zentzuzkoena, baina uztea erabaki da bildumako euskal testu guztiak integratuz gure azterketan, geroago batean eta beste ikerketa batzuetarako baliagarriak izango diren itzaropenarekin.

## 9. Kontzeptukako mapak

VDM programaren lehenengo emaitza kartografikoak kontzeptukako mapak dira. Kontzeptu bakoitzeko mapa bat eraikitzen da; geolinguistikan ohikoa bihurtu den metodologia erabiliz lortzen dira mapa hauek (kontzeptu bakoitzari mapa bat dagokio). Horretarako, lexikoaz dihardugularik, erantzunak lematizatu egin dira lehenik eta ondoren lema bakoitzari kolore bat ipini zaio maparen leiendan agertzen den moduan. Lemak eta koloreak leiendan uztartzen dira (ikus 1. mapa).<sup>9</sup>

Lau erantzun bildu dira mapa honetan: *diru*, *diharu*, *sos* eta *ondasun*. Lehen biek eremu handia hartzen dute: *diru* mendebaldean eta *diharu* ekialdera. *Sos* erantzunak 7 herri biltzen ditu eta bat izan ezik gainerakoak erdialdean kokatzen dira. *On-dasun*, azkenik, herri bakarrean bildu da: Hendaia.

Testuen ustiaketa hau programak eskaintzen duen oinarritzkoena da eta mapa globalizatzaile eta sintetikoaren oinarria da aldi berean. Hauetan kontzeptu bakoitzaren inguruan bildu diren hitzen zerrenda eta hitz bakoitzaren hedadura eta kokagunea agertzen dira. Gorago esan legez, hitz bakoitzaren kolorea aldagarria da eta iker-tzaileak bere moldean erabili ahal ditu.

Bourciez-testuetan agertzen diren 151 kontzeptu landu dira datu-base lexikala sortzeko. 151 mapetatik gainbegirada bat emanez datu deigarri bat agertzen da: erantzun bakarreko mapa edo mapa mononimoen ugaritasuna: 54 mapa mononimo agertzen dira; hau da, mapa guztietarik hirutik bat erantzun bakarrekoa da.

<sup>9</sup> Mapetan herriak zenbakien bidez adierazten dira; hain zuen ere E. Bourciezek «Recueil des Idioms de la Région Gascogne» bilduman eman zizkien zenbakiez. Herriak eta dagozkien zenbakiak eranskinean azaltzen dira. Lexikoari dagozkion mapak kontsultatzeko ikus Aurrekoetxea, Iglesias eta Videgain (2007).

Ipar Euskal Herriko errealitate linguistikoaren egituraketa aztertzen hasteko (datu osoagoen aiduru) datu esanguratsuak agertzen dira.

Mapa honetan 'argent' kontzeptua adierazteko *diru*, *diharu*, *ondasun* eta *sos* lema agertzen dira. Maparen leiendan eta lemaren ondoan lema bakoitzaren maiztasun-kopurua agertzen da: agerrera gehien duen lema *diharu* da, 76 herritan bildu delarik; ondoren *diru* lema da, 54 aldiz. Lematizazio honen arrazoen artean lexikografikoak bigarren mailan gelditzen dira; izan ere, *diru* eta *diharu* lemen eremu geografikoa neurtzeko egindakoa baita. Lehen maparen ezker aldean agertzen da (ohiko lapurtera eta mendebaldeko behe-nafarreraren lurretan gehien bat), bigarrena maparen eskuinaldean (zuberera eta ekialdeko behe-nafarrerarenetan nagusiki). *Sos* lema maila apalago batean erabiltzen da, testu hauei kasu egiten bazaie eta kasu bat izan ezik beste guztietan ekialdeko behe-nafarreraren lurretan agertzen da.

## 10. Hizkeren sailkapena

Asko hitz eta idazten ari da azkenaldion Euskal Herrian diziplina honetan hizkeren sailkapenaz. L. L. Bonaparteren sailkapen eta mapa ez aski, itxuraz, eta iker-tzaile andana batek bere ekarria egin du dagoeneko. Ekarriari dagokionez era desberdinetakoak argitaratu dira: batzuk puntualak dira, beste batzuk orokorrak. Erabilitako metodologiaren aldetik ere badira desberdintasunak: batzuk ohiko dialektologian erabili izan dena baliatu duten bitartean, beste batzuetan metodologia berrizaleagoak erabili dira.

Bataren eta bestearen merituez eta ontasunez hitz egitera jo baino nahiago dugu datu berri batzuekin metodologia berri bat aurkeztea, eztabaida orokorra geroago baterako utziz.

Hemen aurkezten den metodologia gure artean berria bada ere, bere oinarriak ez dira hain berriak izatez. Azken 30 urteotan garatuz joan den metodologia da eta esan daiteke nahiko umotuta eta onartuta dagoela diziplina honetako komunitate zientifikoaren baitan. Bada, hala ere, gure artean oraindik zalantzatan jartzen duenik. Baina esan nahi nuena ondokoa da: eskaintzen dituen emaitzak sendoak direla, hizkuntza askotan frogatuta dagoen legez, eta oro har, molde zaharreko dialektologo batzuk salbu, gainerakoek ontzat ematen dutena dela. Eta hau diot, jakinik, gainerako edozein metodologia bezala, hau ere osatu eta hobetu daitekeela.

## 11. Hizkeren arteko hizkuntz distantzia: DM interpuntuala

Dialektometria interpuntuala edo puntuarteko dialektometria azter eremuko bi herriren arteko hizkuntza distantzia neurtzeko baliatzen den diziplinaren adar bat da. Dialektometria mota honetan bi norabidetan azter daiteke: hizkuntza similitatearen arabera edota hizkuntza distantziaren arabera. Batean hizkeren arteko hizkuntza antzekotasuna neurtzen den bitartean bestean desberdintasuna neurtuko da.

Prozedura globalizatzaile edo sintetikoetan sartzeko gara analisi hauekin. Bakar-kako datuak (kontzeptukako mapak) albo batera utzi eta mapa guztietan agertzen diren datuak batera erabiliz egingo dira hemendik aurrerako azterketak. Ez kontzeptu baterako jaso diren hitzak konparatuz edo aztertuz, landu diren kontzeptu

guztietako emaitzak datu-base batean ezarri eta denak kontuan hartuz egindako prozedura estatistiko-sailkatzaileak erabiliz lortutako mapak dira.

2. eta 3. mapak prozedura globalizatzaileetan arinenak eta konplexotasun gutxien agertzen dutenak ditugu. Tratamendu estatistikoa ere urria da: aurretik aztertutako 151 mapetako datuetan agertzen diren berdintasun eta desberdintasunak, denak batera hartuta, konparatuko dira herriak binaka hartuz.

### 11.1. Elkarren alboko herrien arteko hizkuntz distantzia (2. eta 3. mapak)

Abiapuntu gisa esan behar da geolinguistikan klasikotzat jo den «isoglosa» kontzeptua<sup>10</sup> ez dela ohiko adieran erabiltzen hemen. Erabilera horretatik haratago, baina erabilera hori gaitzesteko inolako asmorik gabe, «isoglosa sintetiko» edo globalizatzailearen kontzeptua dakargu hona.<sup>11</sup> «Isoglosa sintetikoa» jada ez da kontzeptu bati dagokion hitz baten muga marrazten duen lerroa, ezpada ze kontzeptu multzo bati emandako erantzun guztiak bilduz edo ezaugarri gramatikal multzo baten inguruan jasotako datuak bilduz, bi hizkeren arteko desberdintasun graduala adierazten duen marra. Bidenabar, «isoglosa sintetikoa»n muga adierazteko erabiltzen den trikimailua ez da jada ohiko isoglosa adierazteko erabiltzen den marra ondulatu etengabea, mapa eraikitzean burutu diren poligonoen alboek osatzen dutena baino; eta ondorioz, segmentu zuzen eta laburra, poligonoaren hurrengo alboarekin lot daitekeena edo ez.

Dialektometria interpuntualean elkarren alboko herriak bakarrik konparatzen dira. Bigarren mapan agertzen da garden esaten ari garena: mapa hau GIW similaritate indizea erabiliz burutu da. Multzokatze edo tartekatze algoritmoa MINMW-MAX-4 da; hots, multzotan bananduta agertzen dira datuak eta ondorioz 4 kolore agertzen dira mapan. Kolore «bero»ei (gorri eta okrea) dagozkien datuak txikienetik (= 2,65) media aritmetikora (= 60,64) bitartean kokatzen dira. Kolore «hotzak» (urdin argi eta iluna) media aritmetikotik (= 60,64) handienera (= 86,55) doan tartean kokatzen dira. Mapa hori irakurtzean hizkera bakoitzak bere inguruan dituenekin duen hizkuntza harremana bakarrik lantzen dela izan behar da kontuan. Kolore gorri eta okreak agertzen diren eremuan hizkeren arteko hizkuntza harremana handia da, euren arteko similaritate edo berdintasun maila jaso da: Zuberoan eta Lapurdin ematen da era horretako hizkuntza harremana. Nafarroa Behereko lurretan hizkuntza harreman handia duten hizkerak badauden arren hortxe biltzen dira harreman okerrak edo txarrak dituzten gehienak ere. Deigarria da Nafarroa Beherea eta Zuberoako hizkeren artean dagoen harreman txarra: muga administratiboa ia bat dator hizkuntza mugarekin.

Hirugarren mapan, aldiz, similaritate indize bera erabiltzen bada ere (GIW), multzokatze algoritmoa aldatu egin da (MedMw da orain) eta hau ere 4 multzotan banandu da. Bietan, beraz, jatorrizko datu berak erabili dira. Aldaketa bakarra ikusarazpenerako algoritmo desberdina erabiltzen dela da. Horrela, multzokatze algoritmo desberdinek datuetan duten eragina begi bistan gelditzen da.

<sup>10</sup> Ikus adibidez *Hizkuntzalaritza Hiztegia*-1: 268.

<sup>11</sup> Goebel (1992: 451-452 eta 15. mapa); Philips (1984: 295ko mapa).

Aurrera baino lehen, esan behar da ere hizkeren arteko hizkuntza hurbiltasun eta urruntasuna algoritmoak ematen dituen multzoei dagozkien kolorez gain marra loditasunaz ere adierazten dela: zenbat eta marra lodiagoa izan hizkuntza distantzia handiagoa izango da; aldiz, zenbat eta marra meheagoa izan hizkuntza distantzia txikiagoa.

2. mapan hurbiltasun linguistiko ona duten hizkera gehienak bi gunetan agertzen dira: batetik, Zuberoan eta bestetik Lapurdin (hau ez hain argia). Lapurdi eta Nafarroa Beherearen muga inguruetan hizkuntz muga zehatza markatzen da (kolore urdinak, argi eta ilunak). Nafarroa Behereko lurretan ez da espazio kohesionaturik agertzen (ugariak dira urdin argi eta iluneko isoglosak, nahiz badiren tartekamarteka kolore okreak ere eta gorri bakarñoak). Esan daiteke mapa honek erakusten duenetik Nafarroa Behereko lurretan ez dela hurbiltasun handia erakusten duten hizkerez osatutako hizkuntza eremu argirik agertzen. Hizkuntza homogeneousan txikia erakusten du eremu honek.

Nafarroa Beherea eta Zuberoaren arteko mugan isoglosa urdinak dira nagusi bi probintziak ongi bereziz, batez ere hegoaldean, ez hainbeste iparraldean. Linguistikoki ere muga argia dagoela erakusten du honek eta eten garrantzitsuaz hitz egin beharko litzatekeela erakusten du mapak.

Zuberoa da hizkeren arteko hizkuntza hurbiltasun handiena erakusten duen eremua, Pagola salbu. Datu hauen arabera Zuberoako hizkerek homogeneousan linguistiko handia erakusten dute. Hauxe da azter-eremutik gune trinkoena. Pagolako kasua berezia da, bere inguruko herri guztiekin hizkuntza harreman txarrak erakusten ditu. Geroago ikusiko den bezala hurbiltasun handiagoa erakusten du mendebaldeko dauden hizkerekin bere ingurukoekin baino. Kasu berezitzat hartu beharrekoa dela deritzogu. Orain arteko literatura zientifikoan ez da inoiz aipatu berezitasun hau eta honek bideratzen gaitu herri honetako testuaren egileak edo testuak berak ezaugarri propio erabakigarriak dituela esatera. Ez dirudi herri horretako hizkeraren ispilutzat har daitekeenik. Geroago erakutsiko da, hala ere, ahaidetasun handiena zein herritako testuarekin duen.

Hirugarren mapak, MedMW algoritmoaren bidez lortu denak, aldiz, hizkuntza eremu argiakoak erakusten ditu. Algoritmo honek hizkeren arteko kontrasteak areagotu egiten ditu, hurbiltasunak hurbilduz eta urruntasunak urrunduz.

Hizkuntza eremu gardenena, orain gardenago, Zuberoakoa da, zalantzarik gabe. Hemen biltzen dira isoglosa gorri gehienak. Bestalde, Lapurdi eta Nafarroa Beherearen arteko hizkuntza muga mardulago bihurtu da algoritmo honekin eta era berean Nafarroa Beherea eta Zuberoaren artekoa.

Mapa bietan Lapurdi eta Nafarroa Beherea osoko lurretan trantsizio edo zubi eremu handia agertzen da. Aldiz, azken hau eta Zuberoaren artean hizkuntza etena agertzen da, maparen behe aldean (hots, eremu menditsuetan) batez ere.

## 11.2. Similaritate-mapa (4. maparen interpretazio geolinguistikoa)

Mapa hau Delaunay-ren triangulazioan oinarritzen da eta hizkerak funtzio komunikatiboan lantzen ditu; hau da, hizkeren arteko berdintasunaren edo antzekotasunaren azterketa da.



Baina analisi estatistiko honek 2. eta 3. mapek duten muga bera du: elkarren alboko hizkerak bakarrik konpara daitezkeela. Hizkerak binaka eta hurbilen daude baina ez ditu lotzen eta aztertzen. Edo beste era batera esanda, hizkera bakoitza inguruan dituenekin baina ez duela erkatzen. Hala ere, bere mugak eta guzti, oso interesgarria da ikuspuntu geolinguistikotik. Izan ere gune dialektal homogeneousko detektatzeko baliabideak eskaintzen ditu: lotura marra zenbat eta lodiagoa izan, eta zenbat eta «beroagoa», orduan eta hurbiltasun handiagoa izango dute lotzen dituen hizkerak.

Erraz atzemango denez, aurreko maparen kontrako ikuspegia eskaintzen du: hartan desberdintasun edo distantziaren ikuspegia eskaintzen zen bitartean, honako honek similaritate, berdintasun edo antzekotasunen ikuspegia eskaintzen du.

Mapan eremu dialektalen bat aipatu behar balitz Zuberoako erdigune zabalean agertzen dena egin beharko litzateke, Amikuzeko bazter hizkeretaraino heltzen dena hain zuzen ere. Eremu honetan kokatuta dauden hizkerak hizkuntza harreman onak edo oso onak dituztela esan behar da. Eremu trinkoa, homogeneousotasun handikoa eskaintzen dute. Nafarroa Beherean, aldiz, hizkuntza harreman ona duten hizkerak agertzen diren arren ez dute eremu homogeneouso argirik eskaintzen. Harreman onak haranka kokatzen dira: Oztibarren (1306-Ibarrola, 307-Irisarri, 1308-Jutsi, 1309-Landibarre, 1310-Larزابale eta 1311-Izura hizkeren artean), eta 1299-1232-1237, 1233 hizkeren artean. Berdintsu esan daiteke Aldude, Banka, Baigorri, Arnegi, Uharte-Garazi eta Lasaren arteko (1317, 1320, 1315, 1319, 1331, 1323). Baina oro har, Nafarroa Beherean ez da homogeneousotasunik duen hizkuntza eremu handirik.

Lapurdin ere antzeko egoerarekin egiten dugu topo. Eta Nafarroa Behereko hizkeraz esan denak hemen ere balio du. Aipagarria da Lapurdiko ekialdean, Nafarroa Behereko mugetan, oso hizkuntza harreman txarra duten hizkerak aurkitzen direla. Hauetan media aritmetikotik beherako zenbakiak aurkitzen dira.

Hizkuntza harreman txarrak dituzten hizkeren artean kasu berezia da, jada aipatua izan dena, Pagolako (1297) hizkera. Bere inguruko hizkera guztiakin ditu harreman txarrak, salbuespenik gabe. Antzeko kasuan aurkitzen da Lukuze-Altzumarta (1292); honek ere media aritmetikotik beherako zenbakiak agertzen ditu bere inguruko hizkeretik harremanei dagokienez. Hendaia ere berezia da, baina gabetasun handia eskaintzen du herri honetako testuak jatorrizko frantses testuaren itzulpen aski librea izanik. Luhuso (1224) eta Amorotzek (1272) ere ez dituzte bat ere onak, nahiz aurreragokoek baina hobeak diren.

Oro har, esan daiteke, Zuberoako kasua salbu, ez dela eremu dialektal homogeneouso handirik aurkitzen Ipar Euskal Herriko hizkeretan eskuarteko datuen arabera.

### 11.3. Mapa koropletak: hizkeren arteko mapak (5, 6, 7 eta 8 mapak)

Mapa sinoptikoetan hizkera bakoitzak gainerako guztiakin dituen hizkuntza harremanak nolakoak diren aztertzen da; horrela hizkera bakoitzaren profila zehatz daiteke.

Orain arteko mapetan (2., 3. eta 4. mapak) hizkera bakoitza bere hurbilenekekiko harremanetan ezarri da, gune dialektal homogeneousoenak zein diren lortzeko. Baina hizkuntza konparatze edo erkatze horrek ez du urrunago joaterik. Hurbilenekoen arteko harremanak nolakoak diren zehazteko bakarrik balio du.

Mapa sinoptikoekin, aldiz, beste urrats geolinguistiko berri bat egiten da: hizkera bat izanik abiapuntu, azter-eremu osoko gainerako hizkera guztiak aztertzen dira, harekin duten hizkuntza hurbiltasun edo urruntasun zehazteko. Gainera, hizkera guztiakin harremanetan jartze horrez gain, bildu den corpus osoa kontuan hartuz egiten da. Kasu honetan lexikoari dagozkion 151 erregistrotan bildutako datuak erabiltzen dira.

Era honetan «similaritate mapak» deitu izan direnak lortzen dira. Dialektometria interpuntualaren ikusarazpen tipikoak dira mapa hauek. Mapa koropletak ere deituak izan dira, H. Goebel-en terminologian.

Similaritate mapetan abiapuntuko hizkera aukeratuz gero (mapan zuriz agertzen den hizkerari dagokion puntua da) gainerako hizkerekin dituen similaritateak edo antzekotasunak neurtzen dira algoritmo bat erabiliz. Algoritmo horren emaitzak multzo desberdinetan ezartzen dira eta similaritate multzo bakoitzari kolore bat (gorriak similaritate maila handiena adierazten du eta urdin ilunak txikiena) esleitzen zaio mapan marraztuta ager dadin. Horrela, mapan, aukeratu den hizkeratik abiatuz, linguistikoki hurbilenak dituen hizkerak gorriaz agertuko dira eta urrunenak urdin ilunez; tartean kolore larriak, horiak, orlegiak eta urdin argiak aurkitzen dira hizkuntz urruntasuna mailaka adieraziz.

5. mapan (1209-Urruña da abiapuntua) media aritmetikotik goranzko portzentajeak (lelendako 4, 5 eta 6 multzoak) horiz, laranja eta gorri marraztuta agertzen dira eta hizkuntza afinitate handia adierazten dute. Urruña hizkuntza antzekotasun edo afinitate handiena Lapurdiko hizkerekin erakusten du (hurbiltasun handieneko multzoan 13 hizkera kokatzen dira, denak Lapurdikoak, Banka-1320 izan ezik); aldiz, Lapurdiko hizkera guztiak multzo honetan kokatuta ez dauden arren, denak kokatzen dira media aritmetikotik goranzko multzoetan, bi hizkera izan ezik, Bardoze-1243 eta Hendaia-1208. Geografikoki oso hurbil dagoen azken hizkera hau kontu handiz aztertu behar da; izan ere jatorrizko frantses testutik eta gainerako hizkeretako testuetatik asko aldentzen da itzulpenean erabili den prozedura dela medio: oso itzulpen libre eta laburtua da, eta ondorioz jatorrizko testuak dituen kontzeptu asko ez dira agertzen eta erantzun gabeko lauki asko utzi ditu datu-basean.

Nafarroa Behereko hizkerari erreparatzen bazaie bitan banandu beharko lirateke: hori eta laranja koloreztatutakoak (mediatik gorako multzoetan kokatzen direnak), eta orlegiz zein urdinez koloreztatuak (mediatik beheko multzoetan kokatzen direnak). Egia esan, gehiago dira hizkuntza harreman ona erakusten dutenak (hori eta laranja koloreztatutako hizkerak). Ohartu behar dugu mapa honetan media aritmetikoa %76,41ean kokatzen dela. Beraz, 4tik 3tan baino gehiagotan erantzun bera jaso da herri hauetan. Homogeneitate handia erakusten du datu honek.

Zuberoako eta Nafarroa Behereko ekialdeko, eta batez ere Amikuzeko hizkerak dira hizkuntza distantzia handiena erakusten dutenak, abiapuntutik urrunen kokatzen direnak, alegia. Oro har, hizkera honek (Urruña) gainerako hizkerekin alderatuta hizkuntza harreman onak dituela esan behar da; izan ere 70 hizkera kokatzen dira media aritmetikotik goranzko taldeetan eta 67 hizkera media aritmetikotik beheranzkoetan.

6. mapan abiapuntua Nafarroa Behereko mendebalde-mendebaldean kokatu da, Bidarrain (1321) hain zuzen ere. Mapa hau aurrekoan agertzen denaren berrespena

dugula esan genezake: Lapurdi eta Mendebaldeko Nafarroa Behereko hizkerekin du hizkuntza afinitate edo hurbiltasun handiena. Kopurutara joz %54,74 hizkerekin (kopuru errealetan 75 hizkera) harreman onak ditu (kolore hori, laranja eta gorriak dituzten hizkerak). Hala ere, eremu honetan ere, tartean, badira hain hizkuntza harreman ona ez duten hizkerak ere; hala nola Hendaia (1208), Biriatu (1205), Sara (1225), Larresoro (1216), Aldude (1317) eta Azkarate (1319).

Afinitate gutxi erakusten dutenen eremua, aurreko mapan agertzen den bezala hemen ere, Zuberoako hizkerak eta Amikuzekoek osatzen dute.

5. eta 6. mapen similaritate maiztasunen distribuzioa aztertuz esan daiteke antz handia dutela, bietan kurba aski orekatua baita, bi aldeetan hizkera kopuru berdintua agertzen delako.

7. mapan Pagolako (1297) hizkera, edo Pagolan jaso den testua esan beharko litzateke kasu honetan, aztertzeko erabiliko dugu. Beste ezer baino lehen, mapa honek erakusten du herri horretan bildu den testuak ez duela herriko hizkera islatzen, baldin eta inguruetakoa beste herrietako testuek biltzen badute.

Pagolako datuak Zuberoako gainerako hizkeretakoetatik (Domintxaine-1283 eta Etxarri-1284 salbu) guztiz aldentzen dira; oso harreman txarrak erakusten ditu, media aritmetikotik beherako multzoetan kokatu direlarik. Horrez gain, geografikoki hurbil dituen Nafarroa Behereko hizkera askorekin ere berdintu jokutzen du.

Alderantzizko norabidean ere antzeko gauza gertatzen da: inguruko herriek ere hizkuntza harreman txarrak erakusten dituzte Pagolarekin erkatzean.

Linguistikoki hurbilen dituen hizkerak Nafarroa Beherea eta Lapurdin barreiatuak agertzen dira. Aldiz, eremu osoa kontuan hartuz oso portzentaje onak lortzen ditu: hizkera gehiago dira hizkuntza harreman onak lortzen dituztenak, txarrak baino (61 hizkera kokatzen dira media aritmetikotik beherako multzoetan eta 77 gorakoetan).

Honek testuaren egile edo errientaren etorkiaz edo bere izenpean bidalitako testuaren jatorriaz hitz egitera eraman gaitzake. Hemen egin daitekeen baino sakonago aztertu behar den arren, zertzelada batzuk eman ditzakegu:

- a) Pagolan bildu den testuak ez du, itxura guztien arabera, herri horretako hizkera islatzen.
- b) Zaila da, testu horretan agertzen den hizkera kontuan izanik, gune geografiko batean kokatzea eta, beraz, biltzailearen jatorriari buruzko ondorioak ere ezin atera daiteke.
- c) Gainerako hizkera askok baino hizkuntza harreman hobekak ditu azterketa honetan agertzen diren Ipar Euskal Herriko hizkerekin.

8. mapan Zuberoako hiriburuaren profila aztertzen da (Maule-1249). Mapak garbi erakusten du Zuberoako lurraldearen berezitasuna: hizkuntza hurbiltasun handiena erakusten duten bi multzoetan kokatzen diren hizkerak Zuberoakoak dira, Amikuzeko bi izan ezik (Arberatz-Zilhekoa-1273 eta Gamue-Zohazti-1282). Zuberoako hizkeren homogeneousitasun handia erakusten da mapan. Baditu, hala ere, harreman onak Nafarroa Behereko hizkera batzuekin, batez ere Amikuzeko hizkerekin. Harreman txarragoak ditu Lapurdikoekin, baina ez guztiz okerrak. Izan ere urrunen daudenen multzoan ezaugarri bereziak biltzen dituen Hendaia baino ez da kokatzen.

Laburbilduz eta eremu guztia kontuan hartuz, Maulek harreman txar gehiago du, on baino (80 hizkerekin harreman txarrak lortzen ditu eta 57rekin onak).

#### 11.4. Similaritate mapetatik harantz: eremuak lantzen

##### 11.4.1. *Gune dialektalen bilaketan (9. mapa)*

Gune dialektalen kontzeptua (fr. noyau dialectal) oso loturik agertzen da dialektoen izaerarekin. Izan ere hizkuntza ezaugarri batzuen konkomitantzia territoriala dela bide homogeneousan geotipologikoa ematen da; hots, ezaugarri jakin batzuek lurralde batean agertzen direnean homogeneousan linguistikoa bat ematen da ezaugarri horien tipologia duten hizkerak gune geografiko horretan kokaturik agertzen direlarik. Hain zuzen ere, dialektoaren hezurdura osotuko luke gune dialektalak eta ezaugarri propio eta nabariak bilduko lituzke.

H. Goebel-ek dioenez (2003: 80) gune dialektal kontzeptuak bi gauza eskatzen edo behar ditu:

- a) oihal dialektal bat, zeinek gune dialektal bat baino gehiago dituen;
- b) gune dialektal desberdinen artean isoglosa sarea edo iragan eremua («amphizones»), zeinetan hizkuntza ezaugarri desberdinak nahasten diren eta poliki-poliki, hurbilago edo urrunago, beste gune dialektal bat sortzen den.

Gune dialektal hauek bilatzeko estatistika deskriptiboak eskaintzen dituen lanabesetatik distribuzio maximoak kontuan izanik, elkarren arteko harremanetan bakoitzaren similaritate koefiziente handiena aukeratzen da. Izatez, koefiziente handien edo maxima hauek hurbilen dituzten hizkerekin lortzen dira. Datu hauek lortzeko  $IPI(1)_{jk}$  indizea aukeratu da, hedadura txikiena duten ezaugarriak indartzen dituelako, hedadura handia edo unibertsalagoak diren ezaugarrien gainetik jarri;  $IRI_{jk}$  indizea erabiliz gero ezaugarri guztiak berdin tratatuak izango lirarteke eta.

Indize ponderatu hau oso erabilgarria da gune dialektaletan agertzen diren ezaugarri bakar eta isolatuak indartzen dituelako. Hedadura txikiko ezaugarriak izaten dira garrantzitsuenak gune hauek lokalizatzean. Ondorioz, esan daiteke, gune bakoitzak bere ezaugarri propioak izan behar dituela; eta horiek dira, hain zuzen ere, bereizi egiten dutenak.

Prozedura hauen bidez landutako gaia 9. mapara proiektatu da. Ipar Euskal Herriko hizkeren egituraketa geolinguistikokoaren azterketarako lehen hurbilketa azaltzen zaigu bertan: gorri eta laranja koloreztatuta agertzen diren hizkerak dira ingurukoekin erkatzean similaritate indize handienak lortu dituztenak; urdinez agertzen direnak, aldiz, similaritate indize baxuenak dituztenak. Beste berba batzuk erabiliz, gorri-laranja (eta horiz) koloreztatuta dauden hizkerak hizkuntza-gune homogeneousan osatzen dute.

Gauzak horrela, mapan ekarpen esanguratsuak agertzen dira:

- 1) maparen eskuinaldean agertzen den eremu esanguratsua kanpo ez dago gune dialektal garbirik Ipar Euskal Herrian eskuarteke datuen arabera;
- 2) Zuberoan kokatzen da mediatik goranzko multzoetan kokatuta dauden hizkera gehienak. Honek garbi adierazten du aztertzen dihardugun eremuan Zuberoan kokatzen dela gune dialektal argiena.

- 3) Maparen erdialdean, Nafarroa Behereko lurretan agertzen den mozqueta edo etena (urdin ilunak nagusiki eremu honetan agertzen dira).
- 4) Lapurdin gune txiki baten zantzuak agertzen dira Senpere (1217) inguruan kokaturik.
- 5) Nafarroa Beherean agertzen dira ezaugarri hedatuenak biltzen dituzten hizkera gehienak, baina aldi berean ezaugarri propioenak dituzten 9 herri ere. Hizkera hauetako gehienak lurraldearen ekialdean kokatzen dira: (Donibane-Garazi-1326, Aintzila-1329, Hozta-1305, Bunuze-1303, Arhantsusi-1301, Uharte-Hiri-1298, Arberatze-Zilhekoa-1273, Arboti-Zohota-1275 eta Gamue-Zohazti-1282).

Ondorioz, ikerketa honetan erabili diren datuen arabera esan daiteke Ipar Euskal Herrian hedadura handiko gune dialektal markatu bakarra agertzen dela. Gune hau maparen eskuinaldean, Zuberoako lurraldeak hartzen duen eremuan, kokatzen da. Gune argi horrez gain, esan daiteke badirela beste gune txiki batzuen zantzuak, baina ez dutela gune dialektala izateko nahikoa hedadura geografiko.

#### 11.4.2. 10. maparen aurkezpena (*desbideratze tipikoen mapa*)

«Desbideratze tipikoa» (al. Standardabweichung, fr. «écarts-types», ing. «Standard deviation», esp. «desviación típica») estatistika deskriptiboaren lanabes klasi-koen artean kokatzen da (zehaztasunetarako ikus Goebel 2002: 24).

Dialektologian eta areago dialektometriari desbideratze tipikoen sinopsia zubi-hizkerak edo trantsizio-guneak detektatzeko duen gaitasunagatik erabiltzen da. Trantsizio-gune edo zubi-hizkera hitz gutxitan definitu behar balitz bi sistema edo makrosistema dialektalen (edo «gune dialektal» gorago zehaztu legez) artean kokatzen diren hizkerak osatzen duten gunea dela esan beharko litzateke. Hizkera horien ezaugarri nagusia ezaugarri propio gutxi izatean datza; eta poliki-poliki, urratsez urrats edo ezaugarri ezaugarri, hizkuntza sistema batetik bestera joateko edo iragarteko «bidea» eskaintzen dute.

Tresna honek barne kohesio handia duten makrosistema dialektalen artean kokatzen diren hizkerak islatu eta agirian uzten ditu.

10. mapan bi gune dialektal marrazten dira azter-eremuaren bi aldeetan: Lapurdin eta Zuberoan. Bi gune dialektal garbi hauen tartean agertzen diren hizkerak (urdinez eta orlegiz koloreztatuak) osatzen dute makrosistema hauen arteko trantsizio-eremua. Nafarroa Behereko hizkera gehienek osatuko lukete trantsizio-eremua.

Garrantzi handiko ondorioa dela deritzogu hemen aurkezten dugun mapan agertzen den egituraketari. Izan ere, egituraketa geolinguistikoko hau sintesi kuantitativoaren ondorioa besterik ez da. Sintesi kuantitatibo hori lexikoan oinarrituta dago eta 137 hizkeratako 151 mapa edo kontzepturi emandako erantzunak biltzen ditu datuetan.

Bi gune dialektal eta erdialdeko trantsizio-eremuaren arteko harremanetan sakonduz gero Lapurdi eta Nafarroa Behereko hizkeren arteko trantsizioa eten barik gertatzen den bitartean, Nafarroa Behereko eta Zuberoako hizkeren artekoan eten handia agertzen da, Zuberoakoek gune dialektal trinkoa erakusten dutelako.

### 11.4.3. *Fisherren asimetria-koefizienteen sinopsia (11. maparen aurkezpena)*

Orain arteko mapetan behe-eskuinean agertzen den maiztasunen distribuzioa adierazten duen grafikoa ez da kontuan izan mapak aurkeztean. Baina begirada pausatuz gero, ikus daiteke mapa batetik bestera maiztasunen distribuzioa erakusten duen histograma desberdina izan daitekeela. Maiztasunen distribuzioaren kurba simetrikoa izan edo asimetrikoa garrantzia handikoa da datuen irakurketarako. Asimetria hau neurtzeko estatistika deskriptiboak eskaintzen dituen baliabideetan eza-gunena eta dialektometriari erabiliena «Fisher-en asimetria-koefizientea» izenarekin ezagutzen dena da (al. Schiefe; fr. «coefficiente d'asymétrie de Fisher»; ing. «Skewness»; gazt. «coeficiente de asimetría de Fisher»).<sup>12</sup>

Maiztasunen distribuzioa simetrikoa dela esaten da maiztasunen neurketaren emaitzak media aritmetikoaren alde bietan berdintsu banatzen direnean. Distribuzioa asimetrikoa dela esaten da neurketaren emaitzaren parterik handiena alde batean edo bestean kokatzen denean: media aritmetikotik goranzko emaitzak lortzen direnean asimetria eskuinetarantz joango da (balore positiboak) eta media aritmetikotik beheranzko emaitzak lortzen direnean asimetria ezkerretarantz joango da (balore negatiboak). Asimetria koefizientea «0» izango da kurba orekatua denean.

Datuok ikuspuntu linguistikotik irakurriz media aritmetikoaren gainera dauden similaritate emaitzak lortu dituzten hizkerak (balore positiboak) eremu kontserbatzaileak diren bezala, balore negatiboak lortu dituzten hizkerak gainerako hizkerekin komunikatzeko gaitasun onak dituztela esan ohi da.

11. mapan eremu kontserbatzaile nagusi bi agertzen dira (gorriz eta laranja koloreztatuak), bakoitza mutur batean kokatuta: Lapurdin eta Zuberoan. Ikuspuntu geolinguistikotik begiratu bi eremu hauek ingurukoekin hizkuntza harreman txarrak erakusten dituzte eta balore propio asko gordetzen. Orlegiz eta urdinez koloreztatutako hizkerak, ostera, gainerako hizkerekin harreman handi eta onak dituzte eta aldi berean balio propio gutxi. Hain zuzen ere, 10. mapak erakusten duen kontrako ikuspegia erakusten du mapa honek: han trantsizio-hizkerak detektatu diren bezala hemen gutxi bilakatu diren hizkerak dira nabarmentzen.

Fisher-en asimetria koefizienteak datuen ikuspuntu diakronikoa lantzeko aukera ematen du, besteak beste. Hori da, hain zuzen ere H. Goebl-ek erakusten duena (2002: 26-28; eta 54. orrialdeko mapa) ALFko datuak erabiliz, «Oc» eta «Oil» eremuak aztertzen dituenean. Dena den, ikerketa honetan ez da bide hau landuko.

## 12. DM multidimentsionala edo dendrografikoa

Dialektometria multidimentsionalarekin urrats berri bat ematen da hizkeren egituraketa geolinguistikokoaren azterketan. Era honetako azterketa *cluster analysis* edo sailkapen automatikoan oinarritzen da (fr. classification ascendante hiérarchique-CAH; ing. Hierarchical Cluster Analysis; gazt. clasificación automática ascendente).

Sailkapen automatikoetan ezagutzen diren prozedura desberdinen artean hierarkikoak eta metakorrek erabiliko ditugu. Prozedura hauek «dendrograma» izenez ezagutzen den zuhaitz logiko batez irudikatzen da (ikus 13. mapa).

<sup>12</sup> Ikus Goebl (2003: 81-82).

Prozedura sailkatzaile honen oinarrian azter-eremuko hizkera guztiak agertzen dira. Hizkera horiek 13. mapako dendrogramaren ezker aldean kokatzen dira eta marraren hasiera-puntuari egongo lirake. Hizkera hauek hurbiltasun linguistikoa-  
ren arabera elkarren artean biltzen joango dira: lehenengo bilduko direnak hurbiltasun handiena dutenak izango dira (dendrograman urdin argiz koloreztatutako multzoko azken bigarren eta hirugarren lerroak ordezkatzeko dituzten hizkerak dira hizkuntz hurbiltasun edo antzekotasun handiena dutenak). Ondoren bigarren hurbilenak bilduko dira elkarren artean multzo bat eginez. Maila igo ahala geroago eta multzo handiagoak gorpuztuko dira hizkera guztiak multzo batean ezarri arte. Multzo-lotura bakoitzak bere barruan diren hizkera guztiak biltzen ditu.

Sailkapen prozedura hau aspaldidanik da erabilia giza-zientzietan (soziologian...) eta hizkuntzalaritzan (glotokronologian...). Dialektologia edo geolinguistika ere ez da hizkeren egituraketa geolinguistikoa aztertzean prozedura hauek erabiltzean atzean gelditu; Hans Goebel izan zen erabili zituen lehena eta gaur egun hizkuntza desberdinetan ari da aplikatzen (prozeduraren azalpen argiagorako ikus Goebel 1991, 1992, etab.; Aurrekoetxea 1995: 210-220; Bisquerra 1989: 399-434; Jambu 1989: 331-335).

### 12.1. Dendrogramaren azalpena (12. irudiaren azalpen geolinguistikoa)

Irudika dezagun zuhaitzaren kapelaren irudia dugula begi aurrean (zuhaitz eroriaren irudia, hobe esanik) eta irudiaren eskuinaldetik sortzen diren marra horizontalak zuhaitzaren adarrak direla. Adarrotatik beste adartxo batzuk sortzen direla. Horietako adartxo bakoitzak hizkera bat ordezkatzeko du. Dendrogramaren ezkerretatik sortzen den marra bakoitzak hizkera bat ordezkatzeko du. 137 hizkera aztertzen direnez beste horrenbeste kolorezko marra horizontal agertzen dira.

Tresna estatistiko-sailkatzaile honek hizkerak (adartxoak) berdintasunaren arabera multzokatzen ditu: lehen-lehenik biltzen dituenak berdintsuen diren hizkerak dira. Hasieran hizkera bezainbat lerro daude. Hori da hizkerak berdintasunaren arabera multzokatzen hasteko abiapuntua: azter-eremuan ezaugarri guztiak berdina dituzten bi hizkera egongo balira «0» maila horretan elkartuko lirake; zenbat eta eskuinerako elkartu hizkuntza ezaugarri berdina gutxiago izango lituzkete elkarren artean. Prozeduraren amaieran hizkera guztiak multzo batean biltzen dira, baina honek, bistan denez, ez du interes geolinguistikorik.

Dendrograma sortzeko erabiltzen den prozedura ikusteko sortu da 12. irudia. Irudi honetan zuhaitz hierarkikoaren lehen urratsak agertzen dira: biltzen diren lehen hizkerak bakarrik agertzen dira koloreztatuta. Dendrograma honetan gorri koloreztatu dira maila apalenean biltzen diren lehen 7 hizkerak. Berdintasun eta antzekotasun handiena duten bi hizkerak Nafarroa Behereko Gamue-Zohazti (1282) eta Arberatze-Zilhekoa (1273) dira; ondoren Zuberoaren bihotzean aurkitzen diren Lexantzü-Zunharre (1354) eta Etxebarre (1349) aurkitzen ditugu. Hirugarren multzoan Hozta (1305) eta Bunuze (1303). Eta azkenik, banan-banan multzokatzearen azalpenari amaiera emateko, lehenik bildu diren Gamue-Zohazti (1282) eta Arberatze-Zilhekoa (1273) hizkeren Arboti-Zohota (1275) lotzen zaie.

## 12.2. 13. eta 14. mapen azalpena

13. mapan bi gauza aurkezten dira: dendrograma eta mapa. Mapa ezkerretara duen dendrogramaren proiektzio geolinguistikoa besterik ez da. Azter-eremuko hizkera guztiak biltzen dituen dendrograma modu adierazgarri batean argitaratzeko multzo nagusienak koloreztatu egin dira. Bertan agertzen diren 4 multzo handienak kolorez banandu dira multzoak ondo bereizteko. Multzo bakoitzak duen kolorearekin koloreztatu dira mapan multzo horretako hizkerak.

Mapan 4 multzoek oso eremu koherenteak irudikatzen dituzte, salbuespen txiki batzuekin. Ikus ditzagun banan-banan:

- a) Gorriz koloreztatutako multzoak Lapurdiko hizkera gehienak biltzen ditu eta Nafarroa Behereko bakan batzuk. Eremua nahikoa koherentea da, kanpoan kokatuta aurkitzen diren Larribarre (1290) eta Lukuze-Altzumarta (1292) hizkerak baztertzen badira. Bere eremuaren barnealdean ondoko multzoari dagozkion hizkerak ere biltzen ditu; hala nola Ziburu (1206), Mugerre (1194) eta Jatsu (1215).
- b) Orlegiz koloreztatutako multzoak Nafarroa Behereko mendebaldeko hizkerak hartzen ditu bere barnean. Oso gune trinkoa eta hoherentea erakusten du, eremuz kanpo aurkitzen diren hizkera batzuez landa: jada aipatu diren Lapurdiko 3 hizkerak, eta Urketa (1196), Bardoze (1243) eta Bastida (1238) hirukotea, bestetik; eta azkenik, eremuz kanpo aurkitzen diren beste 4 hizkera binaka bilduta (Amenduze-1271 eta Garruze-1286; eta Donaixti-Ibarrre-1313 eta Pagola-1297).
- c) Urdin argiz koloreztatutako multzoan kokatzen dira Nafarroa Behereko ekialdeko hizkerak eta Zuberoako Domintxaine-Berroeta (1283). Guztira 23 hizkera biltzen dira eremu honetan (Oragarre-1294, Arruta-Sarrikota-1277, Ilharre-1288, Arboti-Zohota-1275, Martxuta-1293, Hozta-1305, Ibarrola-1306, Bunuze-1303, Jutsi-1308, Larzabale-1310, Landibarre-1309, Izura-1311, Arhantsusi-1301, Uharte-hiri-1298, Ostankoa-1295, Bithiriñe-1281, Donapaleu-1268, Behaskane-Laphizketa-1280, Aiziritze-1270, Arberatze-Zilhekoa-1273, Gamue-Zohazti-1282, Gabadi-1285, Behauze-1278, Domintxaine-Berroeta-1283). Eremu hau ere geolinguistikoki oso koherentea da; ez dago eremuz kanpo bere multzoko hizkerarik, nahiz bere barruan multzo honi ez dagozkion hizkera batzuk agertzen diren.
- d) Urdin ilunez koloreztatutako multzoan sartzen dira Zuberoako hizkera guztiak, Pagola eta Domintxaine salbu. Multzo hau da geolinguistikoki eremurik eta koherenteena. Harrigarria da eremu honek erakusten duen kohesio linguistiko eta koherentzia geolinguistikoa.

Mapa honetarako dendrogramako datuak 4 multzotan banandu dira. Ikertzailearen hautua da. Kasu honetan 4 multzotan egitea aukeratu denean Bonaparte printzeak burutu eta gaur egun ere, proposamen berriak gora-behera, sinisgarritasun handiena erakusten duen sailkapena izan da gogoan. Jakina da Bonapartek bildutako datuak eta ikerketa honen oinarrian daudenak ia garai berekoak direla.



14. mapak duen desberdintasun bakarra erabilitako algoritmoan datza:  $IPI(1)_{jk}$  algoritmo ponderatua da kasu honetan erabili dena. Ikus daitekeen legez, aldaketa txiki batzuk baino ez daude; gorritz eta orlegiz koloreztatutako eremuetan. Dena den, ikertzailearen esku dago dendrograman zein mailatan egin mozketak. Bere ikerketaren helburuen arabera jokatu behar du. Halere, lau multzotan banandu ordez hirutan bananduko bagenu Zuberoako lurak agertuko lirateke lehen multzoan. Nafarroa Behereko ekialdeko hizkera gehienak osatuko lukete bigarren multzoa; eta Lapurdi eta Nafarroa Behereko mendebaldeko eremuak bat eginik aurkituko genituzke hirugarren multzoan. (ikus 15. mapa).

### 13. L.L. Bonaparteren mapa eta prozedura estatistiko-sailkatzaileak erabiliz lortutako maparen arteko konparaketa

Aukera egokia da Bonaparte printzeak egindako euskalkien mapa ia-ia garaikide diren datuekin, baina irizpide lexikalak erabiliz eta guztiz bestelako prozedura sailkatzaileak erabiliz lortutakoak parekatzeko edo erkatzeko. Dena den, eta horretan hasi aurretik, kontuan izan behar dira pare bat gauza:

- a) Albo batera uzten dugu oraingoz, eta ikerketa honetan, euskalkitasunaren izaeraz eta euskalki edo dialektoen behin behineko ezaugarriei buruzko eztabaida eta irizpide teorikoen azterketa. Har dezagun dialektoaren kontzeptua definizio laxo batean, lainoztaturik utzita xehetasun eta zehaztasunik behinenak, eta uler dezagun orain artean gure artean nagusiki ulertu den eran.
- b) Gorago iradoki bezala, batean erabili den prozedurak (printzearen sailkapenean) ez du bat ere zerikusik bestean (gurean) erabili denarekin, ez landutako gaien edo ezaugarrietan, ezta erabilitako metodologian. Printzeak hizkuntza ezaugarriak landu zituen bere lankideekin izandako harremanen bidez. Irizpide morfologikoak dira, ezaguna den legez, gehien erabili zituenak. Hizkuntzaren ezaugarri nagusi batzuk bildu eta horien araberrako sailkapena egin zuen. Gure kasuan, eta lehen hurbilpen honetan, ezaugarri lexikalak darabiltzagu, eta bestetik, prozedura sailkatzaileak.
- c) Prozedura sailkatzaileek ez diete aurreiritziei lekurik uzten. Datu multzo bat datu-base batean ezarri behar da, zenbat eta datu gehiago izan harik eta emaitza ziurragoak izango dira. Ikerketa honetan erabili den datu-basean 151 kontzeptu ezarri dira eta kontzeptu horiei 137 herritan jasotako testuak hustuz bildu diren hitzak ezarri zaizkie. Ez zaio alde aurretiko hitz bakoitzaren eremua zein den zehazteari ekin, ezta alde aurreko eremuen finkatzeari ere.
- d) Metodologia sailkatzaileei dagokionez bi similaritate indize erabili dira eta ondorioz bi mapa izango ditugu:  $IPI(1)$  similaritate-indizea eta Ward algoritmoa erabiliz lortutako mapa bata; eta IRI indizea eta Ward algoritmoarekin lortutako mapa, bestea. Lehena indize ponderatua da, bazter ezaugarriei garrantzi handiagoa ematen diena; bigarrena ponderaziorik gabea.

Bonapartek Ipar Euskal Herriko hizkerak lau euskalkitan banandu zituen: lapurtera, mendebaldeko nafarrera beheara, ekialdeko nafarrera beheara eta zuberera.

Lapurteran bi azpieuskalki banandu zituen: berezkoa (hiru aldaerekin: Donibane-koa, Sarakoa eta Ainhoakoa) eta nahasia (Arrangoitzekoa). Mendebaldeko nafarrera behegeran hiru azpieuskalki bereizi zituen: Baigorrikoa, Lapurdikoa (Uztaritze eta Lekorneko aldaerekin) eta Aezkoakoa. Ekialdeko nafarrera behegera hiru azpieuskalkitan: Amikuze-Garazikoa (Garazi, Amikuze, Bardoze eta Arberoako aldaerekin), Aturrikoa (Beskoitze eta Urketako aldaerekin) eta Zaraitzukoa. Azkenik, zuberera bi azpieuskalkitan: berezkoa eta Erronkarikoa.

Printzearen sailkapen honetan agertzen diren Hegoaldeko hizkerak ez dira gure ikerketan kontuan hartu eta Iparraldekoetara mugatuko gara. Guk geuk ere, Bonaparteren maparekin erkatzeko helburuz, lau multzotan banatutako mapa eraikitzea erabaki dugu. Horrela harenarekin erkatzea errazago bihurtzen da (ikus 14. mapa).

Bonaparteren terminologia euskaratua erabiliz gero eta gure mapako (IPI(1)<sub>jk</sub> similaritate indizea eta Ward algoritmoa erabiliz) hizkerak 4 «euskalkitan» banandu behar bagenitu ondokoa genuke:

—Lapurtera: ondoko hizkerek egingo lukete muga ekialdetik, Lehuntze (1193), Mugerre (1194), Jatsu (1215), Kanbo (1222) eta Itsasu (1223). Lapurteraren eremuan mendebaldeko nafarrera behegera hizkerek hurbiltasun handiagoa erakusten duten hizkera bi egongo lirarteke: Ziburu (1206) eta Milafranga (1218).

Lapurterari dagokionez bi banaketan artean badira desberdintasunak:

- a) Bonapartek ekialdeko nafarrera behegera hizkeratzat jotzen dituen Lehuntze eta Mugerre gurean lapurteratzat hartzen dira.
- b) Bonapartek Uztaritzeko aldaerako hizkerak mendebaldeko nafarrera behegeratzat hartu zituen bitartean guk lapurteratzat jotzen ditugu ikerketa honetan.

—Mendebaldeko nafarrera behegera: Urketa (1196), Beskoitze (1238), Hazparne (1227), Makea (1229), Luhuso (1224), Bidarra (1321) izango lirarteke mendebaldetik muga hizkerak. Ekialdetik, aldiz, Bardoze (1243), Amorotze (1272), Mehaine (1230), Armendaritze (1302), Iholdi (1299), Suhuskune (1314), Ainhize-Monjolose (1330), Lekunberri (1339) eta Mendibe (1340). Bonaparteren sailkapenarekin erkatuz gero hark bi nafarrera behegeratan kokatu zituen hizkera gehienak gurean mendebaldeko nafarrera behegera deituan agertzen dira. Salbuespenak bi aldeetan aurkitzen dira: mendebaldean gorago lapurteraren barnean kokatu ditugunak; eta ekialdean Amikuzeko eta bere inguruko Landibarre (1309), Lartzabale (1310), Ibarrola (1306) eta Hozta (1305). Hau da, Gure datuen arabera mendebaldeko nafarrera behegera eremu handiago hartuko luke Bonaparterenak baino.

—Ekialdeko nafarrera behegera: Amikuzeko hizketak eta aurreko paragrafoan agertu diren Landibarre (1309), Lartzabale (1310), Ibarrola (1306) eta Hozta (1305) hizkerek osatuko lukete. Azken hizkera hauek gora-behera, Bonapartek Amikuzeko aldaeran kokatzen dituen herriak, hain zuzen ere. Bonaparteren ekialdeko nafarrera behegerako ekialderen diren hizkerek osatuko lukete gure datuen arabera ekialdeko nafarrera behegera.

—Zuberera: Pagolako hizkera (jada ongi aztertu da zergatia) eta Domintxainekoa (1283) salbu Zuberoako hizkera guztiak biltzen dira. Bonaparteren sailkapenarekin erkatuz Lohitzüne (1291) da hizkera bakarra desberdintzen dena: Bonapartek ekialdeko nafarrera behegeran kokatu zuen bitartean gure datuen arabera zubereran kokatzen da.

#### 14. Ondorioak

Ikerketa honek bi helburu nagusi izan ditu: Ipar Euskal Herriko hizkeren egituratze geolinguistikoan sakontzea Bourciez Bildumako euskal testuetan agertzen diren datuetatik abiatuz, eta hizkerak sailkatzeko metodologia berri bat euskal Dialektologian ezartzea.

Ez zen orain artean datu lexikalak erabiliz Ipar Euskal Herriko hizkuntza sailkapenik eginga. Bestelako datuekin egindakoak gutxietsi gabe, batez ere L. L. Bonapartek egindako euskalkien mapa, beste molde bateko datuak erabiliz, saio berri bat egitea diziplina honetan aurrera-pausuak ematea egoki ez eze beharrezkoa dela sinetsiz. Aurkezten dugun hau lehen ahalegina izan arren emaitza interesgarri eta argigarriak eskaintzen ditu Iparraldeko hizkeren egituratze geolinguistikoan sakontzeko. Ikerketa honen emaitzetan orain artean ezagunak ziren datu batzuk egiaztatatu egiten dira eta beste batzuk zalantzatan ezarri. Datu hauek bildu ziren eran bilduak izanagatik, eta bildu ziren garaian eta metodologiagatik kontuz hartzekoak dira eta ez lukete ondorio handiak hartzera inor bideratu behar.

Bonaparteren maparekin erkatzeko aukera izanik ahalegintxo bat egiteak onura ekarriko duela iritzi diogu, jakinik Bonapartek baino 20 bat urte beranduago bildutako datuak direla. Garaikideak nola edo hala. Gure ikerketatik sortutako mapa Printzearenarekin erkatzean, oro har, antzekotasun handiak ikus daitezke: hark egindako sailkapena eta gure datuetatik eratorritakoak hurbiltasun maila handia dute, Iparraldeko hizkerak osotasunean aztertuz.

Xehetasunetara joz, argi dago zubererari dagokionez, hizkera bakanen bat gorabehera Bonaparteren sailkapena eta gurea bat etortzea guztiz zentzuzkoa eta logikoa den arren, Iparraldeko gainerako euskalkien banaketan jada gauzak ez dira hain gardenak; eta, ondorioz, haren banaketa eta gurea ez datoz bat, nahiz lerro nagusietan desberdintasun handiegirik ez dagoen.

Bigarren helburua hizkerak sailkatzeko eta hizkeren arteko hizkuntza distantziak neurtzeko diziplina honetan erabiltzen den metodologia berria aurkeztu nahi izan dugu. Salzburgoko Unibertsitateko irakaslea den Hans Goebel-ek garatu duen VDM (Visual Dialectometry) programa frogatu nahi izan dugu Iparraldeko hizkerak sailkatzeko. Programa hau munduan zehar ezaguna da eta ikerketa talde askok darabilte jada euren ikerketetan. Goebel-en arlo honetako ibilbidea oso luzea da; izan ere, dialektometria diziplinaren sorrera beretik dihardu metodologia hau erabiltzen eta programa informatikoak ontzen. VDM programa estatistiko-sailkatzaille ahaltsua da guztiz integratzen baititu datu-baseak, tresneria estatistikoak eta mapak.

Hizkeren sailkapen eta hizkuntza distantzia gaiak ikertzeko ezinbestekotzat jotzen dugu honelako tresnak erabiltzea.

## 15. Bibliografia

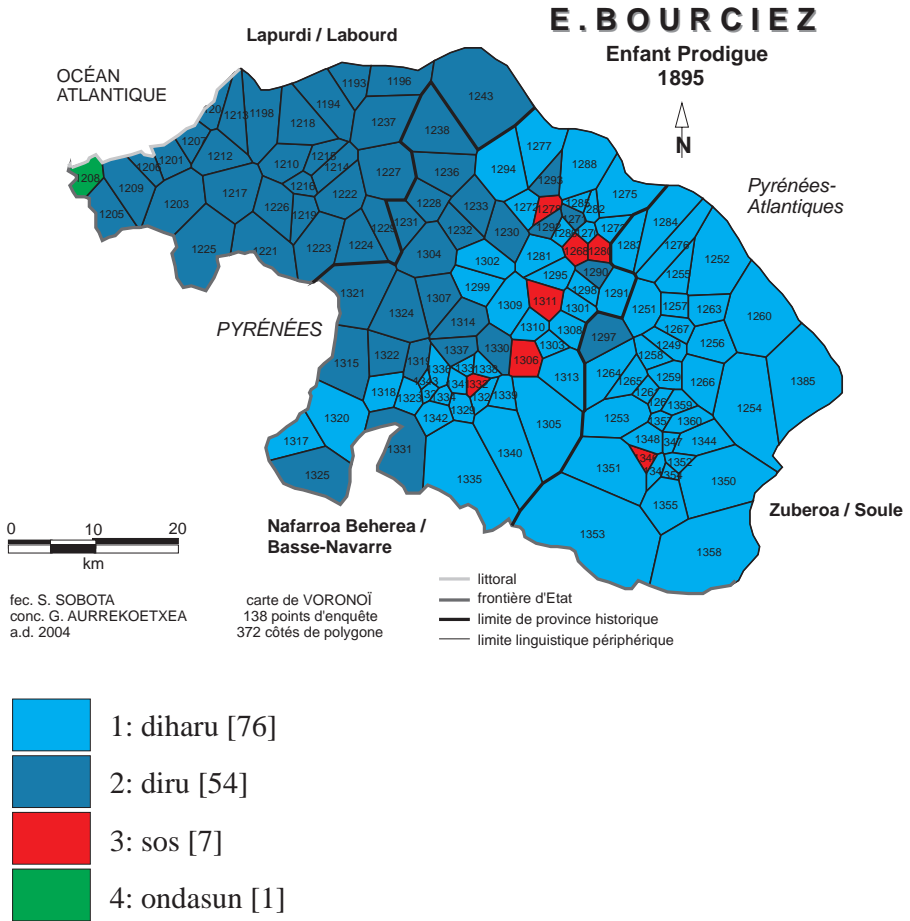
- Allières, J., 1977, «Petit Atlas linguistique basque français Bourciez 1, (20 cartes commentées)», *FLV* 27, 353-386.
- Aurrekoetxea, G., 1992, «Nafarroako euskara: azterketa dialektometrikoa», *Uztaro* 5, 59-109.
- , 1995, *Bizkaieraren egituraketa geolinguistikoa*, UPV/EHUko Argitalpen Zerbitzua, Leioa.
- , eta X. Videgain (arg.), 2004, *Haur prodigoaren parabola Iparraldeko 150 bertsiotan*, ASJU-ren gehigarriak, UPV/EHU, Leioa.
- , A. Iglesias & X. Videgain, 2007, «Bourciez», *bildumako euskal atlasa (BBEA): 1. lexikoa*, (= *ASJU* XXXVIII-2).
- Bailey, K. D., 1994, *Typologies and Taxonomies. An Introduction to Classification Techniques*, Sage, Thousand Oaks / Londres / New Delhi.
- Bisquerra, R., 1989, *Introducción conceptual al análisis multivariable*, PPU, Bartzelona.
- Chandon, J.-L., S. Pinson, 1981, *Analyse typologique. Théories et applications*, Masson, Paris / New York / Bartzelona / Milan.
- Goabl, H., 1981, «Éléments d'analyse dialectométrique (avec application à l'ALS)», *RLiR* 45, 349-420.
- , 1987, «Encore un coup d'oeil dialectométrique sur les *Tableaux phonétiques des patois suisses romands (TPPSR)*. Deux analyses interponctuelles: parquet polygonal et treillis triangulaire», *VR* 46, 91-125.
- , 1991, «Una classificazione gerarchica di dati geolinguistici tratti dall'ALS. Saggio di dialettometria dendrografica», *Linguistica* 31, 341-351.
- , 1992, «Problèmes et méthodes de la dialectométrie actuelle (avec application à l'ALS)», Aurrekoetxea, G. & X. Videgain (arg.), *Nazioarteko dialektologia biltzarra. Agiriak*, Euskaltzaindia, Bilbo, 429-475.
- , 2000, «La dialectométrisation de l'ALF: présentation des premiers résultats», *Linguistica* 40, 209-236.
- , 2002, «Analyse dialectométrique des structures de profondeur de l'ALF», *RLiR* 66, 5-63.
- , 2003, «Regards dialectométriques sur les données de l'Atlas Linguistique de la France (ALF): relations quantitatives et structures de profondeur», *Estudis Romànics* XXV, 59-121.
- Gordon, A. D., 1999, *Classification*, Chapman & Hall, Boca Raton / Londres / New York / Washington D. C.
- Jambu, M., 1989, *Exploration informatique et statistique des données*, Dunod, Paris.
- Philps, D., 1984, «Dialectometrie automatique», in H. Goabl (arg.), *Dialectology Quantitative Linguistics* 21, 275-296.
- Sneath, P. A. & R. R. Sokal, 1973, *Numerical Taxonomy. The Principles and Practice of Numerical Classification*, Freeman, San Francisco.
- UZEI, 1982, *Hizkuntzalaritza Hiztegia*, 2 liburuki, Elkar, Donostia.

## Eranskina:

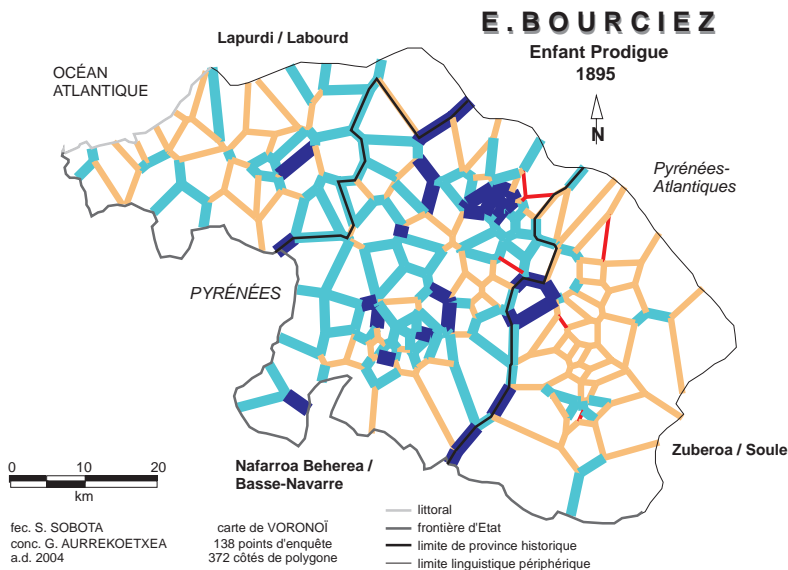
## «Recueil des Idioms de la Région Gascogne» bilduman euskal testuak biltzen dituzten Ipar Euskal Herriko herrien zerrenda

- 1193-Lehunte / Lahonne  
 1194-Mugerre / Mouguerre  
 1196-Urketa / Urcuit  
 1198-Arrangoitze / Arcangues  
 1201-Donibane-Lohizune / Saint Jean-de-Luz  
 1202-Donibane-Lohizune / Saint Jean-de-Luz  
 1203-Azkaine / Ascain  
 1204-Bidarte / Bidart  
 1205-Biriatu / Biriattou  
 1206-Ziburu / Ciboure  
 1207-Getaria / Guéthary  
 1208-Hendaia / Hendaye  
 1209-Urruña / Urrugne  
 1210-Uztaritze / Ustaritz  
 1211-Uztaritze / Ustaritz  
 1212-Ahetze / Ahetze  
 1213-Arbona / Arbonne  
 1214-Haltsu / Halsou  
 1215-Jatsu / Jatzou  
 1216-Larresoro / Larressore  
 1217-Senpere / Saint-Pée-sur-Nivelle  
 1218-Milafanga / Villefranque  
 1219-Ezpeleta / Espelette  
 1220-Ezpeleta / Espelette  
 1221-Ainhoa / Ainhoa  
 1222-Kanbo / Cambo  
 1223-Itsasu / Itxassou  
 1224-Luhuso / Louhossoa  
 1225-Sara / Sare  
 1226-Zuraide / Souraide  
 1227-Hazparne / Hasparren  
 1228-Lekuine / Bonloc  
 1229-Makea / Macaye  
 1230-Mehaine / Méharin  
 1231-Lekorne / Mendionde  
 1232-Donostiri / Saint-Esteben  
 1233-Donamartiri / Saint-Martin  
 1236-Aiherra / Ayherre  
 1237-Beskoitze / Briscois  
 1238-Bastida / Labastide-Clairence  
 1243-Bardoze / Bardos  
 1249-Maule / Mauléon  
 1250-Maule / Mauléon  
 1251-Ainharbe / Ainharp  
 1252-Ürrüstoi-Larrabile / Arrast-Larrebieu  
 1253-Altzurükü / Aussurucq  
 1254-Barkoxe / Barcus  
 1255-Sarrikotapea / Charritte-de-Bas  
 1256-Sohüta / Chéraute  
 1257-Ündüreine / Espès-Undurein  
 1258-Garindaine / Garindein  
 1259-Gotaine-Irabarne / Gotein-Libarrenx  
 1260-Ospitalepea / L'Hôpital-Saint-Blaise  
 1261-Idauze-Mendi / Idaux-Mendy  
 1262-Mendikota / Menditte  
 1263-Mitikile / Moncayolle  
 1264-Muskildi / Musculdy  
 1265-Urdiñarbe / Ordiarp  
 1266-Arrokiaga / Roquiague  
 1267-Bildoze / Viodos  
 1268-Donapaleu / Saint-Palais  
 1269-Donapaleu / Saint-Palais  
 1270-Aiziritze / Aicirits  
 1271-Amenduze / Amendeuix  
 1272-Amorotze / Amorots  
 1273-Arberatze-Zilhekoa / Arbérats-Sillègue  
 1274-Arberatze-Zilhekoa / Arbérats-Sillègue  
 1275-Arboti-Zohota / Arbouet-Sussaute  
 1276-Arüe / Aroue  
 1277-Arruta-Sarrikota / Arraute-Charritte  
 1278-Behauze / Béguios  
 1279-Behauze / Béguios  
 1280-Behaskane-Laphizketa / Béhasque-Lapiste  
 1281-Bithiriñe / Beyrie-sur-Joyeuse  
 1282-Gamue-Zohazti / Camou-Suhast  
 1283-Domintxaine-Berroeta / Domezain-Berraute  
 1284-Etxarri / Etcharry  
 1285-Gabadi / Gabat  
 1286-Garruze / Garris  
 1288-Ilharre / Ilharre  
 1290-Larribarre / Larribar  
 1291-Lohitzüine / Lohitzun  
 1292-Lukuze-Altzumarta / Luxe-Sumberraute  
 1293-Martxuta / Masparraute

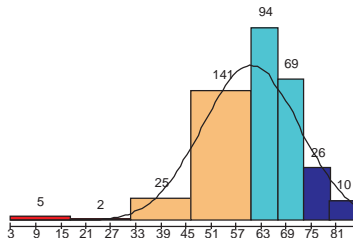
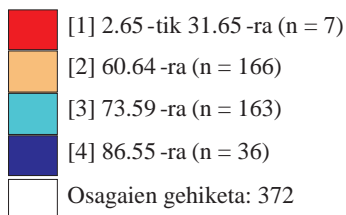
- 1294-Oragarre / Orègue  
 1295-Ostankoa / Orsanco  
 1297-Pagola / Pagolle  
 1298-Uharte-hiri / Uhart-Mixe  
 1299-Iholdi / Iholdy  
 1300-Iholdi / Iholdy  
 1301-Arhantsusi / Arhansus  
 1302-Armendaritze / Armendarits  
 1303-Bunuze / Bunus  
 1304-Heleta / Hélette  
 1305-Hozta / Hosta  
 1306-Ibarrola / Ibarrolle  
 1307-Irisarri / Irissarri  
 1308-Jutsi / Juxue  
 1309-Landibarre / Lantabat  
 1310-Larabale / Larceveau  
 1311-Izura / Ostabat  
 1312-Izura / Ostabat  
 1313-Donaixti-Ibarre / Saint-Just-Ibarre  
 1314-Suhuskune / Suhescun  
 1315-Baigorri / Saint-Etienne-de-Baigorri  
 1316-Baigorri / Saint-Etienne-de-Baigorri  
 1317-Aldude / Aldudes (les)  
 1318-Anhauze / Anhaux  
 1319-Azkarate / Ascarat  
 1320-Banka / Banca  
 1321-Bidarra / Bidarray  
 1322-Irulegi / Iroulégu  
 1323-Lasa / Lasse  
 1324-Ortzaize / Ossès  
 1325-Urepel / Urepel  
 1326-Donibane-Garazi / Saint-Jean-Pied-de-Port  
 1327-Donibane-Garazi / Saint-Jean-Pied-de-Port  
 1328-Ahatsa-Altzieta-Bazkazane / Ahaxe-Alciette-Bascassan  
 1329-Aintzila / Aincille  
 1330-Ainhize-Monjolose / Ainhice-Mongelos  
 1331-Arnegi / Arnéguy  
 1332-Duzunaritze-Sarasketa / Bussunarits-Sarrasquette  
 1333-Buztintze-Hiriberri / Bustince-Iriberry  
 1334-Zaro / Çaro  
 1335-Ezterenzubi / Estérençuby  
 1336-Izpura / Ispoure  
 1337-Jatsu / Jaxu  
 1338-Lakarra / Lacarre  
 1339-Lekunberri / Lecumberry  
 1340-Mendibe / Mendive  
 1341-Donazaharre / Saint-Jean-Le-Vieux  
 1342-Eiheralarre / Saint-Michel  
 1343-Uharte Garazi / Uhart-Cize  
 1344-Atharratze-Sorholüze / Tardets-Sorholus  
 1345-Atharratze / Tardets  
 1346-Altzai / Alçay  
 1347-Aloze / Alos  
 1348-Gamere-Zihiga / Camou-Cihigue  
 1349-Etxebarre / Etchebar  
 1350-Hauze / Haux  
 1351-Lakarri / Lacarry  
 1352-Liginaga / Laguinge  
 1353-Lorraine / Larrau  
 1354-Lexantzü-Zunharre / Lichans-Sunhar  
 1355-Ligi-Atherei / Licq-Athérey  
 1357-Ozaze-Zühara / Ossas-Suhare  
 1358-Santa-Grazi / Sainte-Engrâce  
 1359-Zalgize-Donetztebe / Sauguis-Saint-Etienne  
 1360-Iruri / Trois-Villes  
 1385-Eskiula / Esquiule



1. mapa: 'argent' kontzeptuari dagozkion euskal ordainen Ipar Euskal Herriko banaketa geolinguistikoa.



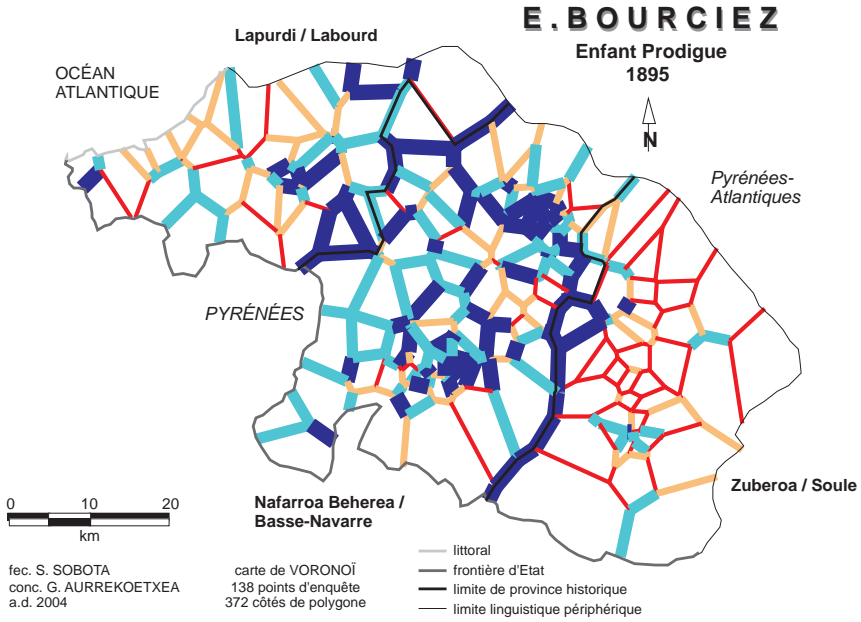
Tartekatze algoritmoa Similaritate maiztasunen distribuzioa  
MINMWMAX 4 multzo MINMWMAX 8 multzo



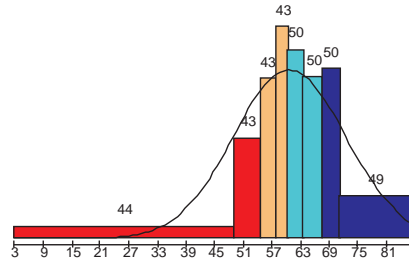
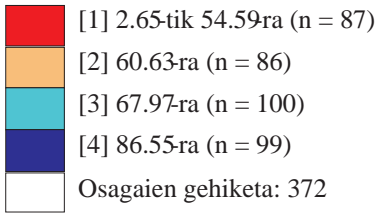
2. mapa: hizkeren funtzio bereizlea erakusten duen puntuarteko mapa (isoglosa sintetiko edo kuantitatiboak)

Corpusa: 151 mapa lexical. Distantzia indizea.  $IPD(1)_{jk}$  edo GIW. Tartekatze algoritmoa: MINMWMAX 4-multzo (ikus 11.1 paragrafoa).



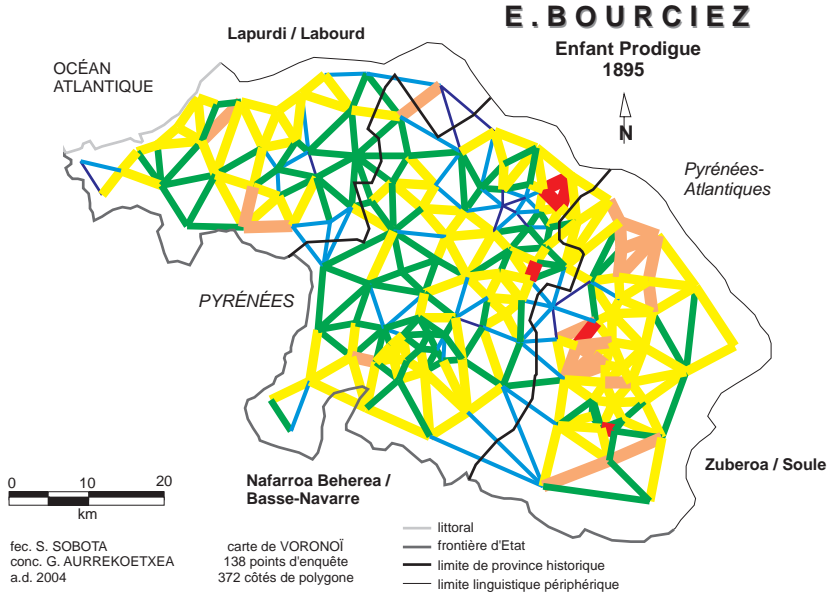


Tartekatze algoritmoa Similaritate maiztasunen distribuzioa  
 MedMw 4 multzo MedMw 8 multzo

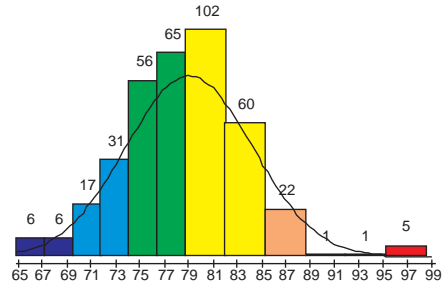
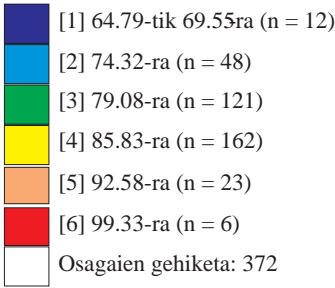


3. mapa: hizkeren funtzio bereizlea erakusten duen puntuarteko mapa (isoglosa sintetiko edo kuantitatiboak)

Corpusa: 151 mapa lexical. Distantzia indizea.  $IPD(1)_{jk}$  edo GIW. Tartekatze algoritmoa: MedMw 4-multzo (ikus 11.1 paragrafoa).



Tartekatze algoritmoa Similaritate maiztasunen distribuzioa  
 MINMWMAX 6 multzo MINMWMAX 12 multzo

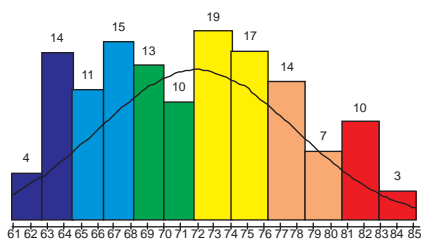
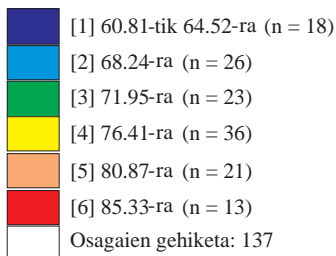


4. mapa: hizkeren funtzio komunikatiboa erakusten duen puntuarteko mapa.

Corpusa: 151 mapa lexikal. Distantzia indizea IRI. Tartekatze algoritmoa: MINMWMAX 6-multzo. Mapa hau Delaunay-ren triangulazioan oinarritzen da (interpretazio geolinguistikorako ikus 11.2 paragrafoa).

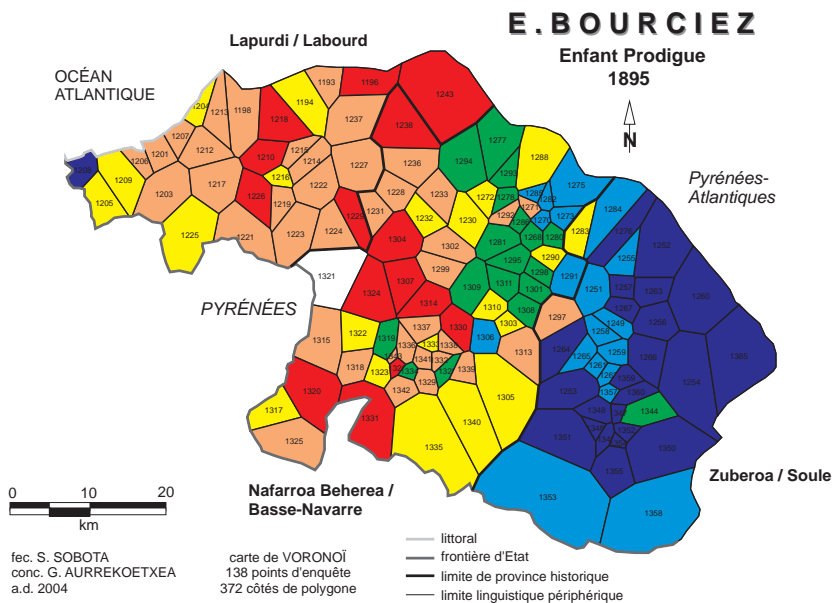


Tartekatze algoritmoa Similaritate maiztasunen distribuzioa  
 MINMWMAX 4 multzo MINMWMAX 8 multzo

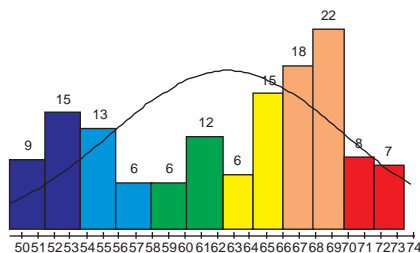
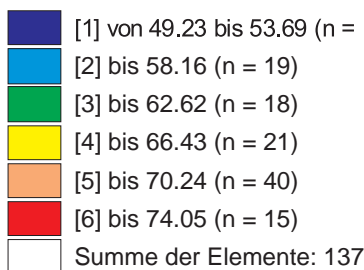


5. mapa: Urruñari (1209) dagokion similaritate ( $IRI_{1209,k}$ ) distribuzioen mapa koropleta.

Corpusa: lexikoari dagozkion 151 mapa. Similaritate indizea  $IRI_{jk}$ . Tartekatze algoritmoa MINMWMAX 6-multzo; 6 kolore desberdin (interpretazio geolinguistikorako ikus 11.3 paragrafoa).

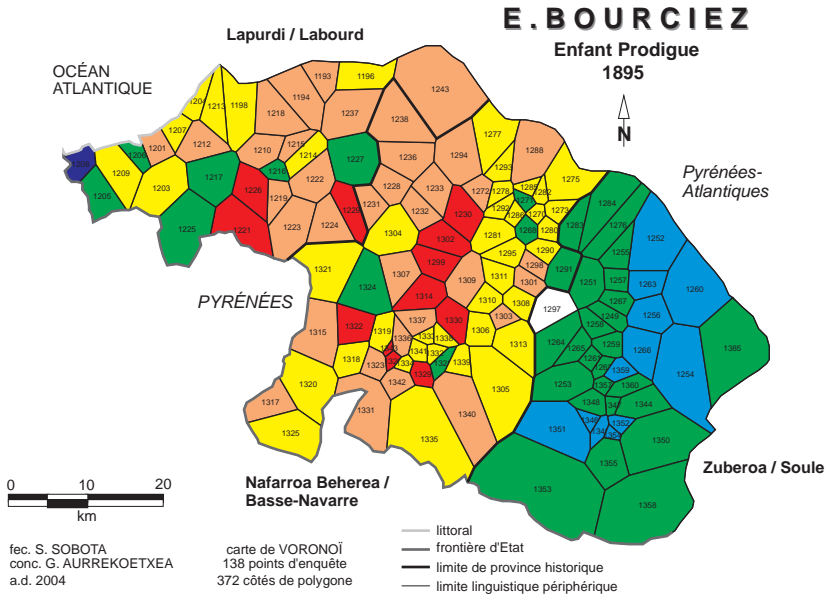


Tartekatze algoritmoa Similaritate maiztasunen distribuzioa  
 MINMWMAX 4 multzo MINMWMAX 8 multzo

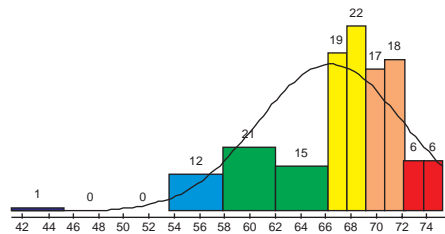
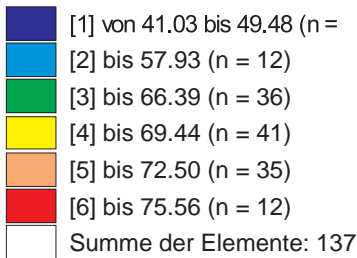


6. mapa: Bidarraini (1321) dagokion similaritate ( $IRI_{1321,k}$ ) distribuzioen mapa koropleta.

Corpusa: lexikoari dagokion 151 mapa. Similaritate indizea  $IRI_{j,k}$ . Tartekatze algoritmoa MINMWMAX 6-multzo; 6 kolore desberdin (interpretazio geolinguistikorako ikus 11.3 paragrafoa).

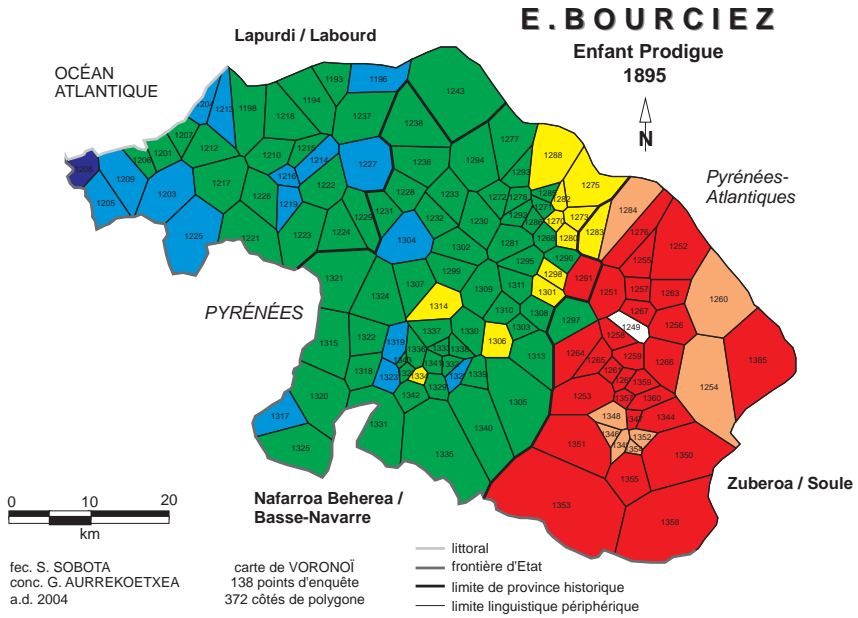


Tartekatze algoritmoa Similaritate maiztasunen distribuzioa  
MINMWMAX 4 multzo MINMWMAX 8 multzo



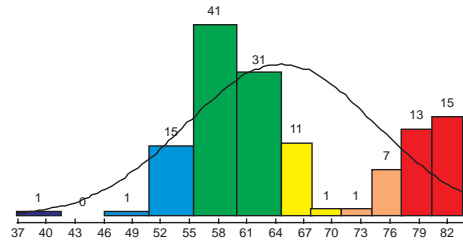
7. mapa: Pagolari (1297) dagokion similaritate ( $IRI_{1297,k}$ ) distribuzioen mapa koropleta.

Corpusa: lexikoari dagozkion 151 mapa. Similaritate indizea  $IRI_{jk}$ . Tartekatze algoritmoa MINMWMAX 6-multzo; 6 kolore desberdin (interpretazio geolinguistikorako ikus 11.3 paragrafoa).



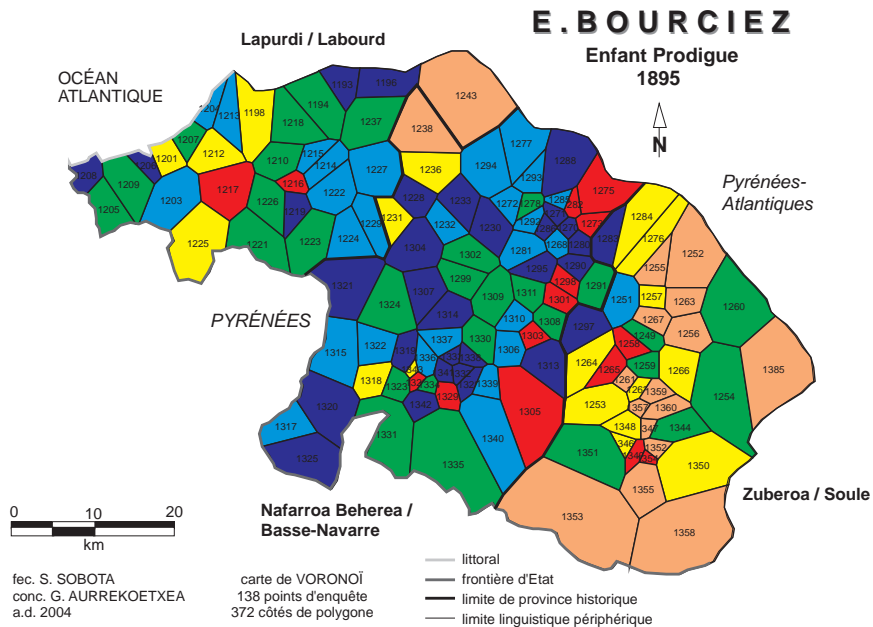
Tartekatze algoritmoa Similaritate maiztasunen distribuzioa  
MINMWMAX 4 multzo MINMWMAX 8 multzo

- [1] von 36.84 bis 46.09 (n = 1)
- [2] bis 55.33 (n = 16)
- [3] bis 64.57 (n = 72)
- [4] bis 71.04 (n = 12)
- [5] bis 77.50 (n = 8)
- [6] bis 83.97 (n = 28)
- Summe der Elemente: 137

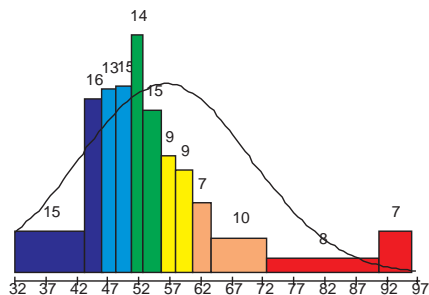
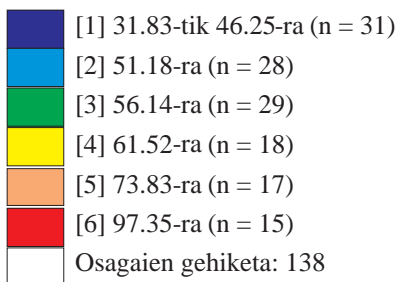


8. mapa: Mauleri (1249) dagokion similaritate ( $IRI_{1,249,k}$ ) distribuzioen mapa koroleta.

Corpusa: lexikoari dagozkion 151 mapa. Similaritate indizea  $IRI_{jk}$ . Tartekatze algoritmoa MINMWMAX 6-multzo; 6 kolore desberdin (interpretazio geolinguistikorako ikus 11.3 paragrafoa).

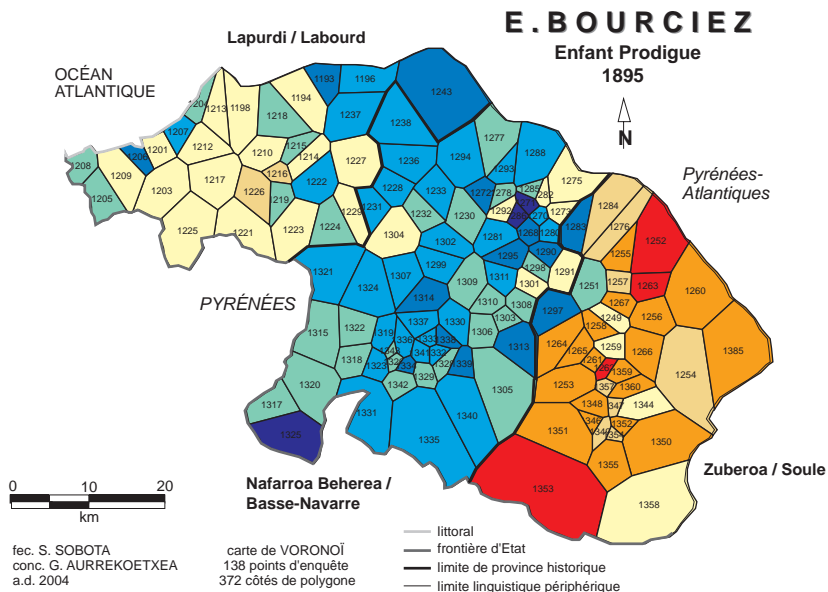


Tartekatze algoritmoa Similaritate maiztasunen distribuzioa  
MedMw 6 multzo MedMw 12 multzo

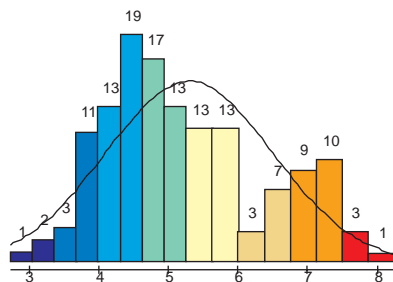
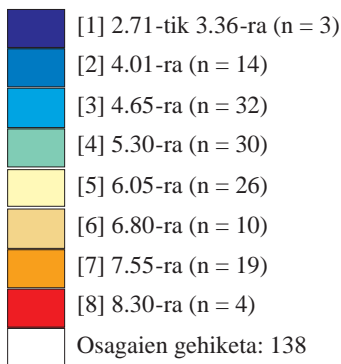


9. mapa: 137 distribuzio maximoen sinopsien mapa koropleta.

Corpusa: 151 mapetan bildutako lexiko osoa;  $IPI(1)_{jk}$  similaritate indizea eta MEDMW 6-multzo tartekatze algoritmoa (6 kolore mapan) erabiliz lortutako mapa (interpretazio geolinguistikorako ikus 11.4.1. paragrafoa).



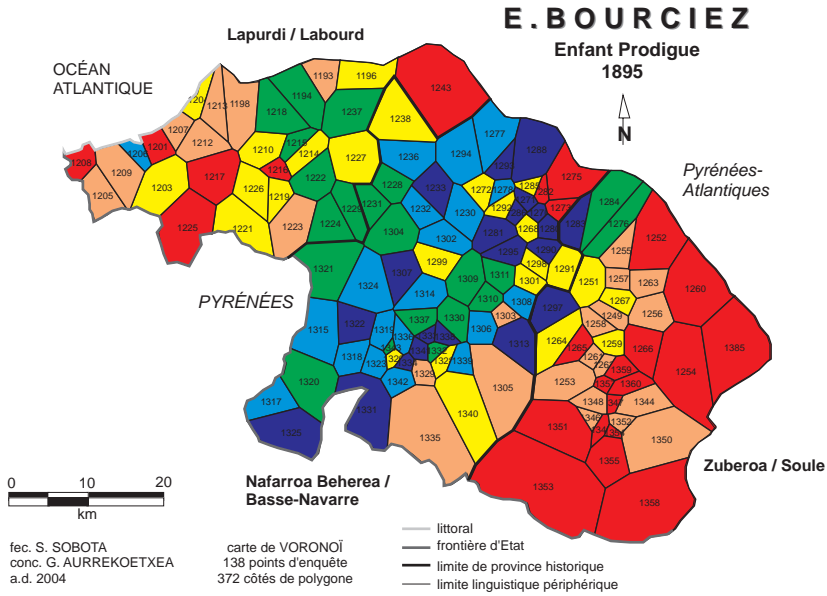
Tartekatze algoritmoa Similaritate maiztasunen distribuzioa  
MINMWMAX 8 multzo MINMWMAX 16 multzo



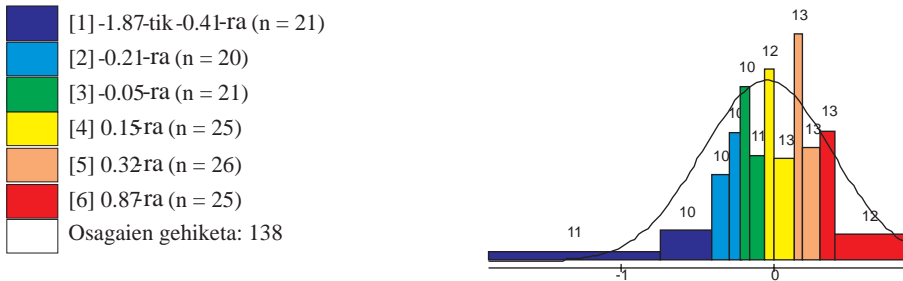
10. mapa: desbideratze tipikoaren mapa.

Corpusa: lexikoari buruzko 151 mapa. Similaritate indizea  $IRI_{jk}$ . Tartekatze algoritmoa MINMWMAX-8 multzo (interpretazio geolinguistikorako ikus 11.4.2. paragrafoa).



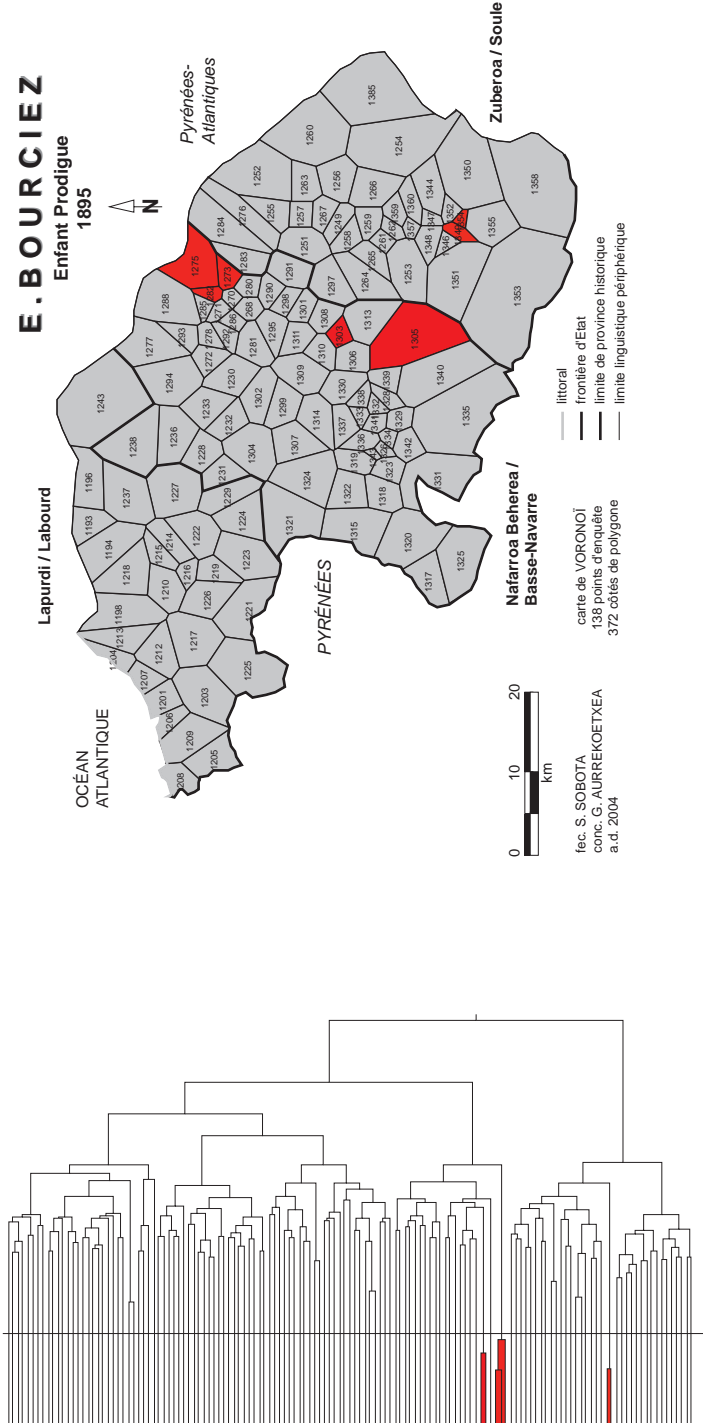


Tartekatze algoritmoa Similaritate maiztasunen distribuzioa  
MedMw 6 multzo MedMw 12 multzo

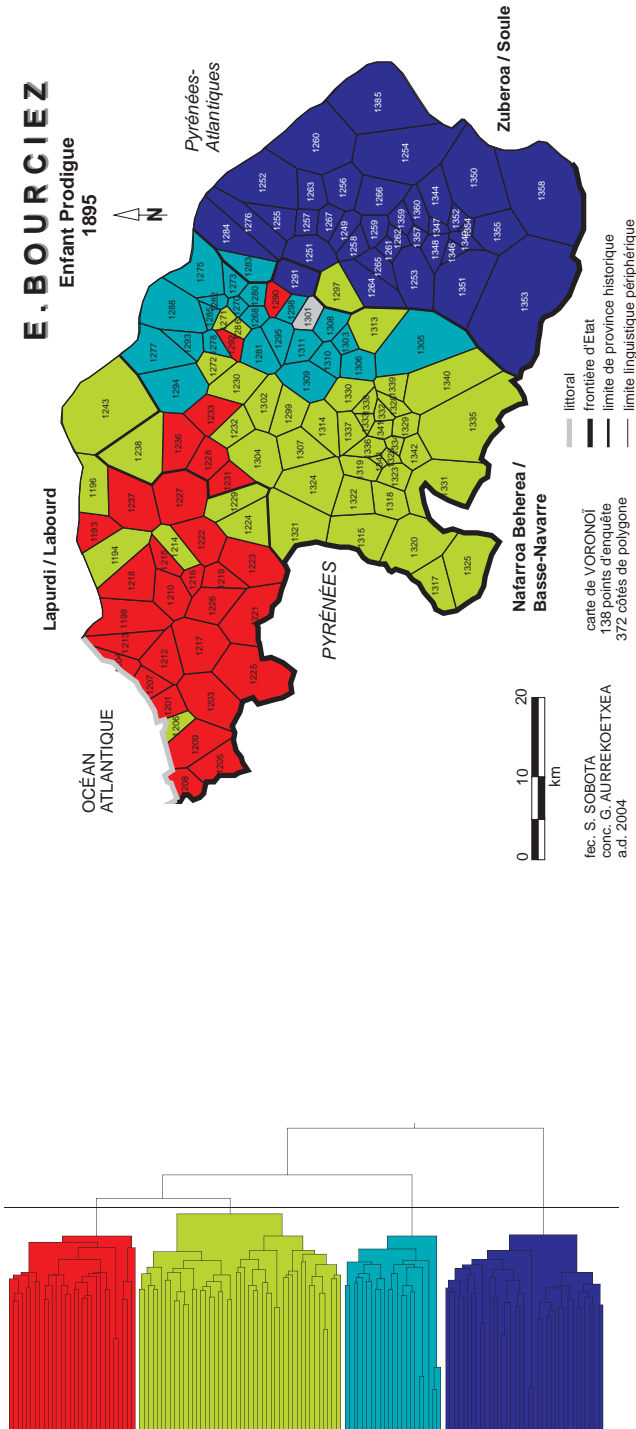


11. mapa: Fisher-en asimetria koefizienteen sinopsiaren mapa koropleta.

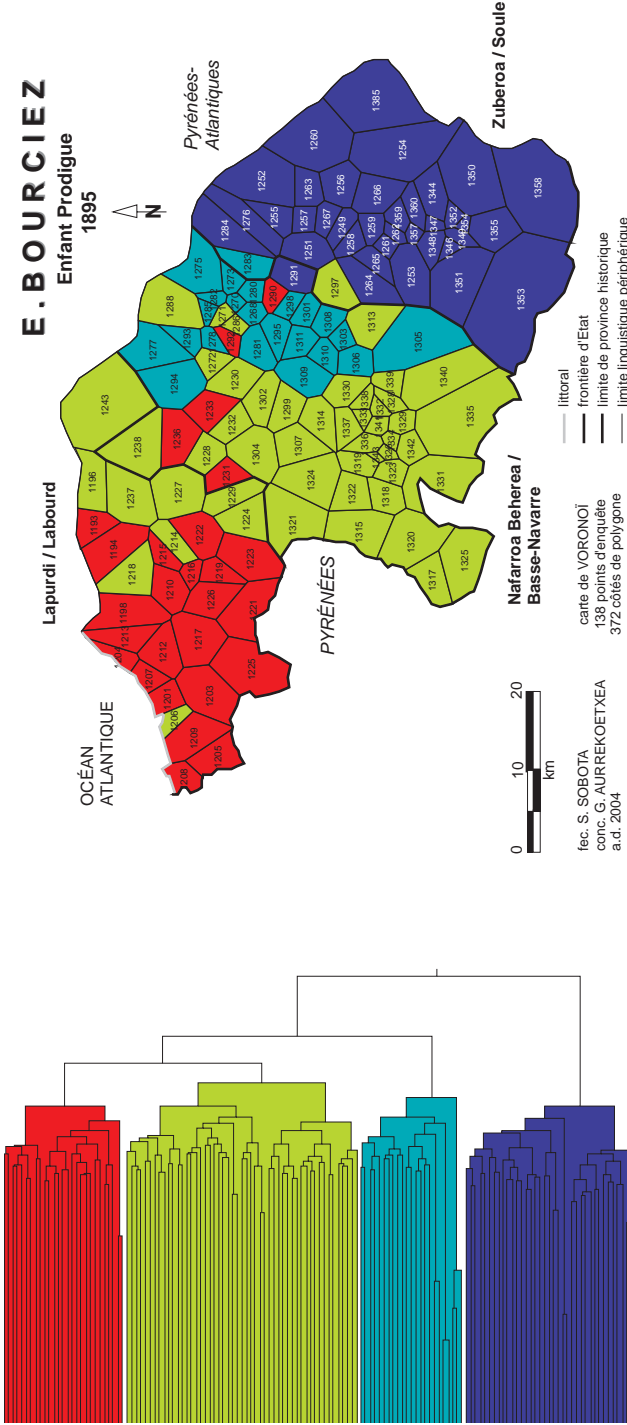
Corpusa: lexikoari dagozkion 151 mapa. Similaritate indizea  $IRI_{jk}$ . Tartekatze algoritmoa MEDMW 6-multzo; 6 kolore mapan (interpretazio geolinguistikorako ikus 11.4.3. paragrafoa)



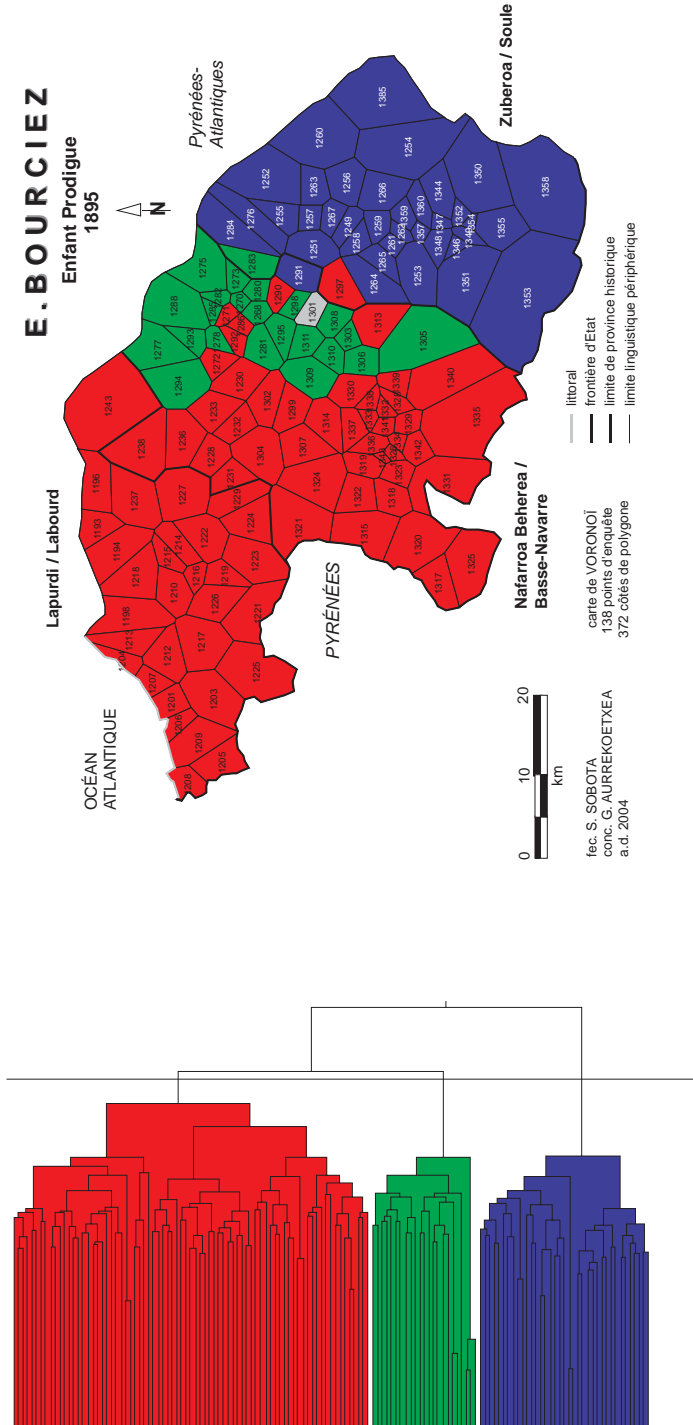
12. irudia: Bourciereren bilduman euskal testuak jaso diren 137 hizkeren sailkapen dendrografikoa. Ward algoritmo metakorra erabiliz. Similaritate indizea  $IRI_{jk}$ . Grafikoan hizkeren multzokatzearen hasiera gorri irudikatutik azaltzen da (interpretatiorako ikus 12.1. paragrafoa).



13. mapa: Dendrogramako datuen egituraketa geolinguistikoa  $IRI_{jk}$  similaritate indizea eta Ward algoritmoa erabiliz



14. mapa: Dendrogramako datuen egituraketa geolinguistikoa IPI(1)<sup>th</sup> similaritate indizea eta Ward algoritmoa erabiliz



15. mapa: Dendrogramako datuen egituraketa geolinguistikoa IRI similaritate indizea eta Ward algoritmoa erabiliz.