

Hezkuntzari buruzko ekarpenak

Contributions on education (EUDIA-8)



EDITOREAK:
Juan Abasolo
Irati de Pablo
Ariane Ensunza

eman ta zabal zazu



Universidad
del País Vasco

Euskal Herriko
Unibertsitatea

Hezkuntzari buruzko ekarpenak (EUDIA-8) [Recurso electrónico] = Contributions on education : Leioan, 2019ko apirilaren 5ean /[editores] Juan Abasolo Isasa, Irati de Pablo Delgado, Ariane Ensunza Aldamizetxebarria. – Datos. - Bilbao : Universidad del País Vasco / Euskal Herriko Unibertsitatea, Argitalpen Zerbitzua = Servicio Editorial, [2020]

1 recurso en línea : PDF (137 p.)

Modo de acceso: World Wide Web.

Textos en euskara e inglés.

ISBN: 978-84-1319-098-3.

1. Euskara (Lengua) - Dialectos. 2. Lenguaje y lenguas - Variación. 3. Prosodia (Lingüística). I. Abasolo Isasa, Juan, coed. II. Pablo Delgado, Irati de, coed. III. Ensunza Aldamizetxebarria, Ariane, coed. IV. Título: Contributions on education.

(0.034)809.169-087

(0.034) 800.87



UPV/EHUko EUDIA Ikerketa-taldea

Euskal Herriko Unibertsitateko ikerketa taldea (GIU 22/16)

Euskal Unibertsitate Sistemako ikerketa taldea (IT 1028/16)

Hezkuntzari buruzko ekarpenak (EUDIA-8)

Contributions on education

Leioan. 2019ko apirilaren 5ean

Juan Abasolo Isasa
Irati de Pablo Delgado
Ariane Ensunza Aldamizetxebarria

erman ta zabal zazu



Universidad
del País Vasco

Euskal Herriko
Unibertsitatea

Aurkibidea

HIZKEREN SAILKAPEN BERRIA	5
1. Sarrera	5
2. Metodologia	7
3. Datuen analisia	9
4. Hizkeren sailkapena	10
4.1. Multzokatze edo cluster analisia	10
Ward Algoritmoa	11
Average algoritmoa	13
Complete algoritmoa	14
4.2. Fuzzy clusterra	15
5. Euskalkien sailkapen berria	16
6. Ondorioak	17
7. Erreferentziak	18
MELODIC ANALYSIS OF SPEECH (MAS): PHONETICS OF INTONATION	20
1. Introduction	20
2. Identification of melodic units	22
2.1. Phonic hierarchy	22
2.2. Melodic units	23
2.3. Establishing the corpus	25
3. Acoustic phase	26
3.1. Determination of relevant frequency values	26
3.2. Standardisation of the frequency data and graphic representation	29
3.3. Praat Script	32
4. Perception phase	34
4.1. Analysis validation	34
4.2. Identification of the melodic features	36
4.3. Forming the hypotheses	38
4.4. Perception tests	38
5. Discussion and concluding thoughts	39
6. References	43
EUSKARA BURUAN ETA IBILI MUNDUAN: AFRIKATUETAN BARNA	48
1. Sarrera	49
2. Corpora eta metodologia	51

3.	Afrikatu motak	52
4.	Euskararen afrikatuez	56
5.	Afrikatuen hedadura	62
6.	Ondorioak	65
7.	Erreferentziak	65
THE SEMANTIC PROFILES OF OLD ENGLISH PRETERITE-PRESENT VERBS OF COGNITION- WITAN AND CUNNAN		68
1.	Introduction	68
2.	The semantics of OE witan	69
2.1.	Witan in pragmatic contexts	71
2.2.	The independent sense of knowing	73
2.3.	Apparently tautological expressions	73
3.	The analysis of the profile of OE cunnan	74
3.1.	The semantic analysis of cunnan in the 10/11th c.	75
3.2.	The semantic analysis of cunnan in the 11th c.	77
4.	Conclusions	78
5.	References	79
AHOZKO TESTUEN PROSODIAREN BILAKAERA JATORRIZKO INGURUNETIK KANPO. EXTREMADURAKO GAZTELANIAREN KASUA BIZKAIAN (BASOURIN)		80
1.	Sarrera	81
2.	Corpusa eta Metodologia	83
3.	Datuen azterketa	86
4.	Ondorioak	94
5.	Erreferentziak	95
DISTANTZIA LINGUISTIKOA NEURTZEKO METODOAK TRANSKRIPZIO MOTAREN ARABERA		96
1.	Sarrera	96
2.	Corpusa	97
3.	Marko teorikoa	98
4.	Metodologia	100
5.	Emaitzak	102
6.	Ondorioak	107
7.	Erreferentziak	108
UMEEI ETA NAGUSIEI EGINEKO TESTUEN PROSODIAREN ALDEEZ		110
1.	Sarrera	111
2.	Metodologia	117
3.	Datuen azterketa	120

3.1. F0	120
3.2. Energia	122
3.3. Iraupena	123
3.4. Bokalen tasunak	129
4. Ondorioak	130
5. Erreferentziak	131

HIZKEREN SAILKAPEN BERRIA

G. Aurrekoetxea (UPV/EHU), I. Gaminde (UPV/EHU), J. L. Ormaetxea (UPV/EHU),
X. Videgain (UPPA/Iker),
gotzonaurre@gmail.com, inaki.gaminde@gmail.com, txipirrin@gmail.com,
charles.videgain@univ-pau.fr

Laburpena

Euskal hizkeren sailkapen hierarkikoa egiten da ekarpen honetan hizkuntzaren alor guztiak kontuan hartuta. Lehen aldiz da egiten honelako sailkapena. Horretarako dialektometria izenez ezagutzen den metodologia erabili da; hau da, EHHA-tik jasotako datuen analisi estatistikoak egin eta emaitzak mapetan ezarri dira. Sailkapen hierarkikoa bezala ezagutzen den teknika estatistikoa erabili da. Lehenik euskalkien eta azpieuskalkien sailkapenak erakusten dira.

Hitz gakoak: dialektologia, dialektometria, sailkapen hierarkikoa, euskalkiak

Abstract

In this work the first classification of Basque dialects is proposed, using techniques known as dialectometry. The statistical technique used is the cluster analysis, with different classification algorithms. The database has been created with the EHHA data (Euskararen Herri Hizkeren Atlas- Basque Linguistic Atlas), which has data from all linguistic categories. The outcomes have been used to create maps of dialects and subdialects.

Key words: dialectology, dialectometry, cluster analysis, Basque

1. Sarrera

Hizkeren sailkapenaz zientifikoki hitz egiten hasteko ezinbestekoa da zientzietan sailkapenak nola egiten diren aipatzea. Egia esan, metodo zientifikoez hitz egitean ez dela metodo bakarra esan behar dugu ezer baino lehen: metodo zientifikoa ez dela bide zurruna, alegia. Gainerakoetan bezala, giza zientzietan edo, zehatzago, hizkuntzalaritzan metodo zientifikoez dihardugunean metodo objektiboez hitz egitea dagokigu. Taxonomia arduratzen da aztertzen dituen objektuak bereizi eta beraien arteko harremanen araberako egitura bilatu eta egitura horretan objektuak kokatzeaz, antzekotasun, berdintasun edo hurbiltasuna kontuan hartuta. Bide honetan, zientziak, gure kasuan zientzia enpirikoak (badira logiko-deduktiboak eta induktiboak ere) ezaugarritu duen prozedura-multzoa hau da: behaketa sistematikoa, neurketa, formulazioa eta analisia; hots, hizkuntza bere osotasunean ezin denez aukeratu, lagin bat, lagin ordezkatzailerik bat aukeratu behar da. Ondoren lagin horren azterketa eta sailkapena edo analisia egiten da.

Ikerketan zehar burutu den jarduera zientifikoki erabiltzen dela ziurtatzeko, bi prozedura erabiltzen dira: neurketa eta arrazoiketa. Batetik, aztertzen diren objektuen desberdintasun eta antzekotasunak zenbatu eta neurtu behar dira. Neurtzeko komunitate zientifikoak onartutako izari edo unitatea(k) erabiltzen d(ir)a. Neurketak berdintasun-desberdintasunak kuantifikatu egiten ditu eta, ondorioz, berdintasun-desberdintasun horiek eskala batean kokatzen ditu. Bestetik, arrazoiketa erabili behar da, datuak zein irizpidek gidaturik bildu diren eta datuen arteko desberdintasunak zein irizpidetan oinarritzen diren zehaztuz.

Horrez gain, metodo zientifikoak bi oinarri nagusi ditu: erreproduzitzeko gaitasuna eta ezeztagarritasuna. Lehenak, analisi bat egiten denean beste edonork analisi bera errepikatzeke gaitasuna ziurtatzen du; hots, analisi bat beste edonork datu berekin eta teknika berak zein beste batzuk erabiliz berregiteko aukera izan behar duela esan nahi du oinarri honek. Bigarrenak, hots, ezeztagarritasunaren oinarriak, edozein proposamen zientifiko faltsutzat jo edo arbuiatua izan daitekeela defenditzen du. Noiz eman liteke hori? Bada, kontraanalisian edo kontrafrogetan berdinak edo antzeko emaitzak lortzen ez direnean.

Zergatik dator hau hizkeren sailkapenari buruzko lan batean? Bada, euskalkien sailkapenean metodo objektiboak erabiltzea aldarrikatzen dugulako. Euskal hizkerak sailkatzea bada kontua, ekin diezaiogun sailkatzeke zientziak garatu dituen tresnak eta irizpideak erabiltzeari. Mekanismo horiek hizkuntzara eta euskarara aplikatuz lortuko da sailkapen zientifikoa edo zientifikoki onartua den sailkapena. Ikerketa honetan metodo objektiboak erabiltzea ondoko hiru ezaugarrietan zehazten da:

- a) Datu-base handiak erabiltzea. Zenbat eta lagin handiagoa izan, orduan eta emaitza fidagarriagoak izango dira. Hori izan da arrazoi bat datu-kopuru handia erabiltzeko.
- b) Irizpide objektiboak erabiltzea ezaugarriak aukeratzean. Datuak aukeratzeko ondoko irizpideak erabili dira:
 - lexikoan, lexiko komuna erabiltzea; irizpide honi jarraituz, bazter utzi dira arrain, txori eta zuhaitzen izenak, adibidez. Hiztun arruntek eguneroko bizitzan erabiltzen dituzten hitzak bildu dira datu-basean;
 - galderak ez ditzala izan erantzunik gabeko herri asko; galderak herri guztien % 5ean erantzunik ez badu, galdera hori ez da kontuan izan.
- c) Estatistika deskribatzailearen sailkapen automatizatuak erabiltzea; datu-basean bildu den informazioa estatistikoki tratatzeko gaur egun zeregin honetarako nazioartean estandarrik diren teknikak erabili dira, hizkerak sailkatzeke *Diatech* eraikitako tresna erabiliz.
- d) Sailkapen teknika desberdinen artean hierarkikoak ere erabiltzea; izan ere, hizkeren arteko loturak hierarkia baten ikusteak beraien arteko loturen egitura erakusten du.

Artikulu honetako 2. atalean erabilitako metodologia inguruko zehaztapenak emango dira. Hurrengoan datuen analisisa egingo da. Laugarrenean hizkeren sailkapenaz arituko gara;

bostgarrenean euskalkien sailkapen berria aurkeztuko da. Azkenik, seigarrenean ondorioak aipatuko dira.

2. Metodologia

Datuak bi iturritatik jaso dira: gehienak Euskaltzaindiaren EHHA egitasmotik; eta egitasmoan bildu ez den informazioa, azentuarena alegia, Gaminderen lanetatik. Hizkerai dagokien sailkapena egiteko hizkuntza atal desberdinetako datuak erabili dira: lexikoari dagozkion 183 kontzeptutako hitzak, izen-morfologiari dagozkion 156 ezaugarri, aditz-morfologiari dagozkion 136 ezaugarri, sintaxiari dagozkion 61 ezaugarri eta fonologiari dagozkion 113 ezaugarri. Guztira 649 ezaugarrietan oinarritzen da hizkeren sailkapen orokorra. Euskal hizkeren sailkapenetan orain artean erabili izan direnetatik oso urrun dago kopuru hau. Hizkeren sailkapen hau orain artekoak baino osoagoa da, ez bakarrik datu-kopuruan, baita hizkuntza ataletan ere. Izan ere, hizkuntza atal guztiak erabili dira sailkapen honetan: lexikoa, izen-morfologia, aditz-morfologia, sintaxia eta fonologia. Izaera honek behar-beharrezko bermea ematen dio lan honi. Ezaugarri hauek, jada ezaguna denez, hizkeretan zehar nola gauzatzen diren zehazteko, euskararen eremuko 145 herritan jasotako erantzunak bildu dira; adibidez, galdera orokorretan morfemarik eransten den ala ez zehazteko (hau da, “al” partikula, “-a” atzizkia ala morfologikoki ezer berezirik ez duten zehazteko) ondoko galdera edo antzekoa egin da herriotan: “zure laguna etorri da?” edo “zure lagunak egin du lana?” eta antzeko besteren bat ere (galderak egitean lexikoan gaia aurkezten euskara erabili den bitartean, gramatika kontuetan erdarazko esaldiak eman izan zaizkie lekukoei eta horiek euskaraz esatea eskatu zaie). Erantzunetan *etorri da*, *etorri al da*, *jin dea* eta antzekoak bildu dira, eremuan eremuko moldeak erabiliz. Basean guztira 100.000 datu baino gehiago bildu dira.

Orain arte egin izan diren sailkapenetan sintaxia nekez agertu izan da, ezaugarriren bat edo beste salbu; lexikoa ere ez da sistematizazio batez erabili, nahiz hitz bat edo beste erabili izan den. Bestalde, gorago aipatu diren 649 ezaugarritan aldaera asko dituztenak eta aldaera gutxi dituztenak bildu dira, baina aipatu behar da hori ez dela izan ezaugarri edo hitz horiek aukeratzeko irizpidea. Ulertzen dugu hizkuntza atalak ez daudela neurri berean ordezkatuak datu-basean: kasu honetan izan dugun muga bakarra EHHA egitasmoa izan da. Ahalegindu gara ahal bezainbat ezaugarri biltzen eta hizkuntza atal bakoitzeko ezaugarriak aukeratzeko eta datu-basea sortzean oinarritzeko irizpidetzat gutxienez 100 ezaugarri biltzea izan da; jakina, sintaxian ez da lortu, hain zuzen ere EHHA egitasmoan ez direlako kopuru horretara heltzen hizkuntza atal honi dagozkion galderak.

Intonazioari dagokion ezaugarririk ez da sartu hizkeren sailkapenerako datu-basean. Ez nahi ez delako izan, EHHako egitasmoan ez delako horren ezaugarririk baizik. EUDIako zenbait ikertzailek intonazioa landuz hizkeren sailkapena egina duten arren (Iglesias & Aurrekoetxea, 2017), ikerketa-metodologiak galarazi du hona ekartzea eta kanpoan gelditu da. Etorkizunean erantsi beharrekoa da.

Datu hauek datu-base batean eratu dira lehenik. Eraiki den datu-base orokorraren ezaugarri nagusiak hauek dira:

- a) Abiapuntua EHHAn jasotako erantzunak idazkera arautuan emanak dira; baina datu-base orokorrean ez da erantzunik eman. Hauek lematizatu egin dira hizkuntza atalkako datu-basean eta hemen datu lematizatuak bakarrik erabili dira.
- b) Atzizkiak (menpeko esaldietan, adb.) edo adizki konposatuak erabili izan direnetan (aditzari dagozkion datuetan) ez dira lexemak edo aditz nagusiak erabili.
- c) Erantzunak eta proposamen onartuak (herrikotzat joak) maila berean tratatu dira; uste dugu metodologia batek ahalbidetu dituen hitzak eta formak lekukoak beretzat hartzen dituenean, nahiz iradokiak izan, herrikotzat jo behar direla.
- d) Erantzunetan agertzen diren desberdintasun fonetiko edo ebakerazkoak izan dira kontuan (*ditut*, *dittut*, *ittut*, adibidez, hirurak onartu dira).
- e) Testuinguruak eragindako aldaketa fonetikoak ez dira kontuan hartu (hauen artean agertzen da *-m* bukaerakoa); hots, honelako kasu guztiak *-n* bihurtu dira.
- f) Azentua ez da kontuan izan: azentuatu gabeko eta azentuatutako formak berdintzat hartu dira. Hori orokorrean, morfologia, sintaxia eta lexikoa aztertu izan direnean, hala izan den arren, fonologia sailean azentuari buruzko 16 ezaugarri erabili dira, hain zuzen ere aurretik ikerketa-talde honetako partaide batzuek (Aurrekoetxea, Gaminde, Iglesias & Gandarias, 2012, 2013, 2014 eta 2015) landutakoa.
- g) Erantzun anitzak: herri berean galdera batentzat erantzun bi edo gehiago izatea.

Analisi hau egiteak denbora eta lana eskatu dio lantaldeari. Lehen datu-basea 2014an sortu zen, izen-morfologiari buruzkoa eraiki zenean; eta harrezkero urtero-urtero datu-base berriak sortzeko lanak ez du etenik izan, hizkeren sailkapen orokorra egiteko datu-base guztiak bildu arte.

Analisiak egiteko *Diatech* tresna informatikoa erabili da (<http://eudia.ehu.es/diatech/index>). Oraindik guztiz amaitu gabe badago ere (gaztelaniaz bakarrik dago *interfacea*, adb.), datu-baseen kudeaketa eta analisi estatistikoak egiteko gai da. Azterketa honetarako *Cluster* edo multzokatze-analisi hierarkikoa aukeratu da hasteko, eta algoritmo desberdinak erabili dira: *Ward*, *Average* eta *Complete*. Hizkuntza distantzia unitatea RIV edo IRD da. Multzokatze hierarkikoaren analisiaren emaitzak hiru dira: dendrograma, mapa eta legenda (multzo bakoitzean zenbat hizkera biltzen diren adieraziz). Dendrogramak erakusten du nola dagoen egituratuta hizkeren hierarkia eta multzo bakoitzaren osagaien arteko harremana. Dendrogramaren egituraren arabera egiten da mozketak; hau da, non edo zein mailatan moztu behar den. Bertsio honek marra baten bidez irudikatzen du eta ikertzaileak aukeratzen du 2, 3, 4... multzotan utzi behar den. Mozketa hori mapan irudikatzen da: 3 multzotan moztea erabaki bada, mapan 3 eremu agertuko dira, bakoitza kolore batekin koloreztatuta. Bertsio honek koloreak aldatzeko aukera ematen dio ikertzaileari. Analisia amaitu ostean datuak (dendrograma, mapa eta legenda) irudiformatuan jaitz daitezke.

Lan honek hainbat egonaldi trinko egitera bideratu du lantaldea. Egonaldi berezitu horietan talde-lan trinkoaren helburua datu-base bat sortzea izan da, egonaldi bakoitzean hizkuntza atal bati dagokion datu-basea. Horrela joan dira urtez urte eraikitzen oinarritzko datu-baseak.

Beraz, hemen aurkezten den hizkeren sailkapen orokorra gauzatzea 5 urteko lanaren emaitza da; 2014. urtean ezarri zen lehen harria eta 2019. urteko hasieran ipini da azkena. Tarte honetan atalkako baseak sortuz joan diren neurrian atalkako argitalpenak ere egin dira; hala nola Aurrekoetxea, 2016a,b, 2018, 2019; Videgain & Aurrekoetxea, 2016; Aurrekoetxea, Gaminde, Gandarias & Iglesias, 2012, 2013, 2014, 2015; Aurrekoetxea, Gaminde, Etxebarria & Iglesias, 2012; Aurrekoetxea, Ormaetxea & Videgain, 2018, 2019).

3. Datuen analisisia

Hizkerak edo dialektoak elkarren artean konparatu eta euren arteko desberdintasunak zehazteak dakartzan arazoak presente egon dira dialektologiaren historian zehar. Hizkeren arteko desberdintasunak kuantifikatzeko Jean Séguyk (1971) proposatutako bideak, *dialektometria* izenez ezagutzen denak, abiarazi du, eta gaur egun nazioarteko dialektologian estandar bihurtu diren prozedurak dira. Uste dut ez dagoela lekuz kanpo kapitulu hau L. Mitxelenaren aipamen batekin hastea, zeinetan hizkeren arteko desberdintasunez ari den (1981: 295):

No puedo entrar aquí a discutir hasta qué punto son divergentes los distintos dialectos vascos. Me limitaré, pues, a decir que una consideración científica de este y otros problemas semejantes exige, como requisito indispensable, algún procedimiento operativo de cuantificar las distancias lingüísticas, y remitiré a la propuesta de nuestro malogrado amigo Jean Séguy, (...). A falta de alguna suerte de medida, habrá que referirse a estimaciones del orden de magnitud, hechas a ojo de buen o mal cubero. Así, es posible que esas diferencias le parezcan muy grandes al hablante ordinario no iniciado a fondo más que en el euskera de su círculo local o comarcal, pero, para un comparatista, y este es el punto de vista que hay que adoptar aquí, las divergencias, como ya he escrito en algún otro lugar, son desesperadamente pequeñas.

Hemen apustu argia dago dialektoak sailkatzeko gaur eguneko nazioarteko teknika estandarrak erabiltzeko. Ikerketa-talde honek, sortu zenetik, helburu hori izan du, euskararen bariazioa ikertzean nazioartean erabiltzen diren metodologia eta prozedurak erabiltzea. Gaur egun bada nazioarteko komunitate zientifikoak onartutako prozedura bat: hain zuzen, gorago aipatu diren datu-base handiak erabiltzea da lehena, sailkatze-teknikak erabiltzea da bigarrena, eta abar.

Hizkeren arteko hizkuntza-desberdintasuna prozedura kuantitatiboak erabiliz zehazki neurtzen da, distantzia-unitate zehatz bat erabiliz. DMn nagusiki bi unitate erabiltzen dira: RIV (distantzia kategorikoa) datu lematizatueterako eta Lv (distantzia numerikoa) datu fonetikoetarako (1. taula). Ikertzailearen hautua da: hemen datu-analisisa lematizatuekin egitea erabakienez, RIV distantzia-unitate kategorikoa da egokiena, eta hori erabili da (bien arteko desberdintasunerako ikus taula). Distantzia-unitate hau erabiliz hizkeren arteko distantzien taula ere lortzen da (ikus 1. taula).

	Ahetze	Aia	Aldude	Alkotz	Altzai	Altzürükü	Amezqueta	Andoain	Ariz	Aramaiio
Abaurregaina	40,61	43,99	35,64	33,05	48	47,8	46,59	45,74	35,07	53,69
Ahetze		41,85	26,25	36,92	40,65	41,56	43,3	43,68	35,81	52,92
Aia			46,54	37,74	53,22	54,53	22,51	21,97	41,89	44,14
Aldude				37,92	38,59	39,15	46,96	47,08	35,96	52,14
Alkotz					50,74	49,44	37,94	37,96	23,7	48,96
Altzai						14,17	56,12	54,77	49,98	60,03
Altzürükü							56,51	55,77	48,9	60,31
Amezqueta								21,31	40,05	43,52
Andoain									41,65	44,76
Ariz										50,15

1. taula: hizkuntza-atal guztietako datuetan oinarritutako hizkuntza-desberdintasunak (taula zatia)

Taulan ikus daitekeenez, Abaurregaineko hizkeratik Ahetzekora, adibidez, % 40,61eko hizkuntza-desberdintasuna dago. Hizkera beretik Aiakora % 43,99. Taulan erakusten den zatian desberdintasun handiena Altzürükü eta Aramaioko hizkeren artean kokatzen da (% 60,31); ondoren oso hurbil Altzai eta Aramaiokoaren artean (% 60,03). Hizkuntza-desberdintasun txikiena erakusten den taularen zatian Altzai eta Altzürüküko hizkeren artean biltzen da (% 14,17), ondoren Aia eta Andoaingo hizkeren artean (% 21,97).

Baina bilaketa taula osora hedatuko balitz, Sohüta eta Altzaiko hizkeren artean biltzen da hizkuntza-desberdintasun txikiena (% 13,7); ondoren, Errezil eta Beizamako hizkeren artean (% 13,65). Aldiz, hizkuntza-distantzia handienak (beraz, urrunen diren hizkerak) Lemoiz eta Eskiulako hizkeren artean jaso dira (% 66,56), ondoren Sohüta eta Etxebarriko hizkeren artean (% 66,07). Horrek esan nahi du lan honetarako aukeratu diren hizkuntza-ezaugarri guztiak kontuan hartuz, aztertzen diren euskara tradizionalaren eremu osoko 145 hizkeren arteko distantzia handiena % 66an kokatzen dela eta distantzia txikiena % 13an.

4. Hizkeren sailkapena

Hizkeren arteko desberdintasun kuantitatibo hauek izango dira emango diren urrats berrien oinarri. Kuantifikotasun honen ezaugarriek ziurtatu behar dute ondoko urratsetan lortzen diren emaitzak. Sailkapen orokorrari dagozkion ezaugarrietan oinarrituz, batetik hizkeren sailkapena burutuko da eta, bestetik, sortu diren hiru euskalkien hizkuntza-ezaugarriak aurkeztu dira. Aurrera baino lehen, gogoratu behar da sailkapen hau hierarkikoa dela; hots, metodo honen bidez aztertzen diren 145 hizkerak hierarkia batean kokatzen direla, antzekoen diren hizkeretatik desberdinen diren hizkeretaraino. Sailkapen hierarkikoaren ostean beste mota bateko estatistika teknika erabiliko da, sailkapen hierarkiko horretan lortutako emaitzak egiaztatzeko: eskalatze multidimentsionala, *MDS* laburdurarekin ezagunagoa.

4.1. Multzokatze edo cluster analisia

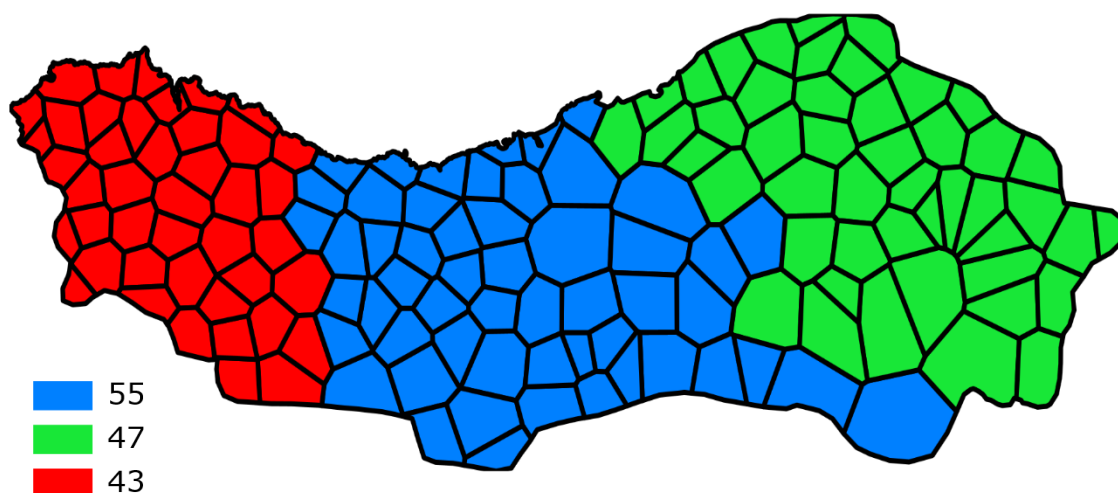
Hizkuntza-ezaugarri guztiak kontuan hartuz, hizkeren sailkapena egiteko multzokatze edo *cluster* analisia deitu teknika kuantitatiboa erabili da. Multzokatze-teknika asko dago, besteak

1. irudia: hizkuntza-ezaugarri guztien dendrograma (*Cluster-Ward* metodoa)

Helburua hizkuntza-ezaugarri guztien datuak erabiliz euskararen eremuan zenbat eremu dialektal agertzen diren izanik, dendrogramak argiki erakusten du hiru eremu nagusi agertzen direla. Hori baieztatzeko aski da dendrogramak dituen hiru adar luzeenak hartzea kontuan. Honek esan nahi du hizkera guztien artean hizkuntza-desberdintasun handiena adar horiek gordetzen dutela. Bariazioaren zein kopuru biltzen duten hiru adar nagusi horiek? Bada, batetik irudiaren goren maila hartu behar da kontuan. Horretarako ezkerrean agertzen den graduazio-eskalari erreparatu behar zaio: azken adar biak eskalaren 8.700 maila inguruan biltzen dira (eskalak 8.000 mailaraino markatzen duen arren, irudia luzeago denez, kalkulua egin behar da). Bestetik, hiru eremu agertzeko (begiratu behealdean agertzen diren herriak hiru kolorez koloreztatuta daudela) eskala 635 maila inguruan moztu da (2.000 maila baino askoz beherago biltzen dira gorri kolorez idatzitako herriak). Hizkuntza-ezaugarri guztiak kontuan izanik, hizkeren artean lortu den bariazio guztia 8.700 mailan biltzen bada eta hizkera horiek hiru eremutan sailkatzeko 635. mailaraino jaitsi garenez, horrek esan nahi du eskalako maila horretan bariazioaren % 92,76 biltzen dela (8.700 mailak % 100 adierazten du eta 635. mailak % 7,30).

Modu honetan, ikerketa honetan erabili diren hizkuntza-ezaugarri guztiei dagozkien datuak erabiliz, hiru eremu dialektaletan bereizten badira, lehena gorritz idatzitako hizkerak osatzen dute. Bigarren multzoan orlegiz idatzitako hizkerak agertzen dira; hauek lehen multzoan bildutakoak baino maila jasoagoan biltzen dira. Azkenik, urdinez idatzitako hizkerak agertzen dira hirugarren multzoan. Hauen biltze-maila gainerako biena baino maila goragoan agertzen da.

Dendrogramaren mozketa eskalako 635. mailan egitea erabaki da eta, ondorioz, hiru multzotan sailkatu dira hizkerak. Maila horretan hiru adar luze agertzen dira; eta hiru adar moztu direnez, hiru eremu dialektaleko mapa eraikiko da. Dendrogramaren mozketa maparatzean kontuan izan behar da mapan marrazten diren eremuak geografikoki koherenteak izan behar direla. Irudikatu den mapan hiru eremuek koherentzia handia erakusten dute (ikus 2. irudia).



2. irudia: hizkuntza-ezaugarri guztiak hartuz eraikitako euskalkien mapa (Ward, RIV, 3 multzo)

Hiru eremuok hiru dialekto irudikatzen dituzte: mendebaldekoa, erdialdekoa eta ekialdekoa (terminologia ez dago finkatua oraindik). Mendebaldekoak, Bizkaiko herri guztiez gain Gipuzkoako mendebaldeko hizkerak (Eibar-50, Bergara-35, Araotz-Oñati-12, Oñati-114, Arrasate-18, Leintz Gatzaga-96), eta Arabako Aramaio-11 hartzen ditu; guztira 43 herri biltzen dira euskalki honetan.

Erdialdeko euskalkiak Gipuzkoako gainerako hizkerak eta Nafarroako guztiak, Luzaide-103 eta Zugarramurdi-145 izan ezik hartzen ditu; guztira 55 herri biltzen dira euskalki honetan. Mendebaldeko euskalkiak bezala, eremu trinkoa osatzen du euskalki honek ere.

Azkenik, ekialdeko euskalkia dugu, zeinetan Iparraldeko hizkera guztiak sartzen diren eta hauekin batera Luzaide eta Zugarramurdi; guztira 47 herri edo hizkera. Oso eremu trinkoa osatzen du euskalki honek ere.

Hiru eremu dialektal, hiru euskalki oso ongi bereizitakoak eta egituraketa geografiko trinko eta orekatua erakusten dutenak. Diogunaz gain, 3 euskalkiek biltzen dituzten eremu geografikoak ere antzekoak dira, mendebaldekoa izanik agian eremu txikiagoa. Hizkeren dagokienetik, 3 euskalkiek antzeko hizkera-kopuruak biltzen dituzte, nahiz erdialdekoa den gehien biltzen dituen.

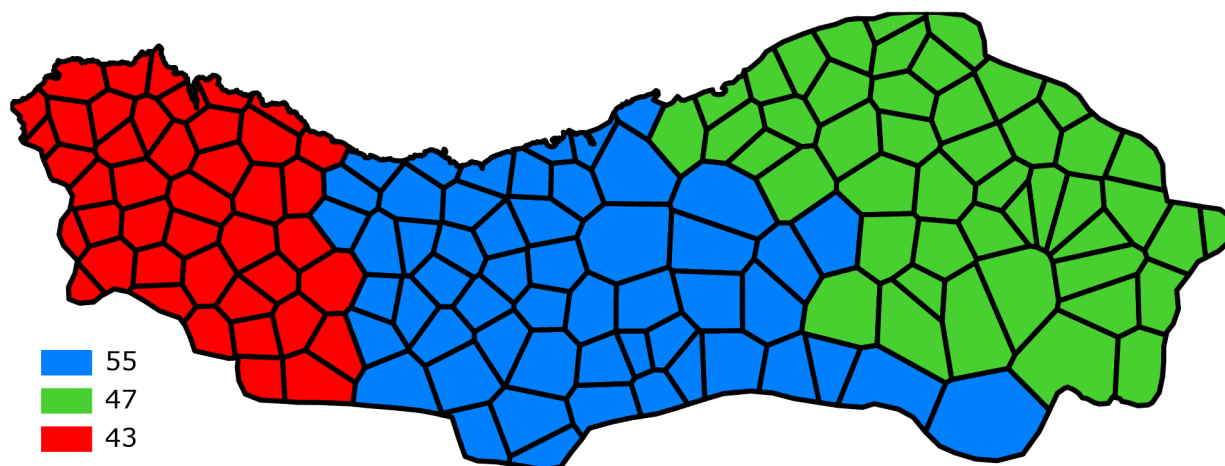
Orain artean inoiz burutu ez den sailkapena dugu hau. Lehen aldia da euskalkiak sailkatzeko ikerketetan 4 gramatika atal (fonologia, izen-morfologia, aditz-morfologia eta sintaxia) eta lexikoari dagozkion ezaugarriak erabiltzen direna. Gainera, inoiz ez bezalako hizkuntza-ezaugarri kopurua erabili da; kapitulu honen hasieran esan bezala, 649 ezaugarri erabili dira lan honetan. Aipatu behar da, halaber, orain artean aztertu diren hizkuntza-alor bakoitzean lortu diren sailkapenen oso-oso antzekoa dela. Honek guztiak egonkortasun eta fidagarritasun handia ematen dio lortu den mapari.

Average algoritmoa

Behin *Ward* algoritmoarekin dendrograma eta mapa lortu ondoren, beste bi algoritmo erabiliko ditugu analisisia egiaztatzeko. *Average* edo *batezbesteko algoritmoa* zertxobait desberdina dela esan da gorago; hizkeren arteko distantziak neurtzean hizkera guztien arteko distantziak hartzen ditu aztergai, eta lortzen diren emaitza guztien batezbestekoa egiten du. Hain desberdinak diren algoritmoen emaitzak bat etortzea ez da ohikoa, hain zuzen ere erabiltzen dituzten prozedurak oso bestelakoak direlako. Algoritmo desberdinetako emaitzak erkatzean helburua ez da emaitza bera lortzea (ia-ia ezinezkoa), desberdintasunak zein diren atzematea baino.

3. irudia: hizkuntza-ezaugarri guztiak hartuz eraikitako euskalkien mapa (*Average*, RIV, 3 multzo)

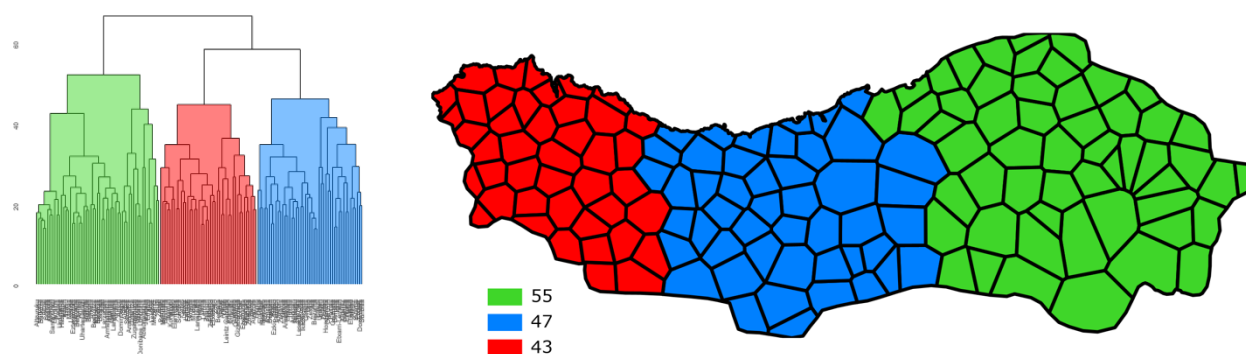
Batezbestekoa algoritmoarekin lortu den mapa (3. irudia) *Ward* algoritmoarekin lortu denarekin (4. irudia) erkatuz gero, konturatuko gara berdin-berdina dela. Hots, bi mapetan



euskalki berak eta euskalki bakoitzean hizkera berak agertzen dira. Hizkerak sailkatzeko metodo desberdin bi erabiliz mapa berdina ateratzeak esan nahi du muga horiek egonkorrak direla.

Complete algoritmoa

Diatech tresnak hirugarren algoritmoa ere inplementatuta duenez, aukera dugu hau ere erabiltzeko eta emaitza gainerako biekkin erkatzeko. Algoritmo honekin aurreko biekkin izandako prozedura eta datu berak erabili dira.

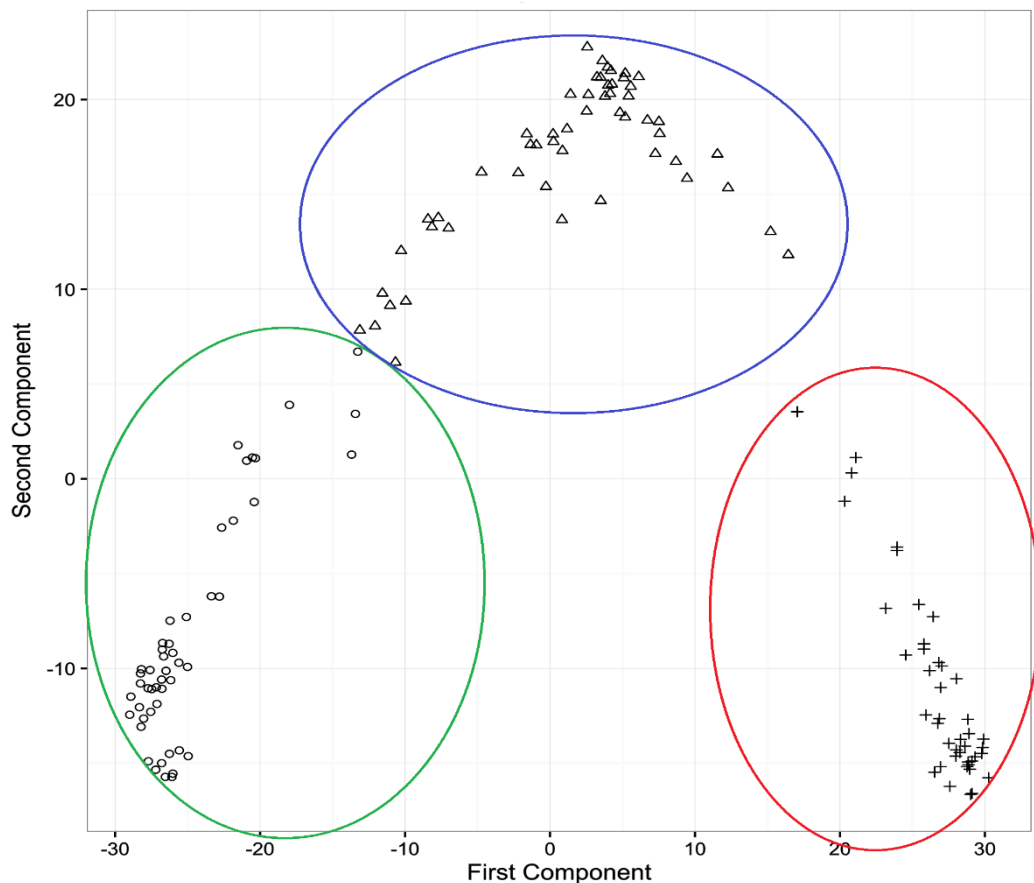


4. irudia: hizkuntza-ezaugarri guztiak hartuz eraikitako euskalkien mapa (*Complete*, RIV, 3 multzo)

Lortu den mapan (4. irudia) hiru euskalki garbi agertzen dira, aurreko bi mapetan bezalaxe; baina desberdintasun bat dago: dendrograman ikusten da mendebaldekoa (gorria) eta erdialdekoa (urdina) lehenago biltzen direla eta gero biltzen dela ekialdekoa (orlegia); aurreko bi mapetako dendrogrametan, aldiz, mendebaldekoa (gorriz) da azken biltzen dena. Horrez gain, bada beste desberdintasun bat, mapak erakusten duena: mendebaldeko eta erdialdeko euskalkien arteko muga bera den arren (mendebaldeko euskalkiak eusten dio bere hizkera kopuruari), erdialdeko euskalkiak ekialderen dauden Nafarroako 8 hizkera galtzen ditu (Jaurrieta, Abaurregaina, Zilbeti, Mezkiritz, Eugi, Erratzu, Lekaroz eta Aniz).

4.2. Fuzzy clusterra

Fuzzy edo multzo lausoen teknika multzokatze ez-hierarkikoa da eta ez ditu hizkerak hierarkia baten barnean sartzen. Bi emaitza grafiko lortzen dira teknika honekin: diagrama eta mapa. Diagramak hizkerak plano desberdinetan ematen ditu: pantailan dugun diagramak hizkera guztiak bi ardatzetan kokaturik erakusten ditu eta beraien artean dauden hizkuntza desberdintasunak distantziarekin irudikatzen ditu. Zenbat eta multzoen artean tarte handiagoa izan orduan eta hizkuntza desberdintasun handiagoa adierazten du (5. irudia).



5. irudia: dispersio diagrama

Teknika hau ona da *cluster* determinista hierarkikoetan jasotzen diren emaitzak baieztatzeko. Gorago esan bezala *cluster* deterministetan ez da hizkeren arteko desberdintasunen maila argiki erakusten; teknika honek, ordea, hori egiten du ondoen. Teknika osatzaileak dira aipatzen ari garen bi hauek.

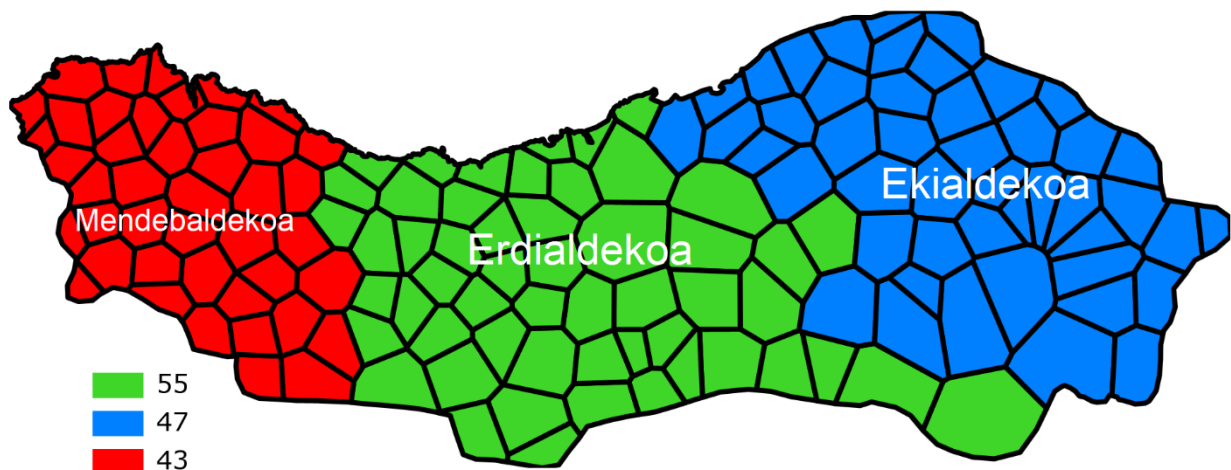
Diagraman ederto agertzen da multzo gorria osatzen duten hizkerak nahiko bananduta aurkitzen direla urdinez bildutakoetatik. Hala ere, multzo horretan berean ere, hizkera batzuk oso hurbil dira (batzuk elkarren artean gainjartzen dira), eta beste batzuei, ordea, urdinez bildu diren hizkeren multzoranzko joera antzematen zaie. Hori horrela izanik ere, multzo kohesionatutzat hartu behar da, bere multzoko hurbilen dituen hizkerak hurbilago dituelako, ondorengo multzoko lehen hizkerak baino.

Ez da hori gertatzen urdinez bildu diren hizkeren multzo eta orlegizkoen artean. Bi multzo hauen artean ez da muga zehatzik marrazten (marra urdin eta orlegiak elkar ukitzen duten gunean badira bi multzoetako hizkerak). Bi multzoek elkar ukitzen duten gunean agertzen diren hizkerak askoz ere hurbilago daude euren artean, bakoitza bere multzoko beste hizkeretatik baino. Hemen argi adierazten da *cluster* hierarkiko-sailkatzaileetan batetik Ward eta *Average* algoritmoekin egindako sailkapena eta mapa eta, bestetik, *Complete* algoritmoarekin egindakoaren artean eremu batetik bestera mugitzen diren hizkeren kasua. Eremu batetik bestera mugitzen diren hizkera horiek beste eremuko hizkerekin harreman ona dute eta hori argiki islatzen du *fuzzy cluster* teknikarekin lortutako dispersio-diagramak.

Cluster hierarkiko-sailkatzaileen eta *fuzzy cluster* teknikak osagarriak dira eta hizkeren arteko hizkuntza hurbiltasunak modu desberdinetan irudikatzen dituzte. Lehenean, hierarkia baten barnean kokatzen dira; bigarrenean, ordea, bi ardatzetako planoan kokatzen dira, hizkuntza hurbiltasuna espazioarekin irudikatuz. Lehenarekin multzoak zurruntasun handiz eraikitzen dira; bigarrenean hizkuntza errealitatean gertatzen diren bariazioa modu bisualago batean eskaintzen da.

5. Euskalkien sailkapen berria

Aurreko analisitik ondorio garbia ateratzen da: euskarak 3 euskalki ditu, ez gehiago, ezta gutxiago ere (6. irudia). Bat mendebaldean; bigarrenak erdialde osoa hartzen du, eta hirugarrenak ekialdea. Euskalkien mapa berria hau da:



6. irudia: Euskalkien mapa berria.

Argumentu bi mapa hau euskalkien mapa izendatzeko:

- Erabilitako *cluster* teknikak erakusten dute hiruko sailkapena dela egokiena. MDS teknikak ere erakusten du hiru multzotan banatzea egokia dela.

- b) Bestalde, eta 1. grafikoa berrartuz, sortu den datu-basean dagoen hizkuntza-desberdintasun guztia dendrogramako eskalan 9.314 mailatan sartzen da. Hori da desberdintasun guztiak metatuta *clusterrean* lortu den kopurua. Dendrogramako hiru lerro handienei erreparatzen bazaie, ikusten da kolore urdinez idatzita dauden hizkerak osatzen duten multzoarena eskalako 1.600 mailan hasten dela. Maila honetan hizkuntza-desberdintasunen % 17a biltzen da.

Euskalkien tasunen inguruan zalantza handiak dauden arren eta eztabaida bere horretan bukatu ez den arren, baieztatu liteke desberdintasunen % 17 baino gutxiago duen eremua ezin dela euskalki izendatu; maila horretatik beherako multzoek azpieuskalki edo beste izendapen bat hartu beharko luketela deritzogu. Egia esan, eztabaida hasi baino ez da egin eta beharko luke aurrerantzean ere ekarpen berriak izatea.

6. Ondorioak

Euskalkien sailkapen berria behar-beharrezkoa izan da euskal dialektologian. Orain artean argitaratutako euskalkien sailkapenak dialektologia tradizionalako metodologia eta prozedurak erabiliz burutu dira. Sailkapen hau, aldiz, dialektologia berriak darabiltzan metodologia eta tresnak baliatuz egin da, nazioarteko dialektologian estandarrak diren dialektoak sailkatzeko teknikak erabiliz.

Horretarako ezinbestekoa izan da EHHA egitasmoa. Egitasmo honek euskararen ohiko eremuari dagozkion datuak bildu izanak ahalbidetu du esku artean duzun hizkeren sailkapen hau. Izan ere, datu dokumentatu eta konstatagarriak ezinbestekoak dira edozein ikerketa-proiektu abian jartzeko. Egitasmo honetan argitaratutako informazioa erabili da; hori gabe ezinezko zatekeen ikerketa honek eskatzen dituen parametroetan aurreratzea.

Gaurko teknikak ahalbidetzen dituzten datu-kopuru handiak erabili dira berton. Ia 100.000 datuko datu-basetik abiatu dira kuantifikazioa eta multzokatze-teknikak. Hain datu-kopuru handiek, batetik, ikerketari sendotasuna ematen diote eta emaitzei ziurtasuna eskaintzen diete; bestetik, ezinezko bihurtzen du ezaugarrietan hautuak egitea; ikertzailearen lanetan sartzen da datuak kudeatzea, baina kudeaketa horretan ez da sartzen ezaugarrien aukeraketa; beraz, subjektibitatek urrunduz objektibotasunean sakontzen da. Hirugarren, datu-kopuru handiak erabiltzeak hizkeren arteko hizkuntza-desberdintasunaren kuantifikazioa tresna estatistikoekin burutu beharra dakar. Tresna hauek edozein zientzian erabiltzen direnak izanik, ez du hizkuntzalaria mesfidantzara bultzatu behar; aitzitik, emaitzen sinesgarritasun eta fidagarritasun maila jasoagoa dutela onartzera baizik.

Hurrengo urratsen artean lehena euskalkien ezaugarriak zehaztea izango da. Ondoren azpieuskalkien zehaztea eta berorien hizkuntza-ezaugarriak lortzea.

7. Erreferentziak

- Aurrekoetxea, G. (2016a). Analysis of the morphological variation of Basque. *Dialectologia et Geolinguistica*, 24, 21-41.
- Aurrekoetxea, G. (2016b). Distantzia geografikoaren eta hizkuntza distantziaren arteko korrelazioa. En G. Aurrekoetxea, J. M. Makazaga & P. Salaberri (arg.), *Txipi Ormaetxea omenduz. Hire bordatxoan*, Bilbo: UPV/EHU, 55-72.
- Aurrekoetxea, G. (2018). Hizkuntza bariazioa eta erantzun anitzak. In A. Etxebarria, A. Iglesias, H. Lejarreta & A. Romero (arg.), *Traineru bete lagun: Iñaki Gaminde omenduz*, Bilbo: UPV/EHU, 171-190.
- Aurrekoetxea, G. (2019). Classification of syntax in Basque. *Language Documentation* 7 (Bolzano 11-13 February, 2019) biltzarrean aurkeztua.
- Aurrekoetxea, G., Gaminde, I., Gandarias, L. & Iglesias, A. (2012). Prosodic variation in the Basque language: stress areas. In X. A. Álvarez Pérez, E. Carrilho & C. Magro (arg.), *Proceedings of the International Symposium on Limits and Areas in Dialectology (LimiAr). Lisbon, 2011*. Lisboa: Centro de Linguística da Universidade de Lisboa, 35-51. [http://limiar.clul.ul.pt/proceedings_en.html]
- Aurrekoetxea, G., Gaminde, I., Etxebarria, A. & Iglesias, A. (2012). Ahozko hiztegia. *Euskalingua*, 21, 20-27.
- Aurrekoetxea, G., Gaminde, I., Iglesias, A. & Gandarias, L. (2013). Prosodic Variation in the Basque Language: Stress Areas. In E. Carrilho, C. Magro & X. A. Álvarez (arg.), *Current Approaches to Limits and Areas in Dialectology*, Cambridge: Cambridge Scholars Publishing, 247-266.
- Aurrekoetxea, G., Gaminde, I., Gandarias, L. & Iglesias, A. (2014). Variación prosódica en vasco: áreas acentuales. En In Y. Congosto, M. L. Montero & A. Salvador (arg.), *Fonética experimental, educación superior e investigación. III. Prosodia*, Madrid: Arco Libros, 11-28.
- Aurrekoetxea, G., Gaminde, I., Gandarias, L. & Iglesias, A. (2015). Euskararen azentuera: hizkera tradizionaletatik euskara estandarera. In M. J. Ezeizabarrena & R. Gómez (arg.), *Eridenen du zerzaz kontenta. Sailkideen omenaldia Henrike Knörr irakasleari (1947-2008)*. Bilbo: UPV/EHU, 71-96.
- Aurrekoetxea, G., Ormaetxea, Tx. & Videgain, X. (2018). Euskalkien sailkapen zientifikoa: lexikoa. In L. Unamuno, A. Romero, A. Etxebarria & A. Iglesias (arg.), *Linguistic Variation in the Basque and Education-III / Euskararen bariazioa eta bariazioaren irakaskuntza-III*, Bilbo: UPV/EHU, 126-140.
- Aurrekoetxea, G., Ormaetxea, Tx. & Videgain, X. (2019). Euskalkien sailkapen zientifikoa: lexikoa (2). In A. Iglesias, N. Eguskiza & L. Unamuno (arg.), *Linguistic Variation in the Basque and Education-IV / Euskararen bariazioa eta bariazioaren irakaskuntza-IV*, Bilbo: UPV/EHU, 41-59.
- Iglesias, A. & Aurrekoetxea, G. (2017). Intonazioaren egituraketa geolinguistikoa. In A. Etxebarria & N. Eguskiza (arg.), *Bariazioa esaldien intonazioan*. Bilbo: UPV/EHU, 277-295.
- Mitxelena, L., 1981, “Lengua común y dialectos vascos”, ASJU XV, 291-313 [orain *Palabras y textos* liburuan 35-55 orrialdeetan ere].
- Séguy, J., 1971, “La dialectométrie dans l'Atlas linguistique de la Gascogne”, *RLiR* 37, 1-24.
- Videgain, X. & Aurrekoetxea, G. (2016). Euskararen bariazio geo-morfologikoaren azterketa. In A. Iglesias, A. Romero, A. Ensunza (arg.), *Linguistic variation in the basque language and education - II / Euskararen bariazioa eta bariazioaren irakaskuntza -II*, Bilbo: UPV/EHU, 25-36

MELODIC ANALYSIS OF SPEECH (MAS): PHONETICS OF INTONATION

Francisco José Cantero Serena, Dolors Font-Rotchés
Universitat de Barcelona
cantero@ub.edu; dolorsfont@ub.edu

Abstract

The acoustic-perception-based Melodic Analysis of Speech method (MAS) is a formal, objective and complete method that encompasses the most relevant aspects to be considered when analysing intonation. It provides researchers who wish to study intonation with criteria for establishing a corpus, the identification of melodic units, the extraction and relevance of acoustic data, its standardisation and the representation and interpretation of graphs, as well as the execution of perception tests and validation of the results obtained. It also offers a way of interpreting the melodic data based on three levels of intonation: one which provides an explanation of the structure of speech, the prelinguistic level, the linguistic level and the paralinguistic level, giving it significance with regard to the speaker's intentions or the communicative context. In addition, a phonological description based on the Autosegmental-Metrical model (AM) and its labelling convention known as Tones and Break Indices (ToBI), which is complementary and totally compatible with the phonetic description of MAS, can be added.

The method is constantly advancing, being reformulated and added to as new requirements arise resulting from its use by researchers in the description of different languages, interlanguages and dialects. It is these developments, which have occurred over more than a decade and particularly in the last two years, together with the basic pillars of the method, which are presented in this article.

Keywords: Melodic Analysis of Speech method (MAS), Intonation, Acoustic analysis, Perception tests, semi-automatic analysis.

1 Introduction

This paper presents the most relevant characteristics of the Melodic Analysis of Speech (MAS) method, together with the developments that have taken place to date¹. It was proposed by Cantero in 1999, published in 2002, and since then has been used by various researchers, initially to describe languages and interlanguages and later dialect variants and aspects of paralinguistic intonation (in relation to the speaker's intentions and the communicative context). Currently, there are open lines of research on description and analysis relating to different aspects of intonation in different languages.

Traditionally, intonation has been one of the aspects of speech characterisation that has been greatly ignored by linguists. The first people to show an interest in describing intonation were authors from the *British school*. Authors such as Jones (1909), Palmer (1922) and Armstrong

¹ This paper is part of the research project financed by the Ministry of Economy and Competitiveness *Analysis of Speech and Teaching Models*, ref. FFI2013-41915-P.

& Ward (1926) aimed to write manuals designed for the correct teaching of English. In their work, they provided melodic curves, their context and meaning, so that learners could imitate them. In later decades, other linguists followed this tradition, including O'Connor & Arnold (1961), Crystal (1969), Cruttenden (1986) and the American, Bolinger (1986, 1989). Close to this tradition, although with original contributions that have greatly influenced intonation studies for the Spanish language, Navarro Tomás published his manual for Spanish intonation in 1944, designed as a practical language manual and a theoretical treatise on Spanish intonation.

Also based on the phonetic characterisation of intonation, authors of the Dutch school, t'Hart, Collier & Cohen (1990) and their *IPO model* (Institute for Perception Research) are of note, as they provide a stylisation of the intonation curves through the selection of relevant pitch movements in the melodic curve, eliminating other pitch variations.

At the same time as these schools, from a structural perspective of the phenomenon and a phonological viewpoint, linguists of the *American tradition* emerged with their level analysis proposal; authors such as Pike (1945) and Trager & Smith (1951), among others, described the curve based on four pitch levels, four stress levels (primary, secondary, etc.) and junctures (marking the beginning and ending of the phrase).

Along this baseline of phonology, and from the theoretical approach of generative phonology, stems the *Autosegmental Metrical model* (Pierrehumbert, 1980, 1987), which, based on the metrical structure of the phrase, suggests the representation of its pitch variations using two levels: H (high) and L (low). For this model, during the 1990s, the ToBI (Tone and Break Indices) transcription system was developed for English (Silverman et al., 1992; Beckman & Hirschberg, 1994), as well as systems for other languages, including Italian (ToBI_{It}, Avesani, 1995), German (GToBI, Grice et al., 1996) and Spanish (Sp_ToBI, Beckman et al., 2002; Prieto & Roseano (eds.), 2010), among others, and a transcription variant for Dutch (ToDI, Gussenhoven et al., 1999).

Later, and influenced by the Autosegmental Metrical model, the Aix-en-Provence model emerged, also generative, which suggested obtaining phonological forms of the prosodic system of different languages. To do so, it follows an automatic styling process using the MOMEL (MOdelling MELody) algorithm and the INTSINT notation model (Hirst et al., 2000).

In this context, the *Melodic Analysis of Speech* (MAS) model aims to provide a robust approach for the acoustic analysis of intonation, regardless of its phonological interpretation. It is a formal and objective method that provides the researcher with a way of analysing intonation, described in detail in the protocol established in 2009 by Cantero & Font-Rotchés² and with a way of interpreting the data from three different levels of intonation: prelinguistic, linguistic and paralinguistic (Cantero, 2002).

² There is an English version in Font-Rotchés & Cantero (2009).

This protocol is divided into three sections: a) the identification of melodic units in a corpus of spontaneous speech; b) the development of an acoustic phase, including the determination of relevant frequency values and standardisation; and lastly, c) a perception phase, during which the analyses and interpretation of data is validated. It remains the same for the procedure described below, although with some new contributions resulting from research, such as matters regarding labelling or the relevance of a number of values in the analysis that have proven to be important for some levels of intonation. Other significant developments include the creation of the semi-automated script by Mateo (2010a), which has also evolved over the years, adapting to the requirements of the group's research.

With regard to levels of intonation in the prelinguistic level (melodic features that structure the discourse and delimit it in comprehensible units), motivated by the development of several studies on dialect variants and interlanguages, the new term *melodic profile* has been coined and defined, as a series of melodic features that characterise an *accent* (such as dialect or foreign accent). On a paralinguistic level, three types of intonation have been defined: emotional, focus and (im)politeness. And finally, given that some melodic features simultaneously inform us of different characteristics (dialect or interlanguage, linguistic or pertaining to the speaker's intentions or the communicative context), the term *melodic syncretism* has been coined, which will be developed further in future research.

Therefore, the following sections explain the steps to be followed for applying the method, which include all of the developments made to date and the interpretation of melodic data, as well as a final discussion chapter where current research being made is described and some possible additions to the method in the near future are set forth.

2. Identification of melodic units

2.1. Phonic hierarchy

Sounds of speech do not occur indiscriminately but are organised in phonic blocks around accents, which are the true nuclei of discourse and have a hierarchical relationship, called *phonic hierarchy* (Cantero, 2002), shown schematically in Figure 1.

According to this hierarchical relationship, consonant sounds are organised around vowel sounds creating syllables and these join to form words where stress is the most relevant component (e.g. the syllable *par-* in *parque* 'park'). This stress is not only combined with the unstressed syllables of the word but also by adjacent unstressed words, such as prepositions, articles, conjunctions, unstressed pronouns, etc. to create *phonic words* or *rhythmic groups* (e.g. *por el parque* 'through the park', where *par-* is the nucleus).

Phonic words or rhythmic groups also form blocks around the accent, which stands out and is higher in the hierarchy. These create the *phonic group* or "the major phonic unit that segments the discourse and the only one that has its own identity and is not conditional upon other units" (Cantero, 2002: 83). The accent or nucleus of the phonic group, which is the most

relevant, is normally found on the last phonic word (e.g. the syllable *par-* of the phonic group *Ayer paseamos por el parque* ‘Yesterday we walked through the park’).

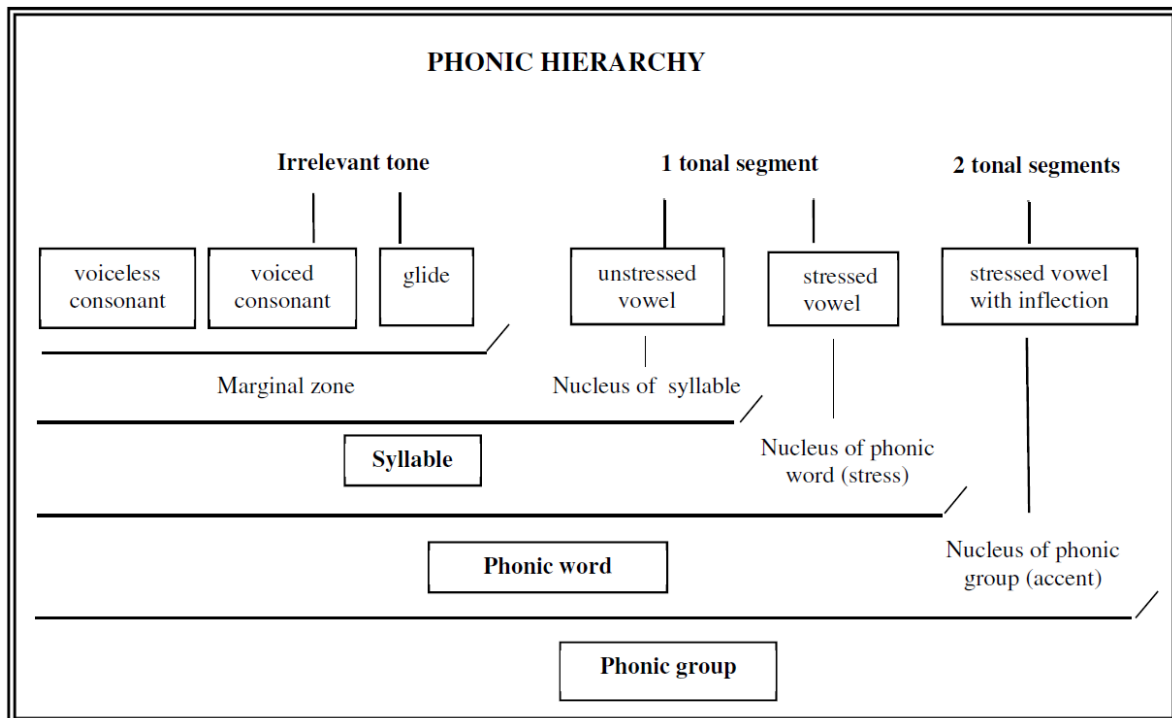


Figure 1. Diagram of phonic hierarchy.

The way of organising speech or phonic hierarchy comprises the personality of each language beyond the mere pronunciation of sounds and has two functions. On the one hand, it integrates the sounds organised hierarchically and forming phonic blocks (syllables, rhythmic groups and phonic groups) by the speaker and, on the other, it delimits these phonic blocks and allows recognition of the words in each block facilitating comprehension of the discourse by the listener.

2.2. *Melodic units*

The phonic group, which we have defined as the major phonic unit of discourse, where sounds of the language are organised hierarchically, contains a coherent lexical-grammatical structure together with a melody. This melody, which represents a significant melodic unit in the intonation of speech, is called the *intonation contour*.

The smallest melodic unit of the intonation contour is the *tonal segment*, in other words, each one of the more or less stable and perceptibly clear phases, which tend to coincide with a *mora* (equivalent to the duration of a short vowel). In this regard, a vowel typically lasts a mora and therefore contains a tonal segment, whilst a long vowel, such as the tonic vowel where the nucleus of the phonic group occurs, tends to last for two moras and as a result contains two tonal segments with different tonal values. This succession of two adjacent and different tonal segments is what we call *tonal inflection*.

The different values of the tonal segments that comprise the melody of the intonation contour follow the hierarchy of the phonic group. Therefore, the only tonally relevant segments are the vowels (these comprise an identifiable tonal segment in the melody of the contour) and, particularly, the stressed vowels (nuclei of rhythmic groups), which stand out among those are unstressed because they present a tonal contrast. Unlike vowels, voiceless consonants are not sounds but noises because they lack phonation. Voiced consonants or glides, despite being considered sounds, tend to be marginal because they always occur associated with a vowel.

Among the different nuclei of rhythmic groups that form a phonic group there is one that stands out for containing the most relevant accent or *nucleus*, melodically characterised for including a tonal inflection with two tonal segments and usually found in the last rhythmic group. Similarly, the first tonic vowel of the contour or *first peak*, is particularly relevant.

Following this hierarchy of the phonic group, which contains the intonation contour, we can see, in Figure 2, that its structure is divided into three parts: the *anacrusis*, the *body* and the *final inflection (FI)*, limited by two accents, the *first peak (1st peak)* and the *nucleus*.

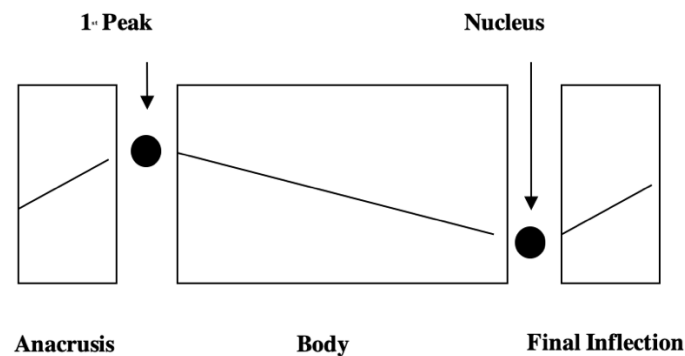


Figure 2. Diagram of the three parts of the contour.

The *anacrusis* is formed by the rising tonal segments prior to the first stressed vowel in the contour, called the *first peak*. Not all of the contours have anacrusis as on occasion it is non-existent. The *body* is the syllables from the first peak to the last stressed vowel in the contour, called the *nucleus*, which consists of a gradual tonal fall of the different segments (although in some emphatic utterances it can be rising) and the point from which the *final inflection* begins. The description of these elements allows us to define the contour melody (especially the description of the *final inflection* and its tonal movements, falling, rising, flat, rising-falling, among others).

To obtain the melodic units or intonation contours of a spontaneous speech corpus, initially it is advisable to choose the utterances that coincide with a turn to speak in the dialogue (normally short and easy to identify) until the researcher has acquired enough confidence in identifying phonic groups and in defining speech melodies. However, it is not always possible to select a turn from the dialogue. For this reason, to identify phonic groups we use the nucleus of the phonic group as a formal criterion for segmenting speech. This nucleus is the

accent that has a tonal inflection and is the most significant part of the contour, usually occurring on the last phonic word.

2.3. Establishing the corpus

To establish a corpus for the purpose of describing a language or a dialect variant, we must obtain models of spontaneous speech that provide an intonation reality, in other words, fragments of speech, uttered by speakers who produce them in a real context. Therefore, the corpus must be obtained outside the laboratory and without any prepared or induced written or spoken support from the researcher. It must also include a significant number of speakers, who vary in age, gender and social-cultural background, who have no speech impediments and who are native speakers of the language.

In certain types of TV programmes such as talk shows, street interviews, debates, discussions and game shows, among others, we have found a resource that meets these basic requirements: multiple native and anonymous speakers of a language who spontaneously express themselves in a context of dialogue. It is worth mentioning that when selecting them we avoid professional speakers and all TV show hosts who follow a script or who are reading. This ensures that the utterances of the corpus are produced by speakers in a spontaneous context.

Another point to be considered in the selection of the utterances is that, in addition to being produced naturally in a real conversation, there must be no overlapping of voices or background noise or music that impede analysis. First interventions by speakers are also disregarded as these could be prepared, selecting those expressed in an uninhibited way and with no form of prior preparation.

Each utterance chosen must be coded, which involves noting down the transcription, its location on the recording, identifying data of the television programme (as well as the time and day it was broadcasted) and, where appropriate, the type of recording, a description of the conversational context in which the utterance was produced and information about the speaker. Once the utterances have been selected and coded, the audio is extracted from the video using suitable software. It is then segmented in Praat (Boersma & Weenink, 2019) and given a code. A maximum of ten utterances per speaker is established so that the corpus is as varied as possible.

The corpus of Peninsular Spanish created by Alfonso (2010), which does not distinguish the origin of the speakers and includes a total of 90 utterances produced by 37 speakers, meets these requirements. Later, corpora were created by Ballesteros (2011) and Mateo (2014), each including samples from five different regions: Asturias, Navarre, the Basque Country, Castile-Leon and Madrid in the case of Ballesteros and Andalusia, the Canary Islands, Castile-La Mancha, Extremadura and Murcia in the case of Mateo. In total the two corpora include 373 hours of recordings, 2941 utterances and 814 speakers (Cantero, 2016). These corpora are the ones used to characterise the linguistic intonation of spontaneous speech in

Cantero et al. (2005), Cantero & Font Rotchés (2007) and Font-Rotchés & Mateo (2013), as well as the dialect intonation in Ballesteros (2011) and in Mateo (2014).

Also of note is the Catalan corpus by Font-Rotchés (2007), which includes 46 hours of recordings, 580 utterances and 160 speakers. This corpus, apart from being useful for describing the linguistic intonation of Catalan and emphatic intonation (Font-Rotchés, 2011), has also been used to describe the intonation of (im)politeness (Devís & Cantero, 2014). Following this same protocol, corpora have also been created of spontaneous speech of German (Torregrosa, 2015), Brazilian Portuguese (Sena, 2013; Araújo, 2014) and Chinese (Kao, 2011).

Other types of corpus also exist that have different objectives, such as the description of an interlanguage, which implies creating a corpus following different criteria to those indicated for spontaneous speech. In studies carried out by our group on interlanguage, corpora have been created using other strategies. In research on Spanish spoken by Taiwanese (Liu, 2005), Hungarians (Baditzné, 2012), Brazilian Portuguese (Fonseca, 2013) and Swedish (Martorell, 2010), conversations between interlanguage speakers have been recorded using microphones outside a laboratory environment. This involved Spanish learning classroom activities and other different situations, focusing the interest on the content: personal experiences, opinions of their relationship with the Spanish language and culture or hot topics in current affairs, among others, with the aim of obtaining conversations that were as spontaneous as possible.

Lastly, with regard to describing models of speech by television news readers, advertising actors and current affairs programme presenters, who produce a very formal and media-oriented model of speech from a written text, a corpus has been created taken from recordings of television broadcasts of each one of these genres (see Font-Rotchés & Paloma, 2012; and Torrent, 2015 for Catalan; and Font-Rotchés & Machuca, 2010 for Spanish).

3. Acoustic phase

3.1. Determination of relevant frequency values

Once the utterances have been separated, we establish the F_0 value of each vowel, of each tonal segment. This involves differentiating the relevant frequency values (the tonal segments) from the irrelevant values (the F_0 of the sonorous consonants and of the glides). This is done using Praat analysis and synthesis software (Boersma & Weenink, 2019), which is a reliable acoustic analysis instrument.

We identify the vowels (guided by a sonogram) and establish their value (the average of the F_0 values of the vowel). Figure 3 shows an example of the extraction of frequency values of the utterance *Tiene usted hijos?* ‘Do you have children?’ obtained by using the Praat application. The dark areas of the spectrogram correspond to each one of the vowel segments (or diphthongs) and for each one of them the F_0 value in Hz is obtained from the average value of the vowel’s F_0 .

When a stressed vowel contains a tonal inflection with a tonal distance of at least 10%³ (100% is equivalent to an octave of the musical scale), a value of two tonal segments must be established (see syllable *-hi-* in Figure 3 and values in Table 5). This constitutes the inflection or three segments if it is a circumflex inflection (see syllable *-xò* in Table 2), which can be found in any part of the contour. These values are calculated from the initial and final stable values or from the extreme values of the inflection (if there is no tonal stability).

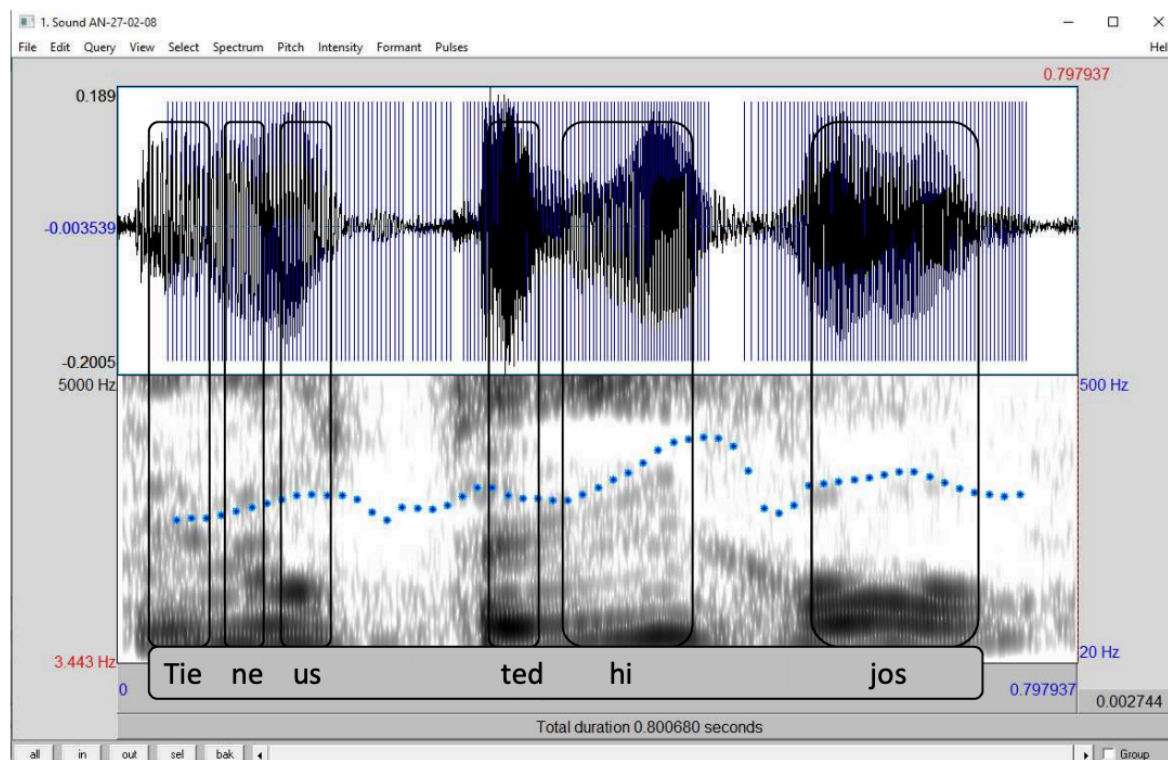


Figure 3. Extraction of frequency values of the utterance *¿Tiene usted hijos?* ‘Do you have children?’

Furthermore, when the inflection ends with a vowel followed by a voiced consonant (specially a nasal, a lateral or a vibrant) with a tonal distance of at least 10% (see an example in Table 6), this consonant usually represents the last tonal segment of the inflection (see perception tests in Font-Rotchés, 2007).

Initially, to describe linguistic intonation we only used very marked tonal inflection data, with more than 20% of movement. However, for research on dialects and interlanguages the need to use inflection data with a tonal distance of 10% for intonation relevance became evident. It is in these small features where some of the characteristics lie that allow us to describe and distinguish one dialect from another (see Ballesteros, 2011; Mateo, 2014) or to describe the transfer features of the interlanguage from the language of origin, as is the case of Brazilian Portuguese (Fonseca, 2013) and Italians (Devís, 2011b) speaking Spanish.

With regard to the extraction of tonal values for the final inflection, in other words, the inflection that occurs in the last tonic syllable of the contour, we can see that there are small

³ This tonal distance has been verified in different perception tests carried out by the Applied Phonetics Lab of the University of Barcelona and applied in MAS research since 2011.

variations according to whether the word is oxytonic, paroxytonic or proparoxytonic. If the word is oxytonic, it may be that the accented syllable has a tonal inflection with two values — we distinguish the second with one asterisk(*)—; this is the case of the syllable *Stadt* 'city' in German, found in the utterance *Schon viel Spass gehabt in der Stadt?* 'Did you have a good time in the city?' (see Table 1). When this syllable has only one value, the final inflection would begin on the preceding syllable.

tim	<i>Schon</i>	<i>viel</i>	<i>Spass</i>	<i>(ge)hab</i> <i>t</i>	<i>in</i>	<i>der</i>	<i>Stadt</i>	<i>Stadt*</i>
Hertz	149	172	190	176	164	146	169	195

Table 1. Example of German whose last stressed vowel has two values⁴.

Table 2 shows the Catalan example of *Bé, crisi és això!* 'Well, crisis is that!' where the last stressed vowel, -ò, has a tonal inflection with three values distinguished by asterisks (*) — one if it is the second value and two if it is the third.

Utterance	<i>Bé</i>	<i>cri</i>	<i>si</i>	<i>és</i>	<i>ai</i>	<i>xò</i>	<i>xò*</i>	<i>xò**</i>
Hertz	328	347	400	365	340	316	426	292

Table 2. Example of Catalan whose last stressed vowel has three values.

On other occasions, the final inflection begins with a stressed vowel and ends with an unstressed vowel, for example, in final inflections that coincide with a word with the stress on the penultimate or antepenultimate syllable. See Table 3, the Catalan example of *Portes vambes?* 'Do you wear sneakers?', where the stress is on the penultimate syllable, and Table 4, the German example of *Kommt das aus Amerika?* 'Is this from America?', where the stress is on the antepenultimate syllable.

Utterance	<i>Por</i>	<i>tes</i>	<i>vam</i>	<i>bes</i>
Hertz	103	97	110	208

⁴ Certain elements of the spoken chain are shown in brackets, such as *(ge)* of *(ge)habt*, to point out that these elements are elided or assimilated and that they do not have an acoustic correspondence in spectrographic analysis.

Table 3. Example of Catalan in which the final inflection takes place in a paroxytonic word.

Utterance	<i>Kommt</i>	<i>das</i>	<i>aus</i>	<i>A</i>	<i>me</i>	<i>ri</i>	<i>ka</i>
Hertz	259	323	269	234	217	296	443

Table 4. Example of German in which the final inflection takes place in a proparoxytonic word.

It is also possible for the accent to be found on the penultimate syllable and that this syllable or the following unstressed syllable could have two values, therefore giving us a circumflex final tonal inflection with three values. This is the case of the Spanish example in Table 5 of *¿Tiene usted hijos?* 'Do you have children?' in which the final inflection has three values: the first and the second in the stressed vowel, *hi-* (293 Hz and 376 Hz), and the third in the next unstressed syllable: *-jos* (305Hz).

Utterance	<i>Tie</i>	<i>ne</i>	<i>us</i>	<i>ted</i>	<i>hi</i>	<i>hi*</i>	<i>jos</i>
Hertz	256	270	289	295	293	376	305

Table 5. Example of Spanish of a circumflex final tonal inflection.

There is another Spanish example in Table 6 *¿Tú tienes internet, Carmen?* 'Do you have internet, Carmen?' in which the final inflection also has three values: the first in the stressed vowel, *Car-*, and the second and the third in the next unstressed syllables: *-men* (123 Hz) from the vowel and *-men** (159 Hz) from the voiced consonant *-n*.

Utterance	<i>Tú</i>	<i>tie</i>	<i>nes</i>	<i>in</i>	<i>ter</i>	<i>net</i>	<i>Car</i>	<i>men</i>	<i>men*</i>
Hertz	234	244	210	198	189	178	171	123	159

Table 6. Example of Spanish of a circumflex final tonal inflection.

3.2. Standardisation of the frequency data and graphic representation

The absolute values obtained in this analysis phase (extraction and determination of F_0 of the tonal segments, measured in Hertz) do not yet constitute the contour melody because they still have the individual phonic characteristics of each speaker, and therefore comparable generalizations can not be extracted. In this sense, the melody uttered by a man that constitutes the succession of absolute values 100Hz–200Hz is not the same as one of values 200Hz-300Hz, uttered by a woman, although the difference is identical: 100Hz in both cases. In relative terms, the tonal interval is different. In the first case, there is a difference of 100% but in the second case, there is a difference of 50%.

Therefore, this data must be relativised, by transforming these frequency values into percentage data that reflect the tonal relationship existing between one absolute value and the

one immediately after it. The formula for the relativisation of the frequency values⁵ is as follows:

$$(1) \text{ Relativisation formula: } REL_{V_{n+1}} = [(ABS_{V_{n+1}} - ABS_{V_n}) / ABS_{V_n}] * 100$$

The variable ABS_{V_n} corresponds to the absolute value of the vowel segment of reference (in Table 7, 259 is the value in Hertz from the first syllable, *Kommt*). $ABS_{V_{n+1}}$ is the frequency value of the next vowel segment at the reference value and from which the relative difference in % with respect to the preceding vowel segment is determined (in Table 7, 323 is the value in Hertz of the next vowel segment). $REL_{V_{n+1}}$ or the relative value expresses the melodic difference as a percentage between the absolute value of reference and $ABS_{V_{n+1}}$ or the value immediately after.

Starting with the arbitrary value 100%, we obtain the relative values of each contour that make up the algorithm, which is the expression of the melody. Table 7 shows the percentages obtained: 100%, 25%, -17%, -13%, -7%, 36% 50% (the rise is expressed as a positive percentage and the fall as a negative percentage). Some authors express this interval using the semitone (st) as a unit, but we consider the use of percentages to be more intuitive.

Utterance	Hertz	Percentage	Standard Curve
<i>Kommt</i>	259	100%	100
<i>das</i>	323	25%	125
<i>aus</i>	269	-17%	104
<i>A</i>	234	-13%	90
<i>me</i>	217	-7%	84
<i>ri</i>	296	36%	114
<i>ka</i>	443	50%	171

Table 7. Absolute (Hertz), Relative (Percentage) and Standard values (Standard Curve) of the contour.

To draw the graphic representation of these melodies, we convert the percentages obtained into standard values: starting, for example, with the value 100 (an arbitrary value). By applying (see Standard Curve on Table 7) the rise percentage seen in the second segment, 25%, we will obtain a second value, 125; to this we apply the fall percentage of the third segment, -17%, to obtain the third value, 104; and so on until the last segment. The formula for standardising the relative values is as follows:

$$(2) \text{ Standardisation formula: } ST_{V_{n+1}} = ST_{V_n} + [(REL_{V_{n+1}} * ST_{V_n}) / 100]$$

⁵Drawn from Torregrosa and Font-Rotchés (2017:1318).

$ST_{v_{n+1}}$ corresponds to the standardised value of $REL_{v_{n+1}}$ with respect to ST_{v_n} . This last value is the standard value that acts as reference value. Using the standard values of each utterance, we can plot the graphic representation of each contour in Excel to compare them.

In Figure 4, showing an utterance produced by a female speaker, the dotted line represents the absolute values in Hertz, with a very marked rise at the end, whilst the lower solid line represents the standard values with a very gentle rise. In contrast, in Figure 5, uttered by a male speaker, once the standardisation has been done, the absolute and standard values are very similar.

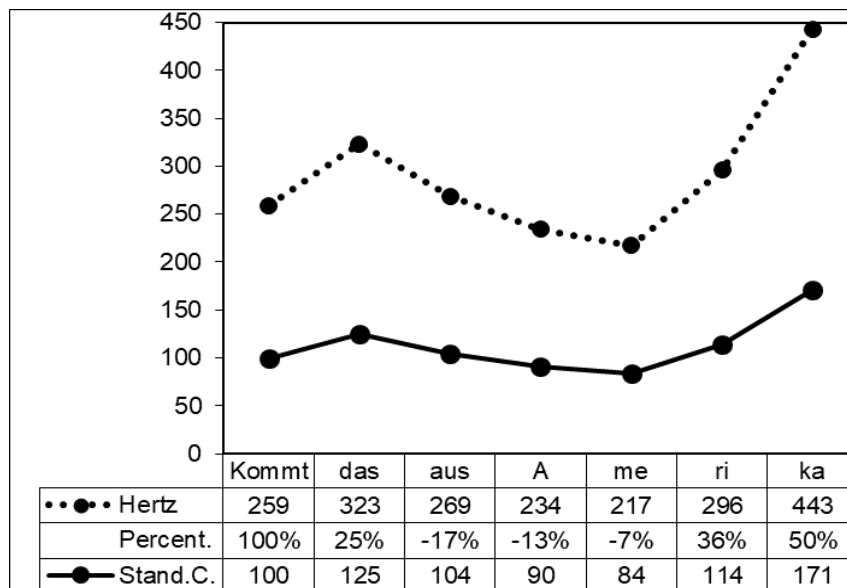


Figure 4. Absolute and standard values of the melody *Kommt das aus Amerika?* ‘Is this from America?’

If we compare the absolute values of both graphs (dotted lines) we can conclude that they have a very similar final inflection; the rise that begins on the last tonic syllable, *-me* and *-bo(ren)*, respectively, and that this is the most significant part of the contour. In contrast, if we compare the standard values (solid lines) we can see that both questions have a final inflection with a very different rise: 104% (from the distance between *-me-*, 217, and *-ka*, 443) in Figure 4 and 148% in Figure 5.

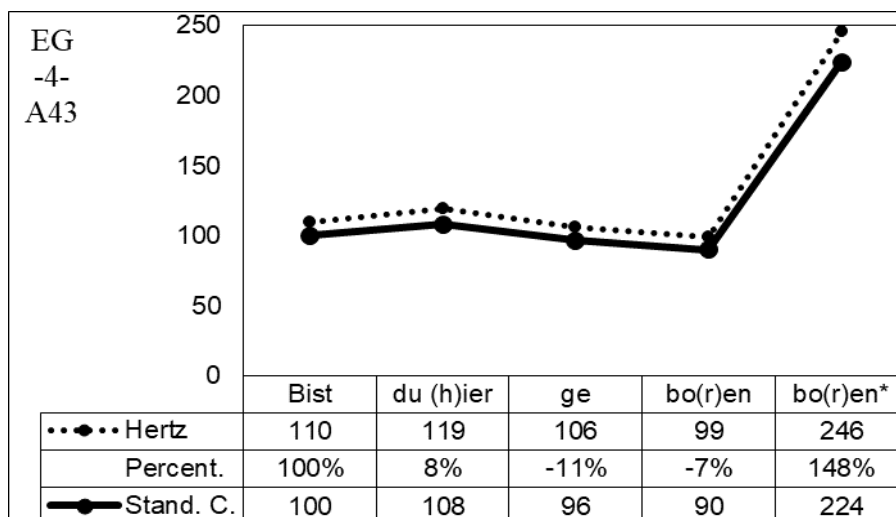


Figure 5. Absolute and standard values of the melody *Bist du hier geboren?* ‘Were you born here?’

With the application of this first phase of the method, or acoustic phase, we obtain the contour standardisation, which can now be compared and classified, regardless of the age, gender or any other characteristic of the speaker, as all the micromelodic variations have been removed and the values have been standardised. This process does not standardise the length of the different contours but isolates and replaces them with the number of tonal segments and their relative value, which is the only relevant data in a melodic analysis. The differences between two contours of different length should be understood from a melodic perspective, regardless of whether the melody contains a word or many words, whether it has a longer or shorter utterance, etc.

3.3. Praat Script

This entire acoustic phase process discussed in sections 3.1 and 3.2, which includes the extraction and standardisation of the data and the creation of graphs, is done manually. To speed up the process, Mateo (2010a, 2010b) designed a script for Praat so that this could be done semi-automatically.

To apply the script for Praat, first the corpus must be established, which can be in any language, dialect variety or register, taken from recordings of short and coded utterances.

Once the script has been installed in Praat (refer to Mateo, 2010b for instructions), the tonal segments for each audio file are established and labelled based on the creation of a TextGrid in Praat. This is a key point in the process as it requires precision and criteria in acoustic analysis. The work must be done using the expanded sonogram (selecting the *Sel* option) and in *Pitch settings* the option of speckles as drawing method must be enabled so that the limits of each tonal segment can be established with the greatest possible precision. At the same time the sound wave of the oscillogram must be considered, revealing the moments of maximum intensity that correspond to the tonal segments, in other words, to the vowels.

To create the labelling in TextGrid (see Figure 6), before the first segment of the utterance we must establish if it is a male (m) or female (f) voice. We can then begin to segment the first vowel of the utterance, transcribing the first syllable. The aim is to search for the stable values of the vowel and mark the limits, which must be located after the first speckle of the vowel and before the last one, to ensure that the values are correct. To mark the limits of a syllable when it is a tonal inflection, either because the vowel has two or three values or because the syllable ends in a voiced consonant, we use the same criteria because the script will use the two, or where appropriate, three values.

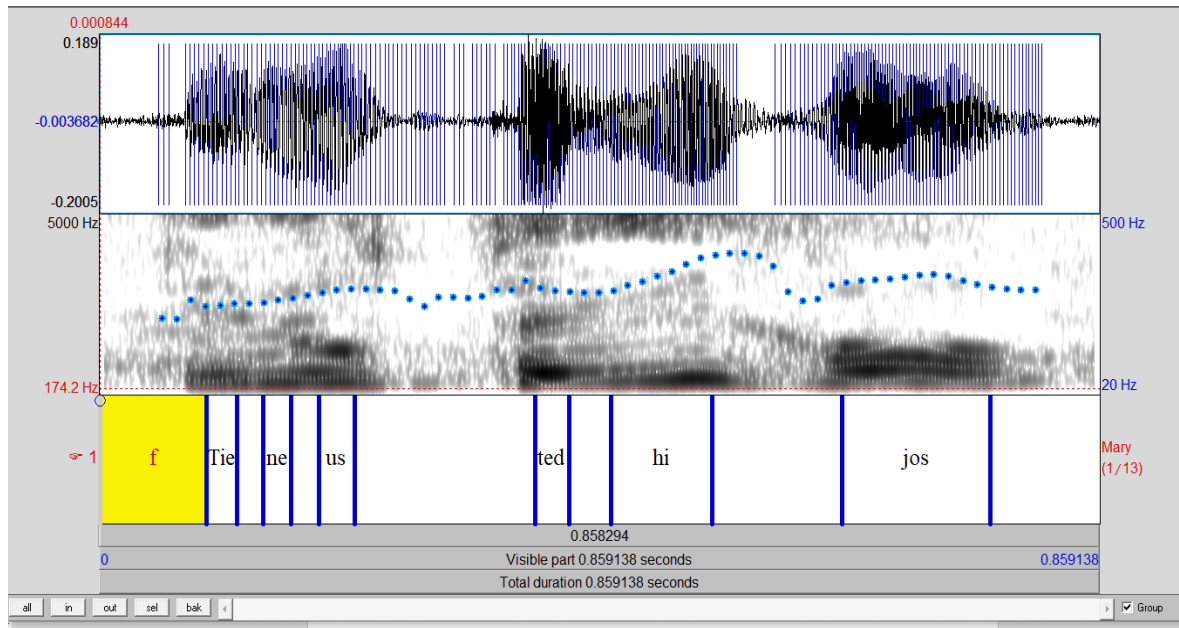


Figure 6. Sonogram and TextGrid of the utterance in Spanish *¿Tiene usted hijos?* ‘Do you have children?’

Thus, in Figure 6 we can see that the utterance is spoken by a woman (f) and that the first syllable is *Tie-*. We have marked the limits taking into account the stable values of the diphthong *-ie-* and the point of intensity on the oscillogram, avoiding using values belonging to the next voiced consonant, *n-*, which is marginal. The second segment is the *-e-* of *-ne*. The voiced nasal *-n-* prior to the vowel is not a relevant value, as previously mentioned. Hence, the limits of the vowel are established respecting the same criteria. The same process is carried out for the other segments. The penultimate vowel *-i* has a tonal inflection with two values; similarly, we establish the limits at the beginning and the end of the vowel and the script will extract the two values.

The next step involves extracting the data, using the audio and the TextGrid of each one of the utterances. This is done by enabling the first of the scripts. This step must include a revision of warnings, such as values for segments that Praat cannot obtain or segments with three values that must be verified, among others, all of which will be flagged by the script. The subsequent modifications to the data, which may be necessary, must also be revised.

Once the correct data has been guaranteed, the standard curve of the utterance is created by enabling the second of the scripts. The graphs of the standard curve are generated using a

macro in an Excel document provided by the script. Once the graphs have been created, comparisons can be made and the intonation models and patterns established⁶. See Figure 7, showing a graph obtained by applying the script.

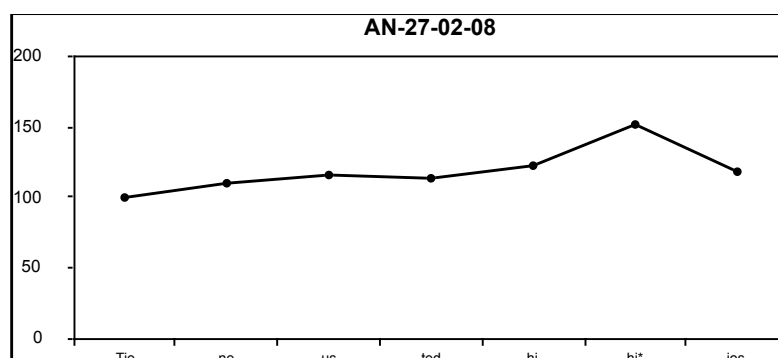


Figure 7. Intonation curve of the Spanish utterance *¿Tiene usted hijos?* ‘Do you have children?’

In short, the Praat script is a reliable and simple tool that can be used to automate the data collection process for the analysis of intonation in any language, saving researchers a considerable amount of time. However, before using the tool, acoustic analysis should be practised manually on a significant number of utterances to obtain criteria and to understand its characterisation.

After analysing all the contours and presenting them graphically, they must be classified into different groups according to the path of the final inflection (rising, falling, rising-falling, falling-rising, among others). For example, the representation of the two German contours (Figure 4 and Figure 5) with a rising final inflection, shows that both belong to the same melodic class: contours with rising final inflection. In some cases, the melodic features of the body and the first peak must be taken into account, such as the Spanish interrogative pattern, the rising body and final inflection pattern (Font-Rotchés & Mateo, 2013; 2018). Subsequently each group of contours must be organised according to the percentage of tonal movement in the final inflection, from greater to lesser, in order to apply the perception phase.

4. Perception phase

4.1. Analysis validation

In the perception phase, first the standard values of an intonation contour are validated in the voice synthesizer by ensuring that it has the same melody as the original, but with a different voice tessitura. We use the option "To manipulate" to open each contour in Praat to obtain this synthesised copy. After opening a contour ready for manipulation (See Figure 9), we erase all

⁶ For more detailed information about the scripts, see Mateo (2010a).

original data of the contour and replace it with our standardised data (See Figure 10)⁷. We can check the validation of our results with a series of perception tests submitted to the listener's judgement from an exact copy (thanks to voice synthesis) of the analysed sentences.

This validation of the melodies obtained during the perceptive phase of the analysis allows us to be completely sure that the analysis has been done correctly and that the obtained data are objective.

It also allows us to check that the validated melodies have been reduced to their most basic expression. Standardisation of the melodic curve provides us with the melody itself, regardless of the speaker's tessitura and clean of any superfluous micromelodic variations. This method is very similar to *close copy* by authors from the Dutch school ('t Hart, Collier & Cohen, 1990).

Following is an example, taken from research that defined the melodic patterns of interrogative intonation in Brazilian Portuguese (Cantero & Font-Rotchés, 2013). The utterance *Duzentos cinquenta os dois?* 'Two hundred and fifty for the two?', where the speaker is giving the price for two calves, revealed the following analysis (see Figure 8).

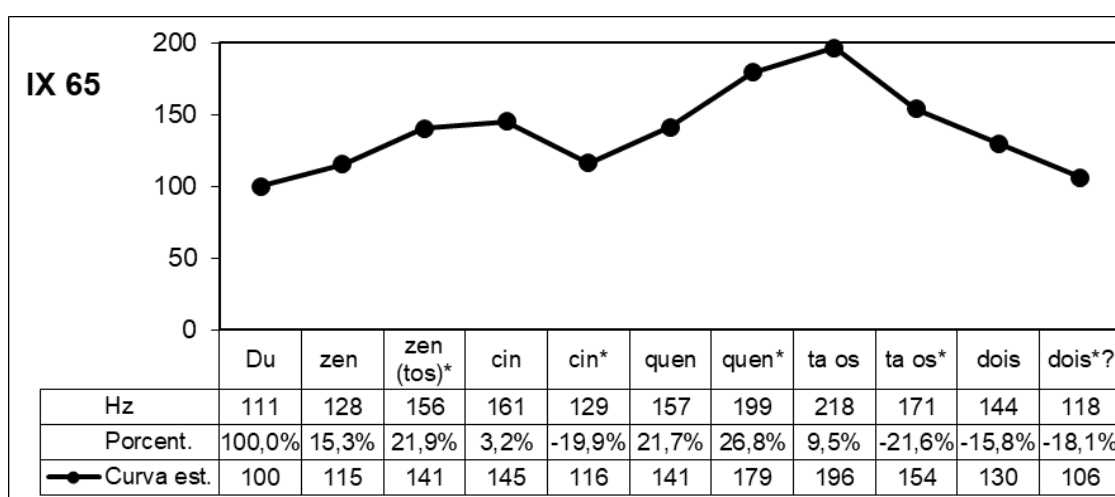
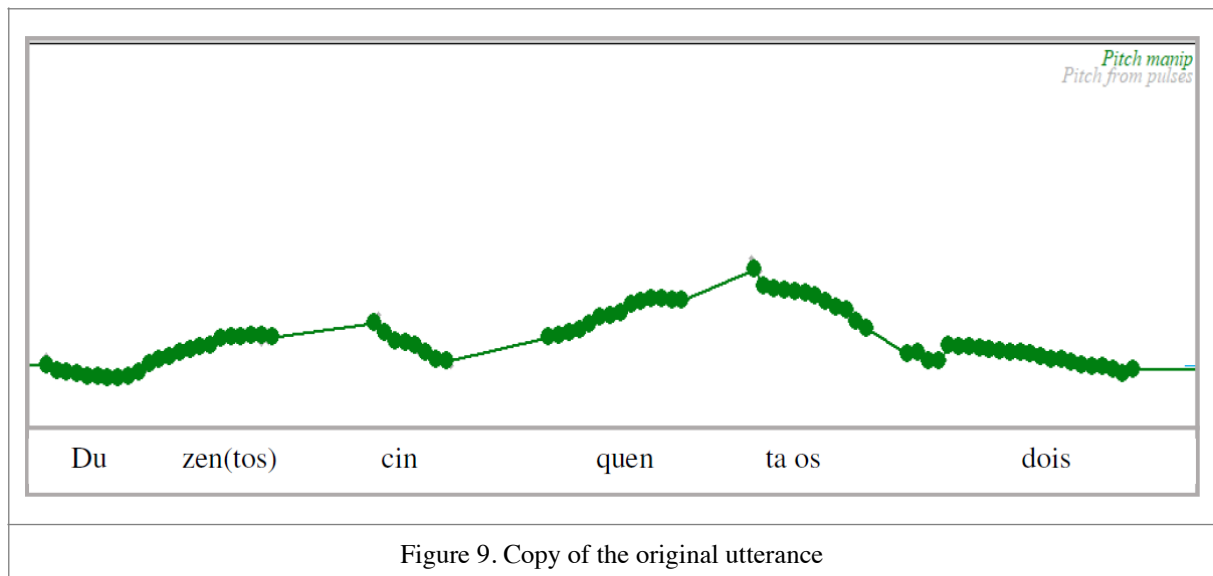


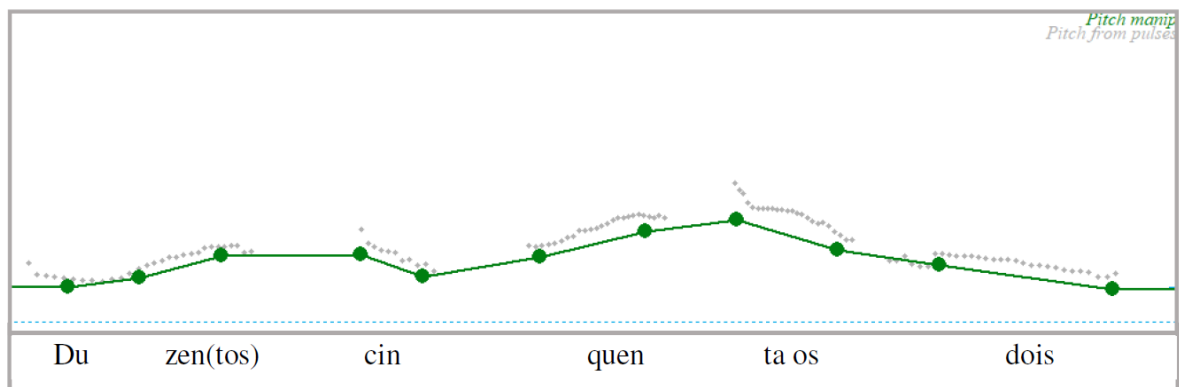
Figure 8. Intonation curve of the Brazilian Portuguese utterance *Duzentos cinquenta os dois?* 'Two hundred and fifty for the two?'

During the Manipulation process in Praat, the copy of the values is obtained (see Figure 9):

⁷ To erase the original data, we must select it, click on 'pitch' and then 'remove pitch points' at the top of the screen. We can also replace it with our standardised data adding a dot where each vowel is located by clicking 'pitch' and 'add pitch point at cursor'.



To this copy the standardised values provided by the analysis are transferred (see Figure 10).



We can now record the synthesis of the result and perceptively validate the melody.

4.2. Identification of the melodic features

Once the melody of the intonation contours, obtained during the acoustic phase, has been validated, we can begin to identify its melodic features: the first peak, body or declination and the final inflection.

Here it is important to distinguish between the different function that the contour's melodic features have in the three levels of intonation analysis:

- Prelinguistic intonation
- Linguistic intonation
- Paralinguistic intonation

Prelinguistic intonation refers to the melodic features that allow the discourse to be structured and integrated in comprehensible units. The *melodic profile* (Cantero & Devís, 2011) of a

language and of a dialect consists in this set of melodic features that characterise an *accent* (such as the dialect or foreign accent). The melodic profile of standard Peninsular Spanish, for example, coincides with the pattern of the intonation contour shown in Figure 2:

- a high first peak, whose centre is normally the first tonic vowel of the group (or the first post-tonic vowel);
- a body or declination that falls evenly, with gradual inflections that systematically begin on a tonic vowel;
- a final inflection that starts on the last tonic vowel of the group and can be falling (with values often greater than -15%) or rising (with values that can reach more than +120%).

Linguistic intonation refers to the level of intonation analysis that allows units that are different on a phonological level to be defined. For example, neutral intonation of interrogative, emphatic or suspended intonations, whose specific manifestations are the *melodic patterns*. In other words, the melodies commonly used by native speakers of a language in their communicative interchanges and which can become models of production and perception.

Paralinguistic intonation refers to the melodic features that provide a meaning related to the speaker's intentions or to the communicative context, but which do not form linguistically coded units (although they can form semi-stable codes: see Cantero (2014)). We consider paralinguistic intonation to have a role in at least three well-defined areas:

- *Emotional* intonation, including the expression of speaker's personal characteristics, his or her emotions and affection.
- *Focus* intonation, which centres on the actual melody of the utterance drawing attention to it, focusing on it or on a part of it to underline its relevance in the discourse. A good example is the intonation of irony.
- *Politeness* intonation, which is focussed on the speaker and calls upon him or her to intensify or attenuate the effect of the message and, in general, to look for confrontation or cooperation.

In any event, there are very few melodic features of an intonation contour, therefore these functions are assumed by the same melodic features simultaneously. This is what we call *melodic syncretism*. One same final inflection, which marks the end of the utterance and its *linguistic* form (for example, an interrogation), can at the same time mark the speaker's dialect accent (*prelinguistic*), express his or her concern (*emotional*), focus on the question so that in this context it is clearly an ironic question (*focus*) and reveal impoliteness to the listener (*politeness*).

4.3. *Forming the hypotheses*

To establish the generalisations that allow us to understand how speech melodies work, we use those obtained during the acoustic phase, which are validated perceptively, identifying and subsequently isolating the melodic features we wish to test. Only a perception test can guarantee which specific melodic feature is responsible for which prelinguistic, linguistic or paralinguistic, communication or discourse function.

For example, to determine the feature of the melody that is responsible for giving the utterance an interrogative value, we can isolate each one of the melodic values found in the analysis of corpus utterances that have been validated as melodies that confer an interrogative value. If we confirm that the melodic feature common to all of them is a rising final inflection, we can determine that this feature is the one we must test in the experiment. Therefore, based on the values that the melodic analysis provides for the questions analysed, we can formulate a hypothesis, such as: *a rising final inflection conveys an interrogative value as of a specific percentage rise*. This hypothesis must be counter-verified in the perception test.

4.4. *Perception tests*

Once the melodic feature to be tested has been identified, it is modified using Praat's synthesis routine (*manipulation*). Using the previous example, we generate various identical utterances where only the value of the rise in the final inflection is modified.

We can modify an utterance in the corpus directly, as long as there are no grammatical, lexical, contextual parameters, etc. that allow the sense or function of the melodic feature hypothesised. For example, we would not use phrases with an interrogative particle, however, we could choose a neutral utterance or create one ourselves by copying the standardised data and modifying only the feature to be tested (in this case, the final inflection). Therefore, the modified melodic feature becomes the variable we wish to counter-verify in the perception test.

This test involves asking a significant group of listeners to react to the manipulated utterances: various unmanipulated utterances, mixed with the manipulated ones (for example, with rises of 20%, 30%, 40%, 50%, etc. up to 120%) and with other non-experimental utterances, in different perception sessions, with various groups of listeners. The results allow us to determine the percentage of rise that led listeners to the conclusion that the utterance is interrogative.

This procedure allowed the melodic patterns of Peninsular Spanish (Cantero & Font-Rotchés, 2007) and Catalan (Font-Rotchés, 2007) to be established for spontaneous speech.

In Figure 11 we can see how the utterance used in the example in section 4.1 (in Figure 8, 9 and 10) is modified:

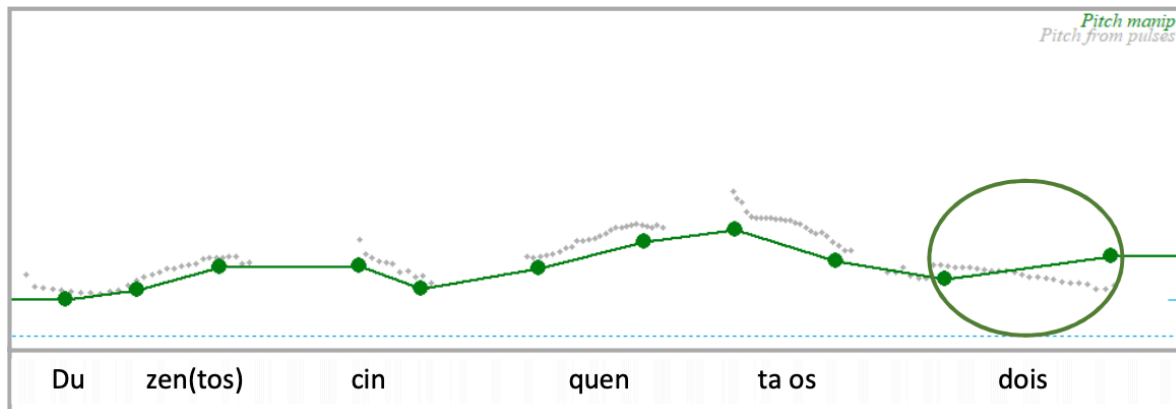


Figure 11. Synthetized copy obtained after modifying the last value (in the circle).

In the original final inflection of the utterance there is a fall of 18.1% (Figure 8). In the standardised copy (Figure 10) we manipulate the last value, *dois** and move from 106 to 158, creating a rise of 21.5% (Figure 11). In this example, the final rise by 21.5% did not allow us to identify the utterance as interrogative (as expected) but did as of a rise of 30% (and up to 50%). In Brazilian Portuguese, therefore, we can determine final rises with very slight interrogative values, which in Spanish or in Catalan are never understood as interrogatives: in Spanish, only as of 70% and in Catalan as of 80% (Cantero & Font-Rotchés, 2013).

5. Discussion and concluding thoughts

The theoretical assumptions of our method are not phonological, but exclusively phonetic, acoustic and perceptive. We determine the tonal segments that form the contour, we standardise their frequency values to express the melody and we validate them perceptively; we identify the melodic features that give the utterance different values (linguistic, pragmatic, discourse...) and we subject them to tests to determine them; we work with tonal intervals expressed as percentages (equivalent to semitones, but more easily and more intuitively handled).

Therefore, it is an independent method, which is at the same time compatible with any phonological focus (structural, notional, autosegmental, etc.), particularly with methods of phonological labelling such as ToBI. The melodies resulting from the analysis have the advantage of allowing any speaker to be compared, of handling large quantities of data and of establishing reliable statistics and precise dispersion margins thanks to the exact quantification of the intervals.

This makes it a solid method: it allows the analysis of spontaneous speech, it requires no preliminary alignments and works well with all types of corpus. It is also objective: the perception test ensures that all of the standardised values and resulting melodies are reliable.

Compared to other intonation analysis models, MAS allows:

- Study of intonation from exclusively phonetic levels of analysis, which do not require the use of other levels of linguistic or grammatical analysis, nor do they depend on them.
- Work with a formal analysis method, based on the acoustic and perceptive analysis of speech. This analysis provides a very detailed description of the intonation from a strictly phonetic viewpoint, regardless of its phonological interpretation.
- Characterisation not only of standard melodies of a language or dialect (whose phonetic generalisation is possible in the perceptive, experimental phase of the method) but also of the individual variants of a single speaker and even the characteristics of pathological speech.
- Work with spontaneous speech: although any type of corpus can be used in the analysis, the solidity of the method allows large volumes of data and a wide range of anonymous speakers to be handled.
- Use of the open phonetic analysis method, allowing any type of notation, analysis or subsequent transcription: pragmatic, grammatical, lexical, as well as any type of phonological labelling. For example, the Autosegmental-Metrical model (AM), which has a labelling convention known as *Tones and Break Indices* (ToBI), is a phonological labelling method for intonation, not a phonetic analysis method. In this sense, it is completely compatible with the MAS method. See two examples in Figures 12 and 13 which demonstrate compatibility.

According to the MAS method, the melodic contour of Spanish ¿*Tiene usted hijos?* (see Figure 12) is characterized by a flat body without anacrusis and first peak and a rising-falling final inflection. This flat body is phonologically transcribed as L* in the AM model because it is phonetically realized as a low plateau at the minimum of the speaker's range.

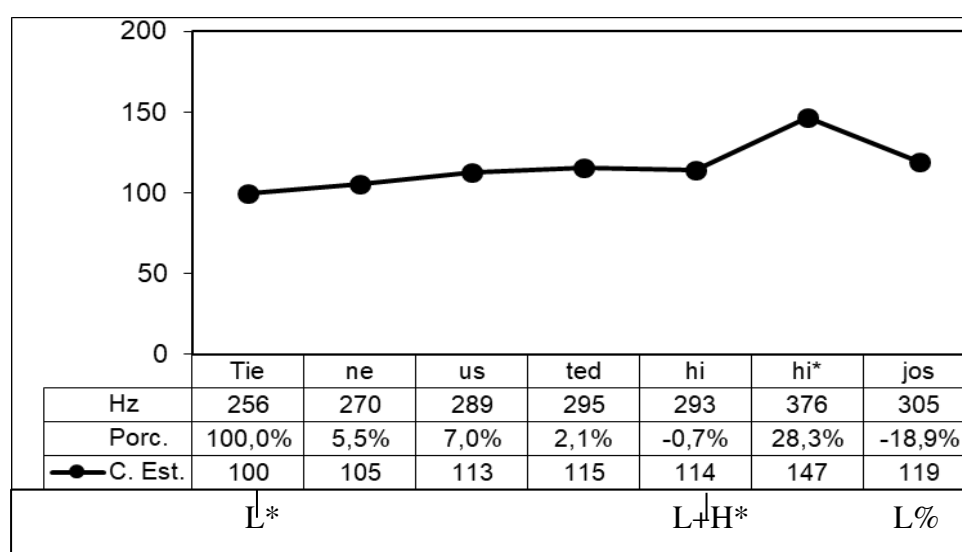


Figure 12. Contour phonetically analysed with the MAS method and labelled with the Sp ToBI system.

The melody of the final inflection has a tonal rise of 28.3% from the first value of the nucleus, *hi-*, to the second, *hi-** and a tonal fall of 18.9% from the syllable *hi-** to *-jos*, which are produced with an L+H* nuclear accent followed by a boundary tone, L%. As we have seen in Figure 12, we can label the melodic contour according to Sp ToBI based on Beckman et al. (2002) and later revised by Face & Prieto (2007) and Estebas-Vilaplana & Prieto (2008) and is totally compatible.

We can see another example in Figure 13. The melodic contour of the question *¿Usted es de Gran Canaria?* ('Are you from Gran Canaria?') starts with an anacrusis with a tonal rise of 68% until the first peak, *es*, which is displaced in the unstressed segment after the first stressed segment of the contour, *-ted*. After that, the melodic line falls abruptly until the nucleus, *-na-* and ends with a falling final inflection of 14.6%. These phonetic features can be complemented with a phonological description in the AM model: an L+ <H* prenuclear pitch accent, which involves a rise with a displaced F₀ peak, an HL* nuclear accent, realized as a F₀ fall within the accented syllable, and an L% boundary tone.

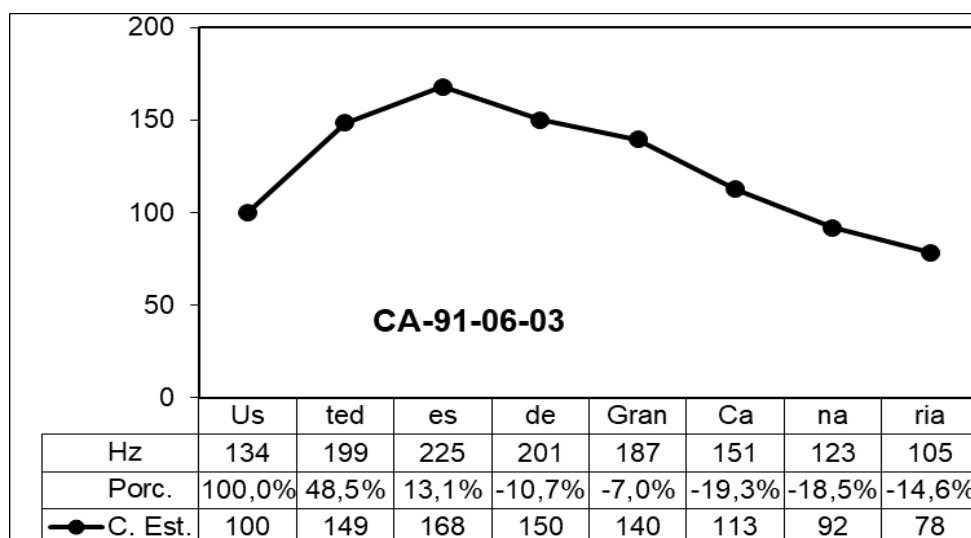


Figure 13. Contour phonetically analysed with the MAS method and labelled with the Sp ToBI system.

In short, MAS is not an alternative method for the analysis of intonation but is simply *the acoustic analysis of the melody*, the phonetic analysis method for intonation that can be complemented with a phonological description.

In fact, carrying out a detailed acoustic analysis of the melodic features of speech is a laborious task and this is undoubtedly the main disadvantage of the method. For this reason, we suggest simplifying the analysis process using the Praat script described in section 3.3, which we continue to optimise to facilitate the task. Many authors have preferred to limit themselves to an intuitive approach to the melodic curve, expressing only the pitch levels that may have a phonological relevance, the assumed stresses and the limits of the utterance. This

is a reasonable method when there is no need for a detailed analysis of the melodies but simply their phonological expression. However, just as the acoustic analysis of the segments cannot be replaced by their simple transcription, on occasion there is a need for a detailed melodic analysis. For example, to characterise the intonation of foreign speakers of a language, to describe dialectal intonations, to distinguish personal characteristics of a speaker, or to specify the melodic features that have pragmatic or discourse functions, etc. In all of these cases, where there is a need to go beyond the previously deduced phonological characterisation, a detailed melodic analysis must be made. This is also the case when the researcher wishes to establish linguistic generalisations based on the mass analysis of new acoustic data.

Since it was proposed (Cantero, 1999) and to date, the method has been used with success in different studies, by our research team⁸ and by different international research groups (particularly South American).

At our research centre, the University of Barcelona's Applied Phonetics Laboratory⁹, we have created a wide corpus of spontaneous speech for Peninsular Spanish and Catalan, among other languages (see section 2.3).

Research on the following areas has been carried out:

Prelinguistic intonation: dialect accents of Peninsular Spanish (Ballesteros, 2011; Ballesteros & Font-Rotchés, 2019; Mateo, 2014) and Catalan (Pons, 2007); as well as for various interlanguages (foreign accents): the intonation of Spanish spoken by Swedish (Martorell, 2010), Hungarian (Baditzné, 2012), Italian (Devís, 2011b), Brazilian (Fonseca, 2013) and Chinese (Cortés, 2004; Liu, 2005) speakers; and the study of speech demarcation units for Spanish (Cabedo, 2013; Martínez, 2015).

Linguistic intonation: melodic patterns and dispersion margins for Peninsular Spanish (Cantero & Font-Rotchés, 2007), Cuban Spanish (García Riverón et al. 2010), Catalan (Font-Rotchés, 2007), Brazilian Portuguese (Araújo, 2014; Cantero & Font-Rotchés, 2013; Sena, 2013; Silva, 2009), German (Torregrosa, 2015; Torregrosa & Font-Rotchés, 2017) and Chinese (Kao, 2011).

Paralinguistic intonation: focusing on politeness and impoliteness for Spanish (Devís, 2011a) and Catalan (Devís & Cantero, 2014); mechanisms of focus in Peninsular Spanish (Cantero et al., 2005), South American Spanish (Pereira, 2014; Ruiz, 2011), Catalan (Font-Rotchés, 2011, 2014) and of professional speakers (Font-Rotchés & Machuca, 2010; Font-Rotchés & Paloma, 2012; Torrent, 2015); and an initial approach to emotional intonation (Cantero, 2014).

⁸ See a compendium of the first ten years of work in Various Authors (2009) and a more recent revision in *Phonic* magazine, vol. 9-10 (2013-2014). <http://revistes.ub.edu/index.php/phonica/index>

⁹ <http://www.ub.edu/lfa/>

Much of this research is still on-going. Currently, work is being made on melodic patterns for Portuguese, German and English, as well as different studies on dialects of Catalan and American Spanish. We have also begun working on the educational applications of our intonation research, particularly on paralinguistic intonation (Devís, 2015; Devís & Bartolí, 2015).

Among the urgent research tasks we propose is the implementation of the script for Praat in a new version that goes beyond melodic analysis and allows the standardisation of intensity (dynamic analysis) and duration, permitting a complete *prosodic analysis* of speech.

In effect, our method only allows the analysis of melodies. However, as we look further into the analysis of spontaneous speech and the prelinguistic and paralinguistic levels of intonation, we find differences that go beyond the strictly melodic, related to the dynamics and the relationships of duration between the melodic units, which doubtless play a very important role in speech.

Therefore, the next step we will take in the analysis of intonation will be the Prosodic Analysis of Speech (Cantero, 2019).

6. References

- Alfonso, R. (2010). *El vocalismo del español en el habla espontanea*. PhD thesis. Barcelona: Universitat de Barcelona.
- Araújo, M. (2014). *Entonação das interrogativas e das declarativas do português brasileiro falado em Minas Gerais: Modelos para o Ensino de Línguas*. (Masters Dissert.). Universidade de Brasília.
- Armstrong, L. E. & Ward, I. C. (1926). *Handbook of English Intonation*. Leipzig/Berlin: Teubner.
- Avesani, C. (1995). ToBI. Un sistema di trascrizione per l'intonazione italiana. In Lazzari, G. (Ed.), *Atti delle V Giornate di studio del gruppo di fonetica sperimentale* (pp. 85–98). Trento: Sevizio editoria ITC.
- Baditzné, K. (2012). *Spanish Intonation of Hungarian Learners of Spanish; Yes or No Questions*. *Biblioteca phonica*, 15. Barcelona: Laboratori de Fonètica Aplicada UB. Retrieved from http://www.publicacions.ub.edu/revistes/phonica-biblioteca/15_Kata.pdf (accessed 9/4/2019)
- Ballesteros, M. P. (2011). *La entonación del español del norte*. PhD thesis. Universitat de Barcelona. Retrieved from <https://www.tdx.cat/handle/10803/109963> (accessed 9/4/2019).
- Ballesteros, M. P. & Font-Rotchés, D. (2019). Acoustic analysis of intonation: Comparison between two dialects of Spanish from the north of the peninsula. *Lingua*, 221, 49-71.
- Beckman, M., Díaz-Campos, M., Tevis McGory, J. & Morgan, T. A. (2002). Intonation across Spanish, in the Tones and Break Indices framework. *Probus*, 14, 9–36.
- Beckman, M. & Hirschberg, J. (1994). *ToBI annotation conventions*. The Ohio State University Research Foundation.
- Boersma, P. & Weenink, D. (2019). *Praat: Doing Phonetics by Computer* (Computer program). Version 6.1.02. Retrieved from <http://www.praat.org>.
- Bolinger, D. L. (1986). *Intonation and its parts*. Stanford: Stanford University Press.
- Bolinger, D. L. (1989). *Intonation and its uses*. Stanford: Stanford University Press.

- Cabedo, A. (2013). Sobre prosodia, marcadores del discurso y unidades del discurso en español: evidencias de un corpus oral espontáneo. *Onomázein*, 28, 201–213.
- Cantero, F. J. (1999). Análisis Melódico del Habla: principios teóricos y procedimiento. Actas del I Congreso de Fonética Experimental (pp.127-133). Tarragona: Universitat Rovira i Virgili.
- Cantero, F. J. (2002). Teoría y análisis de la entonación. Barcelona: Edicions Universitat Barcelona.
- Cantero, F. J. (2014). Códigos de la entonación y entonación emocional. In *Comunicación, Cognición, Cibernétic@*. Actas del 31.er Congreso Internacional de la Asociación Española de Lingüística Aplicada (pp. 618–629). Tenerife: Universidad de la Laguna.
- Cantero, F. J. (2016). Corpus de habla espontánea para el estudio de la entonación. In: 53 reflexiones sobre aspectos de la fonética y otros temas de lingüística (pp. 151–160). Barcelona: Laboratori de Fonètica de la Universitat de Barcelona.
- Cantero, F.J. (2019): Análisis prosódico del habla: más allá de la melodía. In Álvarez, M.R.; Muñoz Alvarado, A. & Ruiz, L. (eds.): *Comunicación Social: Lingüística, Medios Masivos, Arte, Etnología, Folclor y otras ciencias afines*. Vol. II. Santiago de Cuba: Ediciones Centro de Lingüística Aplicada.
- Cantero, F. J., Alfonso, R., Bartolí, M., Corrales, A. & Vidal, M. (2005). Rasgos melódicos de énfasis en español. *Phonica*, 1, 1–40.
- Cantero, F. J. & Devís, E. (2011). Análisis melódico de la interlengua. In Hidalgo, A.; Congosto, Y.; Quilis, M. (eds.): *El estudio de la prosodia en España en el siglo XXI: perspectivas y ámbitos*. Anejo nº 75 de la Revista *Quaderns de Filologia* (pp. 285–300). València: Universitat de València.
- Cantero, F. J. & Font-Rotchés, D. (2007). Entonación del español peninsular en habla espontánea: patrones melódicos y márgenes de dispersión. *Moenia. Revista Lucense de Lingüística y Literatura*, 13, 69–92.
- Cantero, F. J. & Font-Rotchés, D. (2009). Protocolo para el análisis melódico del habla. *Estudios de Fonética Experimental*, 18, 17–32.
- Cantero, F. J. & Font-Rotchés, D. (2013). The intonation of absolute questions of Brazilian Portuguese. *Linguistics and Literature Studies*, 3, 142–149.
- Cortés, M. (2004). Análisis acústico de la producción de la entonación española por parte de sinohablantes. *Estudios de Fonética Experimental*, XIII, 142–149.
- Cruttenden, A. (1986). *Intonation*. Cambridge: Cambridge University Press.
- Crystal, D. (1969). *Prosodic systems and intonation in English*. Cambridge: Cambridge University Press.
- Devís, E. (2011a). La entonación de (des)cortesía en el español coloquial. *Phonica*, 7, 36–79.
- Devís, E. (2011b). La entonación del español hablado por italianos. *Didáctica. Lengua y Literatura*, 23, 35–58.
- Devís, E. (2015). Contextos para una aplicación didáctica de la entonación atenuadora en español. *Rilce. Revista de Filología Hispánica*, 31(1), 154–170.
- Devís, E. & Bartolí, M. (2015). Propostes didàctiques d'entonació de cortesia atenuant per al nivell B1 de català com a llengua estrangera. In Anejo 7 de la Revista *Normas. Revista de Estudios Lingüísticos Hispánicos*. Adrián Cabedo Nebot (ed.): *Perspectivas actuales en el análisis fónico del habla: tradición y avances en la fonética experimental* (pp. 199–210).
- Devís, E. & Cantero, F. J. (2014). The intonation of mitigating politeness in Catalan. *The Intonation of Mitigating Politeness in Catalan*, 10(1), 127–149.
- Estebas-Vilaplana, E. & Prieto, P. (2008). La notación prosódica en español. Una revisión del Sp_ToBI. *Estudios de Fonética Experimental*, XVII, 263–283.

- Face, T. & Prieto, P. (2007). Using accents in Castilian Spanish: a revision of Sp_ToBI. *Journal of Portuguese Linguistics*, 6.1, 117-146.
- Fonseca, A. (2013). Caracterización de la entonación del español hablado por brasileños. PhD Thesis. Universitat de Barcelona. Retrieved from https://www.tesisenred.net/bitstream/handle/10803/134929/01.AFdO_1de3.pdf?sequence=1&isAllowed=y (accessed 9/4/2019).
- Font-Rotchés, D. (2007). *L'entonació del català*. Barcelona: Publicacions de l'Abadia de Montserrat.
- Font-Rotchés, D. (2011). Èmfasis bàsics de l'entonació en els contorns del català central. In Lloret, M.R.; Pons, C.: *Noves aproximacions a la fonologia i la morfologia del català* (pp. 193–216). València: Institut Interuniversitari de Filologia Valenciana.
- Font-Rotchés, D. (2014). Rasgos melódicos de la entonación paralingüística de foco. In Díaz Galán et al. (eds): *Comunicación, Cognición, Cibernético*. Actas del 31 Congreso Internacional de la Asociación Española de Lingüística Aplicada. Tenerife: Universidad de la Laguna.
- Font-Rotchés, D. & Cantero, F.J. (2009). Melodic Analysis of Speech Method applied to Spanish and Catalan. *Phonica*, 5, 33-47.
- Font-Rotchés, D. & Machuca, M. (2010). Énfasis en los eslóganes publicitarios de radio y televisión. *Icono 14. Revista de Comunicación y Tecnologías Emergentes*, Actas n. A4, 591–605.
- Font-Rotchés, D. & Mateo, M. (2013). Entonación de las interrogativas absolutas del español peninsular del sur en habla espontánea. *Onomázein*, 28, 256–275.
- Font-Rotchés, D. & Mateo, M. (2018). Melodías para confirmar, preguntar, sugerir o pedir en español. *Phonica*, 13, 49–67.
- Font-Rotchés, D. & Paloma, D. (2012). Tendencias entonativas en los titulares televisivos en catalán. *Círculo de Lingüística Aplicada a La Comunicación*, 51, 50–81.
- García Riverón, R., Bermúdez, M., Pedrosa, A. & Marrero, A. (2010). El sistema de entonación del español de Cuba a la luz del modelo de análisis melódico del habla. *Phonica*, 6, 3–25.
- Grice, M., Reyelt, M., Benz Müller, R., Mayer, J. & Batliner, M. (1996). Consistency in transcription and labelling of German intonation with GToBI. In Bunnell, T.; Idsardi, W. (Eds.), *Proceedings of the 4th International Conference on Spoken Language Processing* (pp. 1716–1719). New York Institute of Electrical and Electronic Engineers.
- Gussenhoven, C., Rietveld, T. & Terken, J. (1999). ToDI: Transcription of Dutch Intonation. Retrieved January 5, 2019, from <http://gep.ruhosting.nl/carlos/todisun-ah.pdf> (accessed 9/4/2019).
- Hirst, D. J., Di Cristo, A. & Espesser, R. (2000). Levels of representation and levels of analysis for the description of intonation systems. In Home, M. (Ed.), *Prosody: Theory and Experiment* (pp. 51–87). Dordrecht: Kluwer Academic Press.
- Jones, D. (1909). *Intonation Curves*. Leipzig/Berlin: Teubner.
- Kao, W. L. (2011). La entonación de enunciados declarativos e interrogativos en chino mandarín hablado por taiwaneses. (Masters Dissert.). Universitat de Barcelona, Barcelona.
- Liu, Y. H. (2005). La entonación del español hablado por taiwaneses. *Biblioteca Phonica*, 2. Laboratori de Fonètica Aplicada (UB). Retrieved from <http://www.publicacions.ub.edu/revistes/phonica-biblioteca/> (accessed 9/4/2019).
- Martínez, D. (2015). La importancia del factor prosódico en el estudio de los marcadores del discurso: algunos problemas de su análisis acústico-melódico. *Círculo de Lingüística Aplicada a La Comunicación*, 62, 105–124.

- Martorell, L. (2010). Les interrogatives absolutes de l'espanyol parlat per suecs. (Masters Dissert.). Universitat de Barcelona. Retrieved from <http://diposit.ub.edu/dspace/bitstream/2445/32302/1/Treball%20fi%20m%20C3%A0ster%20Laura%20Martorell.pdf> (accessed 9/4/2019).
- Mateo, M. (2010a). Protocolo para la extracción de los datos tonales y curva estándar en análisis melódico del habla (AMH). *Phonica*, 6, 49–90.
- Mateo, M. (2010b). Scripts en Praat para la extracción de datos tonales y curva estándar. *Phonica*, 6, 91–111.
- Mateo, M. (2014). La entonación del español meridional. PhD thesis. Universitat de Barcelona. Retrieved from <http://diposit.ub.edu/dspace/handle/2445/53156> (accessed 9/4/2019).
- Navarro Tomás, T. (1944). *Manual de entonación española*. New York: Hispanic Institute.
- O'Connor, J. D., & Arnold, G. F. (1961). *Intonation of colloquial English*. London: Longman.
- Palmer, H. E. (1922). *English Intonation, with systematic exercises*. Cambridge: Heffer.
- Pereira, D. I. (2014). Patrones y mecanismos prosódicos de cuantificación e intensificación en construcciones exclamativas. PhD thesis. Universidad de Concepción (Chile).
- Pierrehumbert, J. B. (1980). *The Phonology and Phonetics of English Intonation*. PhD thesis. MIT.
- Pierrehumbert, J. B. (1987). *The Phonology and Phonetics of English Intonation*. Bloomington: Indiana University Linguistics Club.
- Pike, K. L. (1945). *The intonation of American English*. Ann Arbor: University of Michigan Press.
- Pons, J. (2007). La ràdio en menorquí i l'entonació de les preguntes parcials. *Phonica*, 7, 100–117.
- Prieto, P. & Roseano, P. (eds.). (2010). *Transcription of Intonation of the Spanish Language*. München: Lincoln Europa.
- Ruiz Mella, M. (2011). Determinación de la incidencia de la entonación en la producción y percepción de oraciones declarativas con focalización de tipo contrastiva en hablantes de español de Concepción. PhD thesis. Universidad de Concepción (Chile).
- Sena, R. (2013). A entonaçãõ no processo de ensino-aprendizagem de PLE. Proposta didática para o ensino de modelos de entonaçãõ interrogativa do português do Brasil-Estado de São Paulo. (Masters Dissert.). Universidade de Brasília.
- Silva, C. (2009). Análisis melódico de declarativas e interrogativas absolutas en español/LE. *Phonica*, 6, 92–113.
- Silverman, K., Beckman, M., Pitrelli, J., Ostendorf, M., Wigham, C., Price, P. & Hirschberg, J. (1992). ToBI. A standard for labelling English prosody. In Ohala, J. et. al. (Eds.), *Proceedings of the 1992 International Conference on Spoken Language Processing* (Vol. 2). Edmonton: University of Alberta.
- t'Hart, J., Collier, R. & Cohen, A. (1990). *A perceptual study of intonation. An experimental-phonetic approach to speech melody*. Cambridge: Cambridge University Press.
- Torregrosa, J. (2015). Una propuesta metodológica para el análisis melódico de enunciados orales espontáneos con partículas modales del alemán. *Revista de Filología Alemana*, 23, 193–212.
- Torregrosa, J. & Font-Rotchés, D. (2017). Melodic patterns of absolute interrogative utterances in northern German in spontaneous speech. *Linguistics*, 55(6), 1311–1335.
- Torrent, N. (2015). Barberà i Cuní: dos models entonatius per a programes informatius. *Phonica*, 15, 87–102.
- Trager, G. L. & Smith, H. L. (1951). *An outline of English Structure*. Norman, Okla: Battenburg Press.

Various Authors (2009): Análisis Melódico del Habla (AMH): 1999-2009. Biblioteca Phonica, 10. http://www.publicacions.ub.edu/revistes/phonica-biblioteca/10_AMH.pdf (accessed 9/4/2019).

EUSKARA BURUAN ETA IBILI MUNDUAN: AFRIKATUETAN BARNA¹⁰

Iñaki Gaminde, Naia Eguskiza eta Irati De Pablo
Euskal Herriko Unibertsitatea (UPV/EHU)

inaki.gaminde@gmail.com, naia.eguskuiza@ehu.eus, irati.depablo@ehu.eus

Laburpena:

Lan honetan euskararen lekua ikertu nahi da munduko hizkuntzetan afrikatuei dagokienez. Horretarako, euskara besteen artean kokatzekotan, 600 hizkuntzaren afrikatuek osatutako datu basea osatu da lau irizpide nagusi erabilita: (a) hizkuntza bizirik egotea XXI. mendean; (b) zonalde geografiko guztiak ordeztuta egotea; (c) hizkuntza familia guztiak ordeztuta egotea; eta (d) bibliografia nahikoa egotea. Euskarari dagokionez, erkaketak egiteko euskara estandarra oinarritzat hartu bada ere, barietateetan barna kausitzen diren afrikatuen tasun artikulatorioak eta akustikoak errepasatzen dira, oraindik ikertzeko dauzkagun arazo batzuk agirian jartzekotzat. Euskararen barietateetako eta gainerako hizkuntzen tasunez jardundakoan, afrikatuen hedadurak alderatzeko hizkuntzen ratio bi kalkulatu dira, bata hizkuntzaren fonema guztien arabera eta bestea, berriz, kontsonanteen arabera. Ratio horien arabera sailkatu ahal izan dugu euskara besteen artean. Ondorioetan ikusi dugu ezen, barietate geografikoak, sozialak eta estilistikoak aztertzean gai hau aurrerantzean ere sakontzeko interesgarria izan daitekeela. Bestalde, ikusi dugu euskararen afrikatuen sistema hizkuntza gehien bide nagusian kokatzen dela, bai ahoskatzeko erei dagokienez, bai eta ratioei dagokienez ere.

Hitz gakoak: Afrikatuak, tipologia, fonologia, fonetika akustikoa eta artikulatorioa

Abstract:

The aim of this work is to investigate the place of the Basque language in relation to affricates in the world's languages. To this end, in order to situate Basque among others, a database of affricates of 600 languages of the world has been compiled using four main criteria: (a) language being alive 21st century; (b) there must be represented all geographical areas; (c) there must be represented all language families; and (d) there must be sufficient bibliography. The standard Basque has been used to analyze the articulatory and acoustic rates of affricates that occur across varieties, in order to reveal some of the problems that are still investigating. In relation to the varieties of the Basque language and the other languages, two ratios of languages to compare the size of the affricates have been calculated, one for all phonemes of the language and the other for consonants. With those ratios we have been able to classify Basque among other languages. In conclusion, we have seen that, when analyzing geographical, social and stylistic varieties, this topic may be interest to deepen in the future. On the other hand, we have seen that the affricate

¹⁰ Ikerketa hau burutu da IT1028/16 (Eusko Jaurlaritza) eta GIU 16/22 (Euskal Herriko Unibertsitatea) proiektuen barnean eta UPV/EHUK doktoretza aurreko formakuntzan dagoen pertsonala kontratatzeko (PIF) emandako laguntza ekonomikoari esker

system of Basque is located in the main path of most languages, both in terms of pronunciation and in terms of ratios.

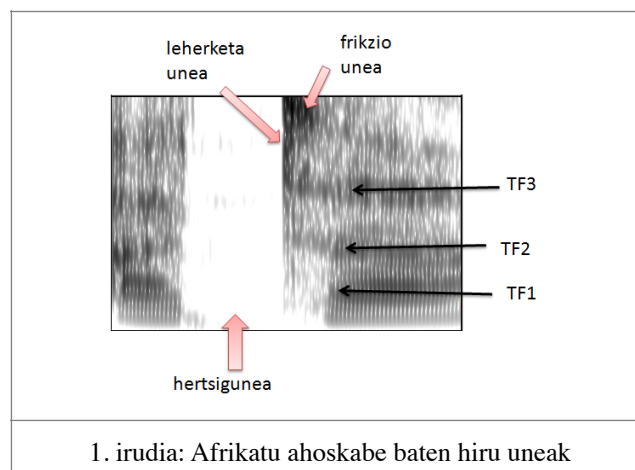
Key words: Affricates, typology, phonology, acoustic and articulatory phonetics

1. Sarrera

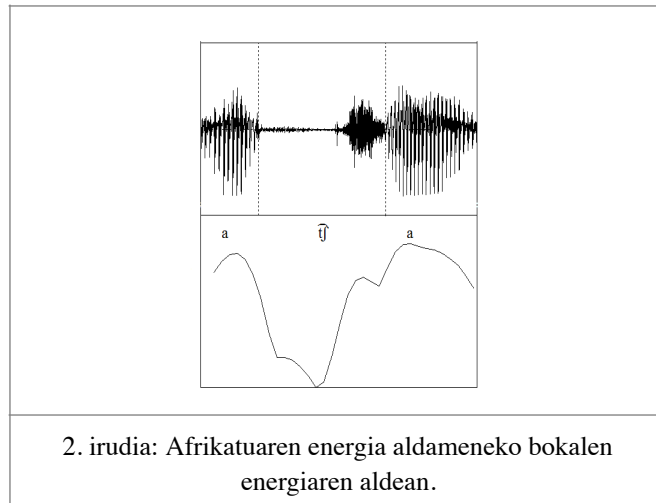
Izenburuan adierazi gura izan den moduan, euskara buruan mundura irten gura izan dugu jakiteko inorenak lagun diezagukeen gurea hobeto ulertzen. Txango hau egin ondoren, konturatu gara asagoko intxaurrak hamalau badauzka ere, parajetuta, lautan geratzen direla eta gehienetan, gurean legez, txakurrak ortozik dabiltzala. Hala ere, lan honen agiriko zein ezkutuko helburuetarako ona izan daiteke hor gaindi dagoenaz zerbait jakitea. Honenbestez, lan honen helburua da munduko hizkuntzetan euskararen lekua ikertzea afrikatuei dagokienez. Egiteko honetarako euskara estandarren eredia aztergaitzat hartu dugun arren, afrikatuen arlo honetan ikertzeke ditugun arazo batzuk agirian jartzen ahaleginduko gara bide batez.

Afrikatuak herskarien eta frikarien arteko soinutzat jotzen izan dira (Martínez Celdrán, 1989; Quilis, 1993). Afrikatuen izaera definitzean ikertzaileak bat datoz adieraztean kontsonante hauek une nagusi bi dauzkatela artikulatzean (Oñederra, 2004); lehen unean hertsigune bat gertatzen da, herskarietan bezala, baina bigarren unean arin zabaldu beharrean, apurka-apurka zabaltzen dira artikulatzaileak eta igurtzi egiten da: "This transitory friction is usually considered a part of the release burst of the stop" (Ladefoged & Maddieson, 1996: 90).

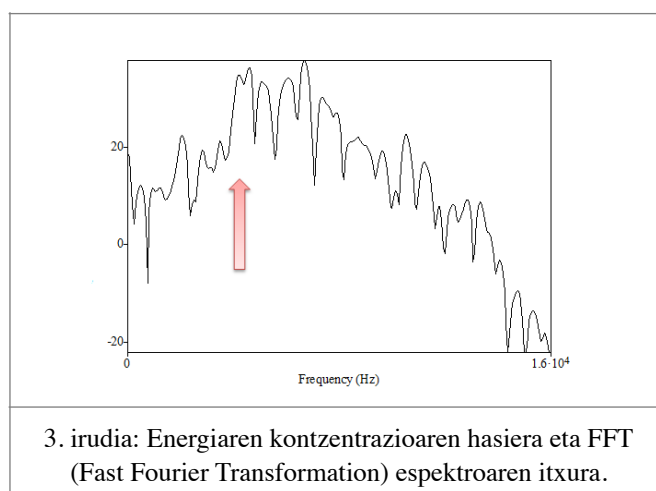
Afrikatuen artikulazioan aipatutako une bi horiez ostera VOT (Voice Onset Time) txikia ere bereiz daiteke (Gaminde et al., 2013; Gandarias et al., 2014). Lehen unean etenaldi bat gertatzen da herskarietan legez, bigarrena leherketa unea da eta leherketa barra bat agertzen da, azkenik, hirugarren unean zarata sortzen da. Herskarietan ez bezala, aire ibilbidea zabaltzen denean, artikulatzaileak oso hurbil geratzen dira eta zarata sortzen da. 1. irudiko espektrograman hiru uneak erakusten dira ondoko bokalerako formakinen trantsizioekin batera.



Afrikatuak akustikoki aztertzeko izari batzuk proposatu dira; izari horiek bost multzotan sailka ditzakegu: iraupena, energia, une espektralak, espektroaren itxura eta ondoko bokalen formakinen trantsizioak. Iraupen arloan hiru neurri proposatu dira (Gaminde et al., 2013): une herskariaren edo hertsigunearen iraupena, leherketaren iraupena eta une frikariaren iraupena. Energia arloan neurri berberak proposatu dira (Gaminde et al., 2013): une herskariaren energia, leherketaren energia eta une frikariaren energia. 2. irudiko espektrograman ikusten den moduan, aldameneko bokalen aldean afrikatuak energia baxuagoa dauka; aldi berean, une herskariaren eta frikariaren arteko aldea ere bertan ikus daiteke.



Une espektraletan lau neurri proposatu dira (Fulop 2011; Hualde 2010; Gaminde et al., 2012, 2013; Gandarias et al., 2014; Jongman et al., 2000; Nissen eta Fox, 2005; Silbert & Jong, 2008), grabitate zentroa, desbiderapena, asimetria eta kurtosia. Une frikariaren espektroaren itxurari dagokionez, 3. irudiko espektrograman ikusten den moduan, energiaren kontzentrazioa altuera desberdinetan gertatzen da afrikatu motaren arabera. Adierazle akustiko hau askotan erabili da (Urrutia et al., 1988, 1989, 1991).



Azkenik, ondoko bokalen formakinen trantsizioen arloan, trantsizioen iraupena eta maldak kalkula daitezke (Gaminde et al., 2013; Quilis, 1993).

Euskara estandarrerako hiru afrikatu ahoskabe onartzen dira \widehat{ts} /, $\widehat{tʃ}$ / eta $\widehat{tʃ}$ / (Hualde, 2003, Oñedera, 2004). Hala ere, badakigu egoera ez dena bat bera bariedade guztietan. Afrikatu ahostunak egon daude eta bariedade askotan ahoskabeen kopurua murriztu da; hau guztia xehekiago aztertuko dugu 4. atalean.

Lan hau sei ataletan banatuta aurkezten dugu; sarrera honen osteko bigarren atalean corpusa eta berau osatzeko metodologia aurkeztuko ditugu, hirugarren atalean afrikatu motak aztertuko dira, laugarrenean euskararen afrikatuek jardungo dugu, bosgarrenean afrikatuen hedaduraz jardungo dugu eta, azkenik, seigarren atalean ondorio nagusiak eta hurrengo lanetarako proposamen batzuk laburbilduko ditugu.

2. Corpusa eta metodologia

Lan hau burutu ahal izateko baliabide bibliografikoetara jo behar izan dugu gehienbat. Munduko hizkuntzetan topa daitezkeen afrikatuak aztertu ahal izateko 600 hizkuntzako datu basea egin dugu. Datu basean agertzen diren hizkuntzak aukeratzeko lau irizpide nagusi erabili dira: (a) hizkuntza bizirik egotea XXI. mendean; (b) zonalde geografiko guztiak ordeztuta egotea; (c) hizkuntza familia guztiak ordeztuta egotea; eta (d) bibliografia nahikoa egotea.

Lehen irizpideari garrantzia eman diogu, izan ere, eraz dauzkagun datu base batzuetan ez dira bereizten desagertu diren hizkuntzak eta bizirik daudenak (Gary eta Fenning, 2018; Moran et al., 2014; Lev et al., 2015; Maddieson et al., 2018). Datu horiek datu baseetan agertzen diren arren, hizkuntza guztien informazioak batera agertzen dira. Ez da erraza jakitea zer den hizkuntza bat bizirik egotea edo ez (Ethnologue-n, Gary eta Fenning, 2018, irizpide batzuk ematen dira); gure kasuan, XXI. mendean hiztunak eduki baditu hizkuntza batek, bizirik dagoela kontsideratu dugu.

Munduko zonalde guztietako hizkuntzak izateko, irizpide kuantitatiboak erabili ditugu; horrela bada, aukeratu ditugun hizkuntz kopuruak eta ehunekoak 1. taulan agertzen direnak dira kontinenteka (4. irudiko mapan banaketa erakusten da).

Kontinentea	Kopurua	%
Afrika	180	30
Amerika	90	15
Asia	180	30
Europa	30	5
Ozeania	120	20
1. taula: Hizkuntzen kopuruak eta ehunekoak kontinenteka.		



Era berean, hizkuntz familia guztiak eta isolatuta dauden hizkuntzak ere ordeztuta agertzen dira; hemen ere ahalegindu gara ordezkari proportzionala izan dadin. Familien arabera sailkapenean WALSek (Dryer & Haspelmat, 2013) erabiltzen duenari jarraitu diogu.

Azkenik, bibliografia nahikoa egotea ere irizpide garrantzitsua izan da hizkuntza aukeratzeko unean, izan ere, hizkuntza askoren gainean izena eta kokagunea baino ez dakigu. Inventario fonologikoak osatzeko nagusiki hiru datu base erabili dira PHOIBLE (Moran et al., 2014), SAHPHON (Lev et al., 2015) eta LAPSyD (Maddieson et al., 2018). Bestalde, hizkuntzak geolokalizatzeko GLOTTOLOG datu basea (Hammarström et al., 2018) erabili da.

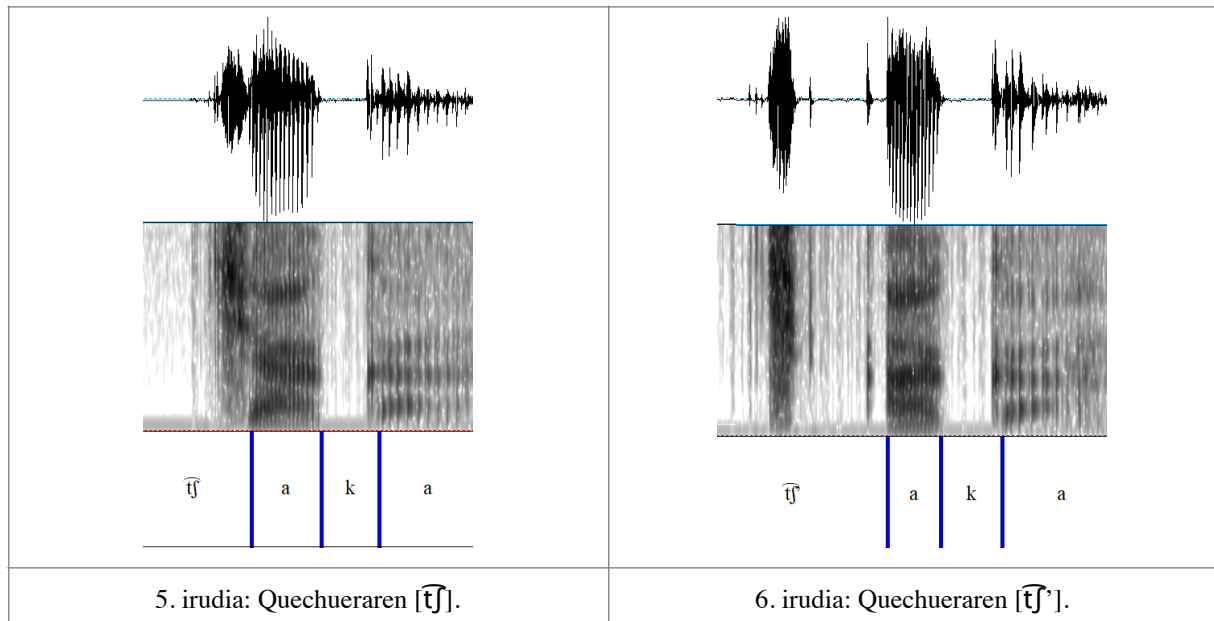
3. Afrikatu motak

Afrikatu motak sailka daitezke eurak biltzen dituzten tasun batzuen arabera. Klikekin batera erabiltzen diren afrikatuak bazter utzita, datu basean batu ditugun 1109 afrikatuak ondoko tasunen arabera sailka daitezke:

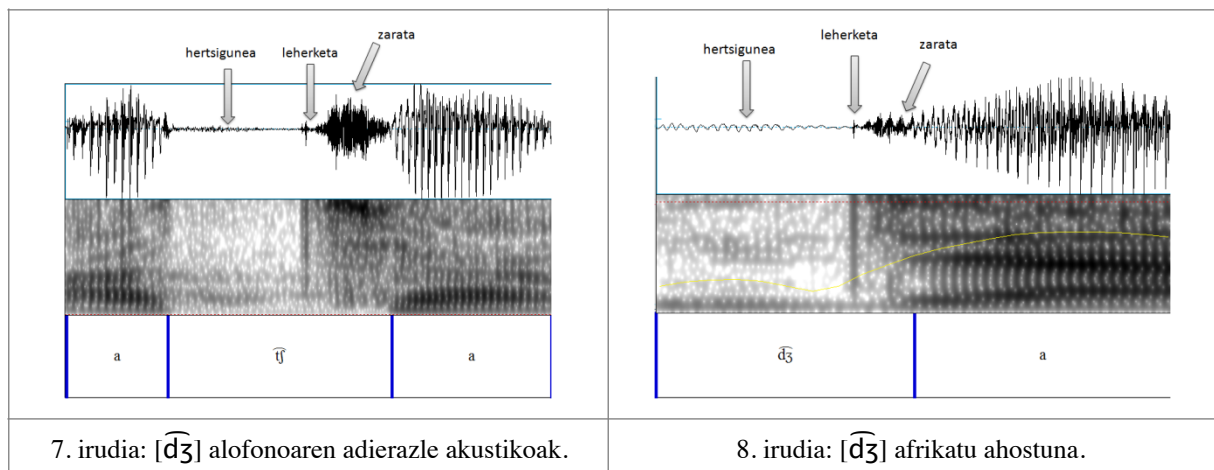
- (1) Aire-etorria
- (2) Ahotsa
- (3) Ahoskunea
- (4) Zabalgunea
- (5) Bigarren mailako artikulazioa
- (6) Fonazio mota
- (7) Iraupena

Lehen tasunari dagokionez, esan dezakegu afrikatu gehienak biriketarik datorren aire-etorriaren bidez egiten diren arren (1060, % 92,41), badaudela ejektiboak ere hainbat hizkuntzatan (87, % 7,59). Honelakoetan glotisaren mugimenduak erabiltzen dira airearen presioa aldatzeko aho barrunbean aire-etorriaren ibilbidea itxi edo estutzen den artean (Ladefoged & Maddieson, 1996; Oñederra, 2004). Gure corpusean afrikatu ejektiboak 38

hizkuntzatan gertatzen dira (% 6,33). 3. irudian quechueraren¹¹ ejektiboaren espektrograma paratu dugu une frikariaren ostean dagoen hutsunea ikusteko eta 6. irudiko espektrograman agertzen den ez-hasperendunarekin alderatzeko.



Ahotsari dagokionez, afrikatu gehienak ahoskabeak dira (745, % 64,95), ahostunak gutxiagotan agertzen dira (402, % 35,05). Euskararen bariedade batzuetan hasierako gunean ahoskabea /tʃ/ eta ahostuna /dʒ/ topa daitezke. 7. eta 8. irudietan bakoitzeko espektro bana ematen dugu.



Ahoskuneari dagokionez, afrikatuak ezpain-horzkarietatik ubularretaraino topatzen ditugu, 2. taulan mota nagusiak agertzen dira ahoskune bakoitzeko.

¹¹ Soinuak UCLA Phonetics Lab Data orritik hartu dira.

Ez_hor	Hor_Ar	Horkari	Albeolar	Post	Erretro	Alb_Sa	Sabai	Belokoa	Ubular
p̄f	t̄θ	t̄s	ts	t̄j	t̄ɕ	t̄c	c̄ç	k̄x	q̄ɣ
b̄v	d̄ð	d̄z	d̄z	d̄ʒ	d̄ʒ	d̄z	ʃ̄j	ḡɣ	-

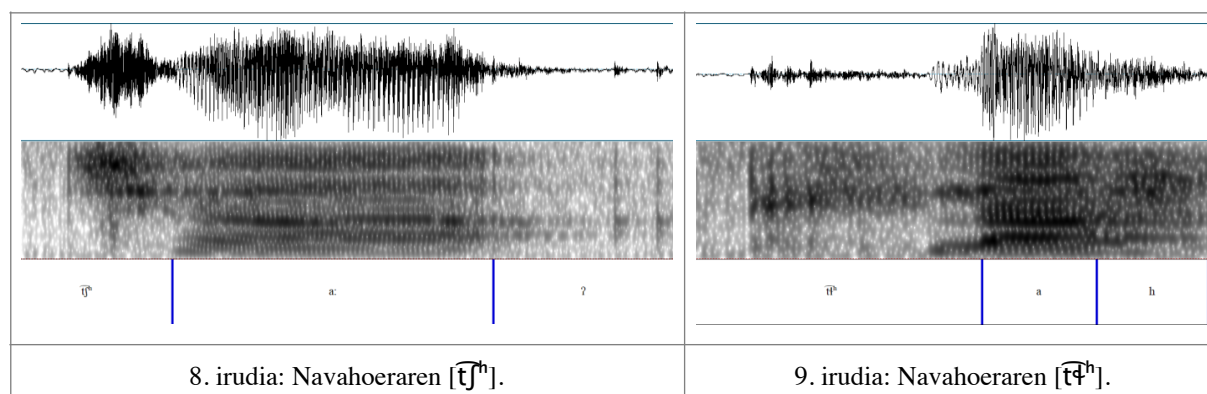
2. taula: Afrikatuak ahoskuneen arabera, goikoak ahoskabeak dira eta behekoak ahostunak.

3. taulan ahoskune bakoitzaren arabera agertzen diren kopuruak eta ehunekoak ematen ditugu. Bertan agertzen den moduan, post-albeolarrak dira ehunekorik handiena daukatenak; bestalde, hartz eta albeoloetan egiten direnekin batera guztien % 81,08 biltzen dute.

Ahoskunea	Kopurua	%	Ahoskunea	Kopurua	%
Ezpain-horzkariak	23	2,01	Erretroflexuak	39	3,40
Hortzartekoak	4	0,35	Aleolo sabaikariak	34	2,96
Horzkariak	97	8,46	Sabaikariak	70	6,10
Albeolarrak	298	25,98	Belokoak	23	2,01
Post-albeolarrak	531	46,29	Ubularrak	5	0,44

3. taula: Afrikatuen kopuruak eta ehunekoak ahoskune bakoitzeko.

2. eta 3. tauletan ikusi ditugun afrikatuen zabalguneeetan une frikariak baino ez ditugu ikusi; une frikarien ordez albokari-frikariak ere ager daitezke, hala nola /t̄ɕ/ eta /d̄ʒ/. Era honetakoak gutxitan agertzen dira (29, % 2,53). 9. irudian navahoeraren [t̄ɕ^h] paratu dugu; berorren espektrograma erkatu daiteke 8. irudiko hizkuntza beraren [t̄j^h]ren espektrogramarekin.



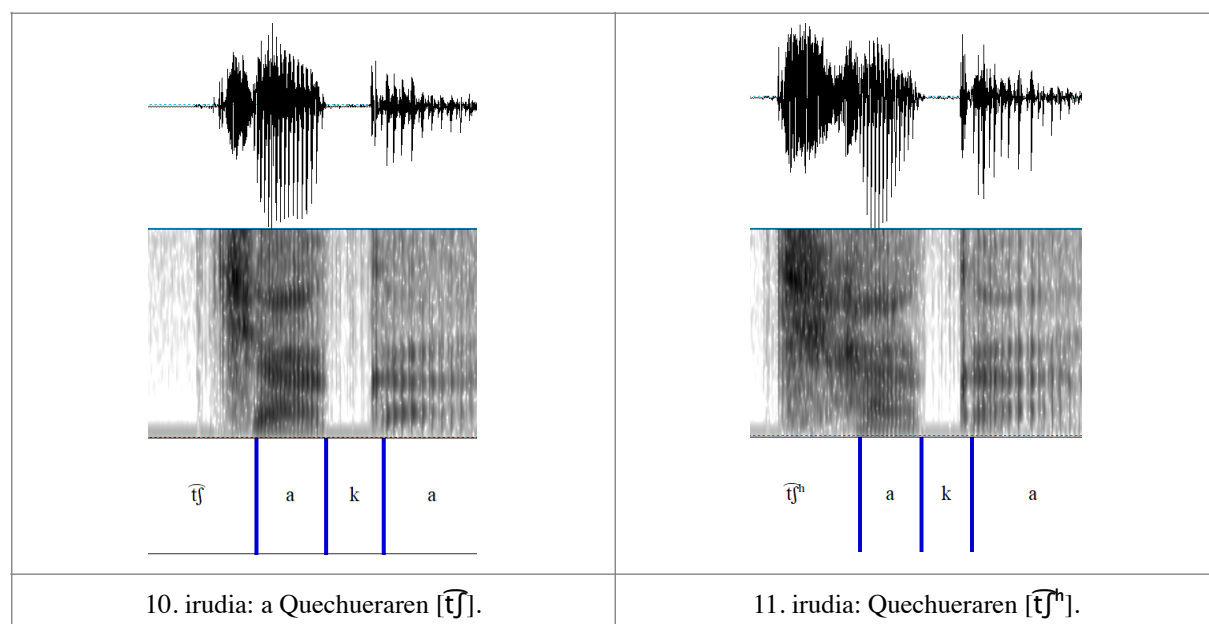
Bigarren mailako artikulazioak gerta daitezke afrikatuekin batera; batzuk aurretik gertatzen dira (79, % 6,89) eta beste batzuk aldi berean (185, % 15,95). Aurreko artikulazioetan gehien agertzen direnak sudurkariak dira (4. taulan gehien erabiltzen direnak paratu ditugu). Aurrehasperendunak (/^ht̄s/ eta /^ht̄j/) eta aurre-herskariak (/p'k̄x'/ eta /t'k̄x'/) behin baino ez

zaizkigu agertu; aurre-hasperendunak Hopieraz eta Ojiweraz, hurrenez hurren, eta aurre-herskariak Xoo hizkuntzan.

Ez hor	Horkari	Albeolar	Post	Erretro	Belokoa
$mp̪f̪$	$nts̪$	$nts̪$	$ntʃ̪$	$nts̪$	$ŋkx̪$
$mbv̪$	$ndz̪$	$ndz̪$	$ndʒ̪$	$ndz̪$	$gɣ̪$

4. taula: Afrikatu aurre-sudurreztatuen mota nagusiak.

Aldi berean egiten diren artikulazioak lau dira: ezpainkariak, sabaikariak, hasperendunak eta ezpainkari-hasperendunak. Gehien agertu direnak hasperendunak dira ($/p̪f̪^h/$, $/ts̪^h/$, $/tʃ̪^h/$, $/kx̪^h/$, e.a.), (148, % 12,9), 10. eta 11. irudietan quechueraren $[tʃ̪]$ eta $[tʃ̪^h]$ soinuen espektrogramak erakusten dira une frikariaren luzeran dagoen aldea ikusteko. Ezpainkarietan ($/d̪z̪^w/$, $/ts̪^w/$, $/tʃ̪^w/$, $/q̪x̪^w/$, e.a.) 23 jaso dira (% 2,01), sabaikarietan ($/ts̪^j/$ eta $/tʃ̪^j/$) 7 kasu jaso dira (% 0,61) eta ezpainkari-hasperendunetan ($/ts̪^wh/$ eta $/tʃ̪^wh/$) 5 kasu (% 0,44).



Fonazioari dagokionez, afrikatu murmurikatuak agertzen dira ahots arruntaren aldean (esate baterako: $/b̪v̪/$, $/d̪z̪/$, e.a.); gure datu basean 18 kasu gertatu dira (% 1,57). Afrikatu murmurikatuak 10 hizkuntzatan agertu dira, Asian 8tan (Darmiya, Gujaratera, Hindi, Jaunsariera, Kurush, Newarera, Parauk eta Sauhrastrera) eta Afrikan 2tan (Tsongera eta Xhosea).

Azkenik, kasu gutxitan bada ere (13, %1,13), afrikatu luzeak fonematizat jotzen dira hizkuntza batzuetan (Kaukasoarrak diren Avarera eta Lakera, Dravidarra den Malayalamera, Niger-Kongoarra den Soninkera eta Khoisan den Kung hizkuntzan).

4. Euskararen afrikatuez

Euskara estandarerako hiru afrikatu ahoskabe onartzen dira; barietateetan, ordea, zehar afrikatuen fonema ahostunak eta ahoskabeak agertzen zaizkigu (5. taula). Orain arteko lan gehienetan afrikatu ahoskabeaz arduratu dira egileak (Aranbarri, 1996; Arretxe, 1994; Barrutia, 1996; Bilbao, 2002; Errazti, 1994; Etxebarria, 1991; Gaminde, 1992, 2001, 2003a, b, 2004, 2007; Gaminde et al., 2004, 2005; Gilisasti, 2003; Hualde & Bilbao, 1992; Salazar, 2001; Txillardegui, 1980; Uriarte, 1995; Yárnoz, 2001; beste askoren artean).

	Albe. apikaria	Albe. laminarra	Sabaiaurrekoa
Ahoskabeak	$\widehat{t}_{\square} (\widehat{t}_{\square})$	\widehat{t}_{\square}	$\widehat{t}_{\square} (\widehat{t}_{\square})$
Ahostunak	\widehat{d}_{\square}	\widehat{d}_{\square}	\widehat{d}_{\square}
5. taula: Fonema afrikatuak.			

Afrikatu ahostun sabaiaurrekoa Mendebaldeko zonalde batzuetan erabiltzen da (Gaminde, 2007) eta Zuberoko beste bakan batzuetan. Afrikatu ahostun apikaria Sakanan agertzen da (Hualde, 1996). Azkenik, afrikatu ahostun laminarra Ondarroan eta Lekeition erabiltzen da, gehienbat berba berezi batzuekin “dzanga”, “dzapart”, “dzingo”, e.a. (Lekeitioko Berbak Lantaldea, 2001).

$[\widehat{d}_{\square}]$ hitz hasieran edo sudurkariaren ostean agertzen izan da, baina bokalen artean gertatuz gero (1) arauaren bidez frikari bihurtzen izan da.

(1) +Afri > +Frik / V__V

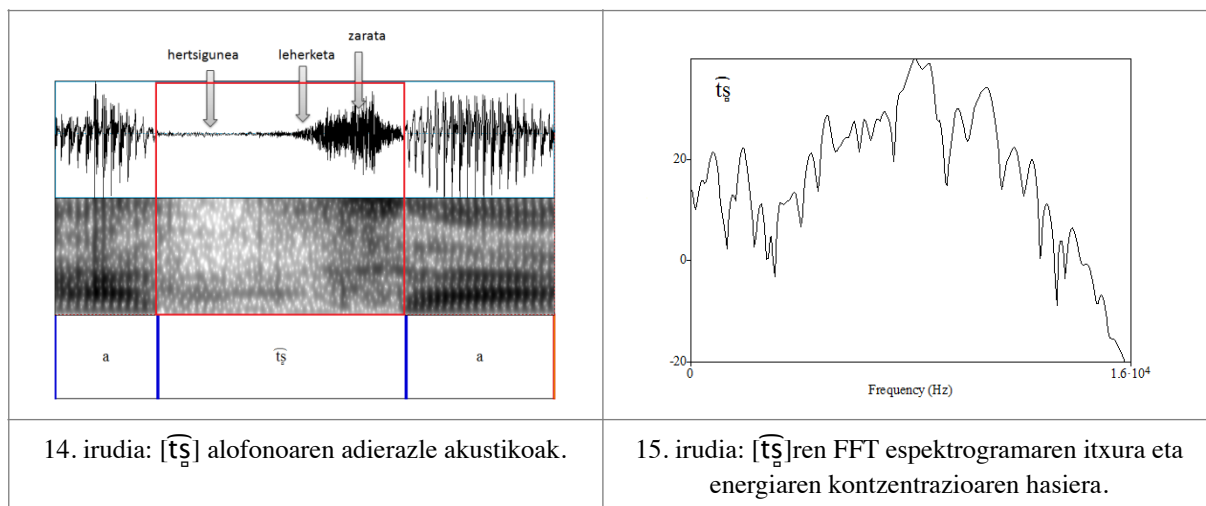
Alofonoen deskripzioari dagokionez, $[\widehat{t}_{\square}]$ artikulatorioki afrikatu albeolar bizkarkari ahoskabetzat hartzen da (Oñederra, 2004; Pagola, 1992) Hualdek laminar izatea aipatzen du bereziki (Voiceless lamino-alveolar affricate, Hualde, 2003). 12. eta 13. irudietako palatogrametan aho sabaian mihiak uzten dituen lorratzak ikus daitezke. Mutilaren kasuan artikulazioaren lorratza aho barrunbearen aurrealdean baino ez da egiten. Bestalde, kasu guztietan goiko hortzetan ere ukitzen dela ikus dezakegu, artikulazio aldetik bizkarkaria eta hortz-albeolarra dela esan dezakegu; baliteke horzkari hutsa ere egitea $[\widehat{t}_{\square}]$.



12. irudia: Mihiak uzten duen lorratza $[\widehat{t\zeta}]$ egitean, (neska batek egina).

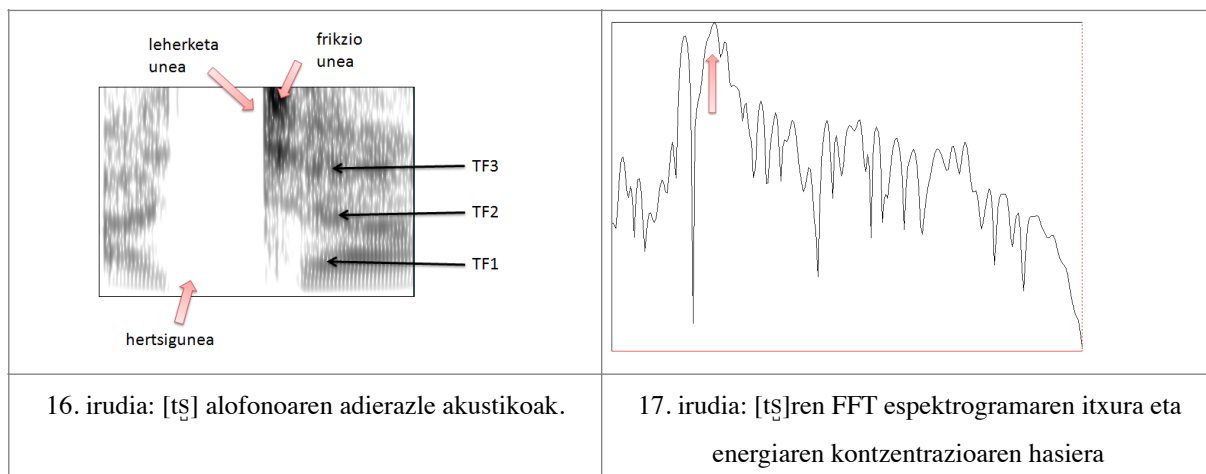
13. irudia: Mihiak uzten duen lorratza $[\widehat{t\zeta}]$ egitean, (mutil batek egina).

Akustikoki une frikariaren energia goiko frekuentzietan indartzen dela ikus dezakegu 14. irudiko espektrograman. Espektrograman hertsigunearekin eta une frikariarekin batera leherketa barra ere ikus daiteke. 15. irudiko FFT espektrograman energiaren kontzentrazio handiena non gertatzen den erakusten da. Aldameneko bokalaren trantsizioen iraupenek eta norabideek laguntzen dute soinu mota bereizten.



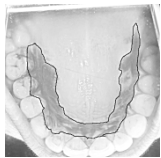
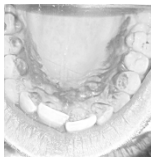
Artikulatorioki $[\widehat{t\zeta}]$ afrikatu albeolar apikari ahoskabetzat jotzen da (Oñederra, 2004). Nafarroako barietateei dagokienez, hainbat egilek eretroflexutzat $[\widehat{t\zeta}]$ jo izan dute (N´Diaye, 1970; Pagola, 1992; Salaburu, 1984; Yárnoz, 2001).

Akustikoki une frikariaren energia beheko frekuentzietan indartzen dela ikus dezakegu 16. irudiko espektrograman. Espektrograman hertsigunearekin eta une frikariarekin batera leherketa unea ere ikus daiteke. 17. irudiko FFT espektrograman energiaren kontzentrazio handiena non gertatzen den erakusten da.











Artikulatorioki $[\widehat{t\zeta}]$ afrikatu sabaiaurreko ahoskabetzat hartzen dute Oñederrak (2004) eta Hualdek (2003); Pagolak (1992) sabaikotzat jotzen du eta Yárnoz (2001), Bortzirietan

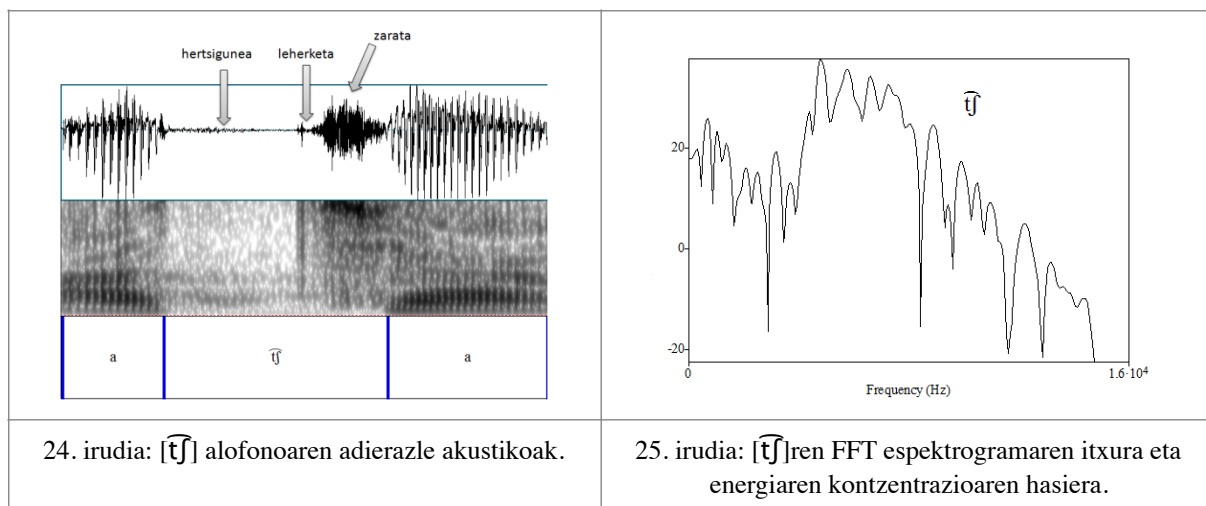
egindako lanean, albeolo sabaikaritzat [t̪]. 18. eta 19. irudietako palatogrametan aho sabaian mihiak uzten dituen lorratzak ikus daitezke.

	
18. irudia: Mihiak uzten duen lorratza [t̪] egitean, (neska batek egina).	19. irudia: Mihiak uzten duen lorratza [t̪] egitean, (mutil batek egina).

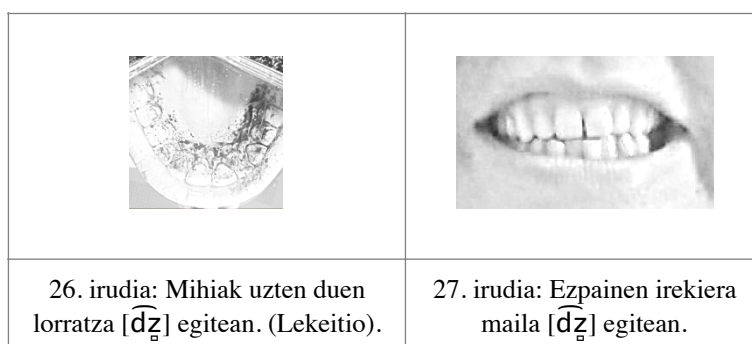
Afrikatu ahoskabe biak bereizten dituzten informatzaileen adibideak paratu ditugu 20, 21, 22. eta 23. irudi multzoetan. Artikulazioa aho barrunbean atzerago egiten bada, orduan ezpainak irekiago agertzen dira. Bestalde, sabaiaurrekoaren kasuan ezpainak arean biribiltzen dira [t̪^w] egiteraino.

			
[t̪s̪]	[t̪]	[t̪s̪]	[t̪]
20. irudia: Ezpainen irekiera maila afrikatu biak egitean (Abadiño).		21. irudia: Ezpainen irekiera maila afrikatu biak egitean (Lekeitio).	
			
[t̪s̪]	[t̪]	[t̪s̪]	[t̪]
22. irudia: Ezpainen irekiera maila afrikatu biak egitean (Mallabia).		23. irudia: Ezpainen irekiera maila afrikatu biak egitean (Arrigorriaga)	

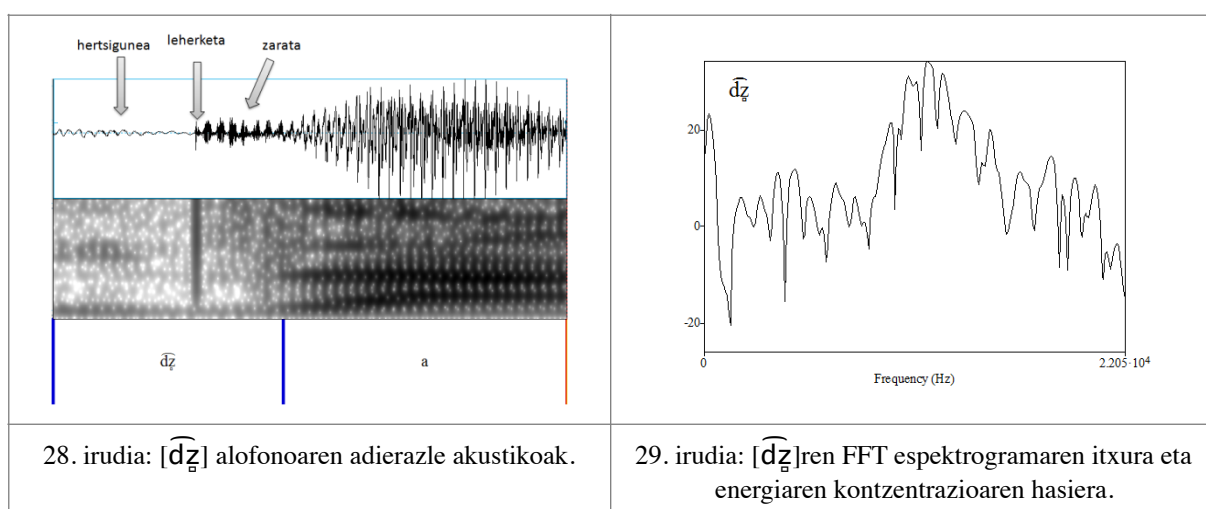
Akustikoki une frikariaren energia laminarraren kasuan baino beherago agertzen dela ikus dezakegu 24. irudiko espektrograman. Espektrograman hertsigunearekin eta une frikariarekin batera leherketa barra ere ikus daiteke. 25. irudiko FFT espektrograman energiaren kontzentrazioak handiena non gertatzen den erakusten da.





Artikulatorioki [dʒ] afrikatu ahostun bizkar edo laminar albeolartzat jo dezakegu. Gorago esan dugun moduan, Ondarroan eta Lekeitioan erabiltzen da berba berezi batzuetan. 26. irudiko palatograman mihiak ukitze gunean uzten duen lorratza ikus daiteke; era berean, 27. irudian ezpainen irekiera maila erakusten da.



Lekeitioko [dʒ] 28. irudiko espektrograman erakusten dugu soinu honen adierazle akustikoak agirian uzteko. 29. irudiko espektrograman energiaren kontzentrazioaren hasiera ikus daiteke.

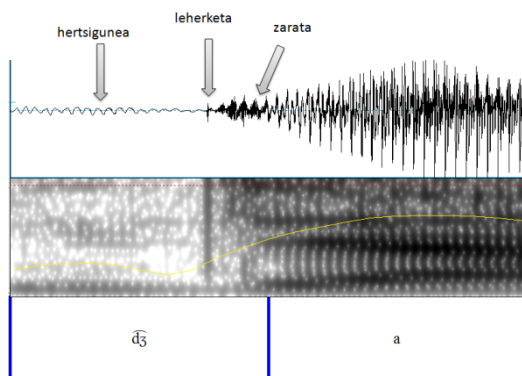


Artikulatorioki [dʒ] afrikatu sabaiurreko ahostuntzat jotzen da (Oñederra, 2004). 30. irudiko palatograman mihiak ukitze gunean uzten duen lorratza ikus daiteke, 31. irudian ezpainen irekiera maila erakusten da.

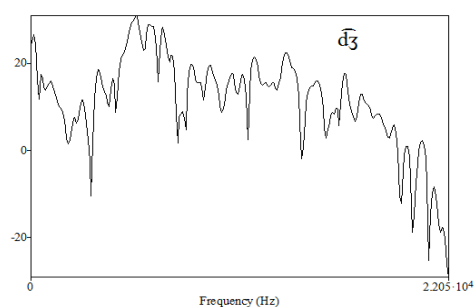
	
30. irudia: Mihiak uzten duen lorratza [d̪] egitean (Lekeitio)	31. irudia: Ezpainen irekiera maila [d̪] egitean.

[d̪] afrikatua erabiltzen duten informatzaileek berau berba hasieran eta, Bermeon esate baterako (Gaminde et al., 2012), sudurkariaren ostean baino ez dute erabiltzen. Bokalen artean egonez gero, afrikatua frikari bihurtzen da gorago aipatu dugun arauaren arabera.

Akustikoki, 32. irudiko espektrograman soinu honen adierazle akustikorik garrantzitsuenak ikus daitezke. 33. irudiko espektrograman energiaren kontzentrazioaren hasiera agertzen da. Datu akustiko zehatzetarako Gandarias et al.en (2014) lana ikus daiteke.

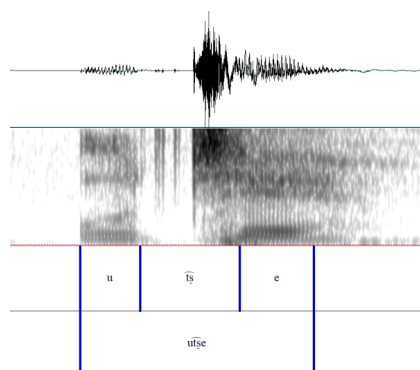
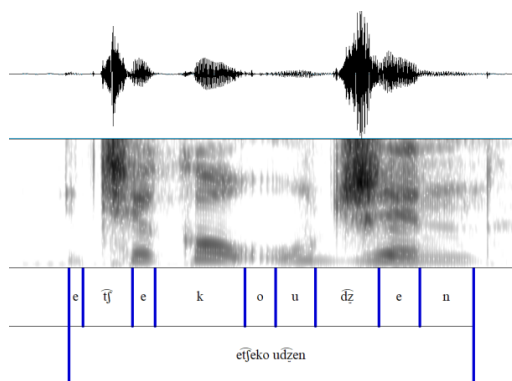


32. irudia: [d̪] alofonoaren adierazle akustikoak.



33. irudia: [d̪] ren FFT espektrogramaren itxura eta energiaren kontzentrazioaren hasiera.

Gorago esaneko moduan, [d̪] Arbizurako deskribatu da “ura” esateko (Hualde, 1996). 34. irudiko espektrograman Arbizuko esaldi batean agertzen da eta erkatu daiteke esaldi berean agertzen den [t̪]rekin. Herri batzuetan ahostunaren ordeztasuna [t̪] erabiltzen da; Lizarragako 35. irudiko espektrograman ikusten den bezala.

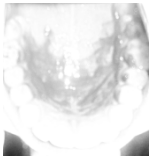



34. irudia: Arbizuko [d̪z̪]ren espektrograma.

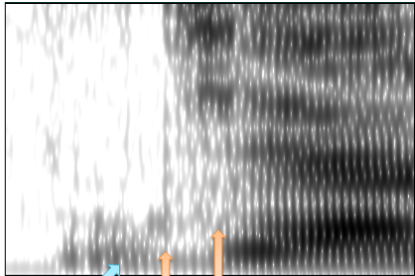
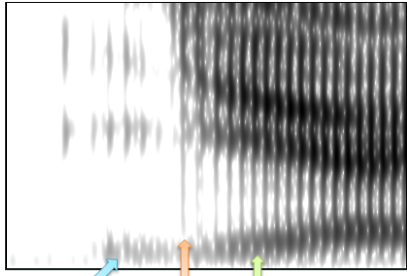
35. irudia: Lizarragako [t̪s̪]ren espektrograma.

Aditz partizipioen hasieran ager daitekeen [j] egile batzuen arabera (Mitzelena, 1977; Holmer & Abrahamson, 1968; Pagola, 1992; Oñederra, 2004) herkarizat jotzen da. Beste batzuek, sail honetan barik, afrikatuen sailean kokatzen dute (Sopelarako Markaida et al., 1993). Azkenik, egile batzuek bietako ezaugarriak dituela adierazten dute (Navarro Tomas, 1925). Hualdek (2003: 16) ahoskeraren arabera honako definizioa ematen du: “Voiced palatal approximant, fricative or stop”.

36. irudiko palatograman gauzapean afrikatuaren lorratzak erakusten dira. Bertan ikusten den moduan, aire-etorriaren ibilbidea guztiz itxita agertzen da. 37. irudian ezpainen irekiera maila agertzen da.

	
<p>36. irudia: [j]ren gauzapean aho sabaian uzten duen marka.</p>	<p>37. irudia: Ezpainen irekiera [j] egitean.</p>

Afrikatuaren gauzapenean glidea agertzea berorren adierazle akustiko garrantzitsutzat jotzen da (Martínez Celdrán, 2013: 273). 38. irudiko espektrograman ez da glidea agertzen eta 39. irudiko espektrograman bai; horrenbestez, lehena [j] herkaria dugu eta bigarrena [j̠] afrikatua.

	
<p>38. irudia: [j] herkariaren adierazle akustikoak (Zamudio).</p>	<p>39. irudia: [j̠] afrikatuaren adierazle akustikoak (Gatika).</p>

Gure iritiz euskararen bariazioaz dihardugunok badugu hemen arlo interesgarri bat ikertzeko, izan ere, afrikatu kopuruaren murrizteak aukera gehiago ekarri du euren gauzapenetan, Bermeon adierazi genuen bezala (Gaminde et al., 2012) eta afrikatuen banaketan faktore sozialek, hala nola generoak (Gaminde, 2010), estatusak, e.a.-ek garrantzi handia daukate.

5. Afrikatuen hedadura

Gure corpusean agertzen diren 600 hizkuntzetan afrikatuak, neurri batean edo bestean, 374tan (% 62,33) erabiltzen dira eta beste 226tan (% 37,67) ez. 6. taulan kontinenteka ematen ditugu afrikatuak dauzkaten hizkuntzen kopuruak eta ehunekoak; bertan ikusten den moduan, Asian eta Ameriketan afrikatuak dauzkaten hizkuntzen ehunekorik altuenak gertatzen dira eta Ozeanian, berriz, txikiena (40. irudiko mapa).

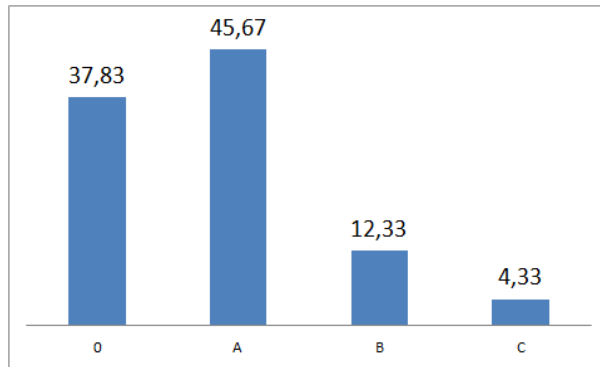
Kontinentea	Hizkuntzak	Ehunekoa
Afrika	125	69,44
Amerika	66	73,33
Asia	133	73,89
Europa	19	63,33
Ozeania	25	20,83

6. taula: Afrikatuak dauzkaten hizkuntza kopuruak eta ehunekoak kontinenteka.



40. irudia: Afrikatuak dauzkaten (gorririk) eta ez daukaten (urdinik) hizkuntzen hedadura.

Afrikatuen hedadura zenbaterainokoa den jakiteko kalkulu batzuk egin daitezke. Batetik zenbaki absolutuak erabil daitezke; afrikatuak agertzen diren hizkuntzetan euron kopurua 1etik 20ra bitartekoa izan daiteke. Hiru multzo eginez gero, hau da, (A) bat eta hiru bitartean, (B) lau eta sei bitartean eta (C) zazpitik gorakoak, ikusten dugu hizkuntza gehienak A multzoan kokatzen direla (274, % 45,67), gero B multzokoak datoz (74, %12,33) eta, azkenik, C multzokoak (26, % 4,33). 41. irudiko grafikoan multzo hauetako ehunekoak agertzen dira afrikaturik ez daukaten hizkuntzen ehunekoarekin batera (Euskara A multzoan agertzen da).



41. irudia: Multzo bakoitzeko agertzen diren ehunekoak.

Afrikatuak azter ditzakegu hizkuntzen inbentarioetan agertzen diren fonema kopuru osoen arabera, izan ere, ez da gauza bera afrikatu bakarra edukitzea inbentario txiki batean edo handi batean. Afrikatuak dauzkaten hizkuntzen azterketa eginda, ikusten dugu 1,69 eta 28,57 bitartean kokatzen direla. Lau talde egiten baditugu, A (1,64_8,37), B (8,38_15,1), C (15,2_21,83) eta D (21,84_28,57), hizkuntzak 7. taulan erakusten den bezala sailkatzen dira. Euskara B taldean kokatuta agertzen zaigu, afrikatuak % 10,71 direla.

Taldea	Hizkuntza	Ehunekoa
A	261	69,79
B	93	24,87
C	12	3,21
D	8	2,14

7. taula: Talde bakoitzeko hizkuntza kopuruak eta ehunekoak.

C taldean agertzen diren hizkuntzak Afrikan Duruma, Hadzera, Tivera eta Tsongera dira, Asian Akhera, Armeniera, Gyalrong, Karbiera, Mandarinera eta Tshangla, eta European Albaniera eta Poloniera. D multzokoak Afrikan Ronga, Tswanera eta Whe dira, Asian Chuanqiandan, Naxiera eta Nivkhera, eta Ameriketako Candoshi-Shapra eta Slaveyera (42. irudiko mapan erakusten dira).



Hizkuntzetan bokalen eta kontsonanteen kopuruak oso desorekatuta daudenez gero, kalkulua errepikatu dugu, baina hizkuntza bakoitzaren kontsonante kopuruaren arabera. Horrela eginda hizkuntzen ehunekoak 2,33 eta 34,29 bitartean kokatzen dira. Lau talde eginez gero, A (2,33_10,32), B (10,33_18,31), C (18,32_26,3) eta D (26,4_34,29), hizkuntzak 8. taulan ikus daitekeen bezala sailkatzen dira. Euskara B taldean kokatzen da, haren ehunekoa 13,04 izanik.

Taldea	Hizkuntza	Ehunekoa
A	224	59,89
B	122	32,62
C	20	5,35
D	8	2,14

8. taula: Talde bakoitzeko hizkuntza kopuruak eta ehunekoak.

Kasu honetan C taldean agertzen diren hizkuntzak hurrengoak dira: Afrikan Kutebera, Lucahzi, Ronga, Tivera eta Tsongera; Ameriketako Fulnio eta Navahoera; Asian Akhera, Armeniera, Burushaski, Gyalrong, Karbiera, Kaxmirera, Kok Borok, Tshangla eta Tuvera; Europan Albaniera, Poloniera eta Hungariera eta Ozeanian Isebe. D multzokoak Afrikan Tswanera eta Weh dira, Ameriketako Candoshi-Shapra eta Slaveyera, eta Asian Chuanqiandan, Mandarinera, Naxiera eta Nivkhera. (43. irudiko mapan ikus daitezke multzo biotako hizkuntzen kokaguneak).



6. Ondorioak

Ez da erraza honen moduko lantxo baten ondorioak laburbiltzea, hala ere, aipatu ditugun gauzarik nabarmenenak ekarriko ditugu hona beratu mukurutzeko.

Lehenengo eta behin esan behar dugu afrikatu mota gehienak aho barrunbearen zonalde zehatz batean gauzatzen direla, hau da, hortzetan eta albeoloen inguruan. Zonalde horretan afrikatuak horzkariak, albeolarrak, albeolo sabaikariak eta albeolo atzekoak edo post albeolarrak izan daitezke, ahoskune zehatzari dagokionez. Mihiaren erabilerari dagokionez apikariak, laminarrak, bizkarkariak eta eretroflexuak izan daitezke.

Euskararen kasuan egin diren ikerketen arabera, barietate geografikoak, sozialak eta estilistikoak aztertzean ez da arazo hau guztiz argitu, hau da, zenbat afrikatu eta zelan egiten diren ikerketa gai interesgarria da.

Euskara estandarren afrikatuak aztertzean munduko gainerako hizkuntzen aldean ikusten dugu alde batetik, afrikatuak besteetan gehien egiten diren modukoak direla, hau da, gauzatzen direla aho barrunbearen zonalde berean eta, bestetik, hedaturari dagokionez, hizkuntza askorekin batera B taldean agertzen dela, hau da, ratioaren arabera erdiko aldean, ez ratio txikienetan ez eta handienetan ere.

7. Erreferentziak

Aranbarri, F. (1996). *Ermua eta Eitzako euskara*. Ermua: Ermuko Udala.

Arretxe, J. (1994). *Basauriko euskara*. Basauri: Basauriko Udala.

Barrutia, E. (1996). *Bermeo eta Mundakako arrantzaleen hiztegia*. Iruña: Udako Euskal Unibertsitatea.

Dryer, Matthew S. & Haspelmath, M. (arg.) (2013). *The World Atlas of Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. Eskuragarri: <http://wals.info>, Accessed on 2018-10-25.

Errazti, A. (1994). *Iurreta elizateko euskara eta toponimia*. Iurreta: Iurretako Udala.

- Etxebarria, J.M. (1991.) *Zeberio haraneko euskararen azterketa etno-linguistikoa*. Euba: Ibaizabal.
- Fulop, S. A. (2011). *Speech Spectrum Analysis*. Berlin: Springer.
- Gaminde, I. (1992). *Urduliz eta Gatikako Herri hizkeren Azterketa Linguistikoa*. [Doktoretza Tesia]. Bilbo: Deustuko Unibertsitatea.
- Gaminde, I. (2001). *Mungia Berbarik Berba*. Mungia: Mungiaiko Udala.
- Gaminde, I. (2003a). Mungialdeko herskarien eta afrikatuen txandaketaz. *Euskalingua* 2:10-17.
- Gaminde, I. (2003b). *Zaldibar Berbarik Berba*. Zaldibar: Zaldibarko Udala.
- Gaminde, I. (2004). *Berango Berbarik Berba*. Berango: Berangoko Udala.
- Gaminde, I. (2007). *Bizkaian Zehar: Euskararen Ikuspegi Orokorra*. Bilbo: Mendebalde Kultura Alkartea eta Bizkaiko Foru Aldundia.
- Gaminde, I. (2010) *Bizkaiko Gazteen Prosodiaz: Euskaraz eta Gaztelaniaz*. Bilbo: Mendebalde Kultura Alkartea eta Bizkaiko Foru Aldundia.
- Gaminde, I., Goikoetxea, A., Sanz, Z., Astobieta, O., Elgoibar, E. & Gaminde, J. (2005) *Derioztarren Euskara*. Derio: Derioko Udala.
- Gaminde, I., Legarra, H. & Romero, A. (2012). *Gramatika eta Hizkuntz Bariazioa Bermeon*. Bermeo: Bermeoko Udala.
- Gaminde, I.; Unamuno, L., Iglesias, A. & Gandarias, L. (2013). Bizkaiko neska gazteen kontsonante afrikatuen izari akustikoez. *Euskalingua* 23, 6-13.
- Gandarias, L., Plaza, J. & Gaminde, I. (2014). Lekeitioko Txistukariez: Frikariak eta Afrikatuak. *Euskalingua* 24, 6-21.
- Gary, F. & Fennig, CH. D. (eds.) (2018). *Ethnologue: Languages of the World, Twenty-first edition*. Dallas, Texas: SIL International. Eskuragarri: <http://www.ethnologue.com>.
- Gilisasti, I. (2003). *Urduliz aldeko Berba Lapikokoa. Léxico del euskera de Uribe Kosta*. Bilbo: Euskara Mankomunitatea.
- Hammarström, H., Forkel, R. & Haspelmath, M. (2018). *Glottolog 3.3*. Jena: Max Planck Institute for the Science of Human History. Eskuragarri: <http://glottolog.org>, Accessed on 2018-10-25.
- Holmer, N. M. & Abrahamson, V. (1968). Apuntes vizcaínos. *ASJU* 2, 87-141.
- Hualde, J. I. (1996). Arbizuko hizkeraren zenbait soinu bereizgarriak. *Uztaro*, 18, 49-60.
- Hualde, J. I. (2003). Segmental Phonology. In J. I. Hualde & J. Ortiz de Urbina (arg.), *A Grammar of Basque*, Mouton de Gruyter, Berlin. 15-33.
- Hualde, J. I. (2010). Neutralización de sibilantes vascas y seseo en castellano. *Oihenart* 25, 89-116.
- Hualde, J. I. & Bilbao. X., (1992). A Phonological Study of the Basque Dialect of Getxo. *ASJU*, XXVI-1.
- Jongman, A, Wayland, R. & Wong, S. (2000). Acoustic characteristics of English fricatives. *Journal of the Acoustical Society of America* 108, 1252-1263.
- Ladefoged, P. & Maddieson, I. (1996). *The sounds of the world's languages*. Oxford: Blackwells.
- Lekeitioko Berbak Lantaldea (2001). *Itsasoa*. Lekeitio: Lekeitioko Udala.
- Lev, M., Stark, T., Clem, E. eta Chang, W. (2015). South American Phonological Inventory Database v1.1.4. Survey of California and Other Indian Languages Digital Resource. Berkeley: University of California.
- Maddieson I., Flavier S., Marsico E., eta Pellegrino F. (2018). LAPSyD: Lyon-Albuquerque Phonological Systems Databases, Version 1.1. Hemendik berreskuratuta: <http://www.lapsyd.ddl.cnrs.fr/lapsyd/>.
- Markaida, E., Gaminde, I. eta Markaida, B. (1993). *Sopelako Euskaraz*. Sopela: Urdulizko Euskara Taldea.

- Martín Celdrán, E. (1989). *Fonética*. Barcelona: Teide.
- Martín Celdrán, E. (2013). Los sonidos obstruyentes en la cadena hablada. In Penas, M^a. A. (arg.), *Panorama de la fonética española actual*, Arco/Libros, 253-289.
- Mitxelena, K. (1977). *Fonética histórica vasca*. ASJU.
- Moran, S.; McCloy, D. eta Wright, R. (arg.) (2014) *PHOIBLE Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. Hemendik berreskuratuta: <http://phoible.org>, Accessed on.
- Navarro Tomás, T. (1925). Pronunciación guipuzcoana. Homenaje a Menéndez Pidal III. Madrid, 593-653.
- N'Diaye, G., (1970). *Structure du dialecte basque de Maya*. Paris : De Gruyter Mouton.
- Nissen, S. L. eta Fox, R. A. (2005). Acoustic and spectral characteristics of young children's fricative productions: A developmental perspective. *Journal of the Acoustical Society of America* 118, 2570-2578.
- Oñederra, M. L. (2004). *Fonetika Fonologia Hitzez Hitz*. Bilbo: UPV/EHU.
- Pagola, R. M. (1992). *Euskal Fonetika Nafarroan*. Iruñea: Nafarroako Gobernua.
- Quilis, A. (1993). *Tratado de Fonología y Fonética españolas*. Madrid: Gredos.
- Salaburu, P. (1984.) *Hizkuntz teoria eta Baztango euskalkia: fonetika eta fonologia. Hizkuntzaren soinu-egitura*. Bilbo: UPV/EHU.
- Salazar, B. (2001) *Ugaoko Euskara*. Bilbo: Ediciones Beta.
- Silbert, N. H., eta Jong de, K.J. (2008). Focus, Prosodic Context, and Phonological Feature Specification: Patterns of Variation in Fricative Production., *Journal of the Acoustical Society of America* 123, 2769-2779.
- Txillardegi, (1980). *Euskal Fonologia*. Donostia: Ediciones Vascas.
- UCLA Phonetics Lab Data (2018). Hemendik berreskuratuta: <http://www.phonetics.ucla.edu/index.html>.
- Uriarte, J. A. (1995), *Fruizko Hizkera: Azterketa Linguistikoa*. [Doktorego tesia]. Bilbo: Deustuko Unibertsitatea.
- Urrutia, H., Etxebarria, M., Turrez, I. & Duque, J. C. (1988). *Fonética Vasca I. Las Sibilantes en el Vizcaino*. Bilbo: Deustuko Unibertsitatea.
- Urrutia, H., Etxebarria, M., Turrez, I. & Duque, J. C., (1989). *Fonética Vasca II. Las Sibilantes en el Guipuzcoano*. Bilbo: Deustuko Unibertsitatea.
- Urrutia, H., Etxebarria, M., Turrez, I. & Duque, J. C., (1991). *Fonética Vasca III. Las Sibilantes en los dialectos Orientales*. Bilbo: Deustuko Unibertsitatea.
- Yárnoz, M. B. (2001). *Sibilants in the Basque Dialect of Bortziri: An Acoustic and Perceptual Study*. Philipps-Universität Malburg.

THE SEMANTIC PROFILES OF OLD ENGLISH PRETERITE-PRESENT VERBS OF COGNITION- *WITAN* AND *CUNNAN*

Agnieszka Wawrzyniak
The Department of Pedagogy and Fine Arts, Kalisz, Poland
agnieszka78kaga@wp.pl

Abstract

The present paper offers a historical account of the semantic profiles of Old English verbs of cognition, namely of *witan* and of *cunnan* with a particular attention paid to the emergence of the epistemic senses in *witan* and the lack of epistemicity in *cunnan*. The contexts for the analysis of the OE verbs were taken from *The Toronto Corpus*. Moreover, the paper attempts to analyse the variety of concepts evolving on the path to epistemicity in *witan* and to account for the factors which precluded the emergence of such senses in *cunnan*. The present study aims to show that by relying solely on metaphor itself, the process of the emergence of epistemicity would be distorted. Instead, the paper offers the approach of a gradual shift via metonymy and metaphor in the light of the grammaticalization process.

Key words: cognition, epistemic, grammaticalization, profile, reanalysis

1. Introduction

The aim of the present paper is to analyse two Old English preterite-present verbs, *witan* and *cunnan*, the first of which disappeared in the 19thc., except in the phrase *to wit* 'namely' while the second one developed into PDE modal *can*. The study indicated that the two analysed verbs belonged to the verbs of cognition in Old English and that they underwent different conceptual paths of development. Hence, the underlying assumptions of the analysis are as follows:

Firstly, the study shows that despite similar meanings in Old English, *witan* and *cunnan* underwent different conceptual lines of development, which started in Old English period.

Secondly, the study applies to the contemporary theory of metaphor and metonymy postulated by Barcelona (2003), Radden (2003), Koivisto-Alanko (2000) among others to analyze the semantic profile of the these verbs, thereby bridging historical linguistics and modern semantics.

As for *witan*, the study will prove that the morphological reanalysis, whereby the past has acquired a present meaning was semantically motivated. The paper will investigate the semantic consequences of the Proto-Indo-European **weid*, denoting 'to see', and its perfect form **woid-/wid-*, 'I have seen', reinterpreted as 'I know' upon the meaning of the Old English verb *witan*.

Secondly, the paper will also illustrate that the attributes related to perception are present in OE *witan*. In other words, the semantic reanalysis, whereby not only the form but, above all, the meaning changed, took much more time than previously assumed, extending well into the Old English period.

Moreover, the present study will also indicate that the development of *witan* goes parallel to the subjectification process, thereby reflecting tendencies of directionality exemplified by Traugott (1989) with regard to the grammaticalization process. Hence, the directionality of the semantic change projected by *witan* reflects a path of change from the initial, objective meaning of knowing underpinned by perception, thus from the real-world, more concrete domain to the more abstract, epistemic one, which focuses on the internal world of speaker's belief and knowledge status (Traugott, 1989: 37).

In the analysis of the semantic profile of Old English *cunnan*, the aim will be to account for factors which precluded the development of epistemicity in the Old English period. Hence, the paper will juxtapose the semantic developmental paths of *witan* and *cunnan*, in order to show the gradual emergence of epistemicity in *witan* and the lack of epistemic senses in *cunnan* in the Old English period.

The present analyses of OE verbs are present on *The Toronto Corpus* compiled by Antonette di Paolo Healey (1986). The Corpus is an online database consisting of about three million words of Old English. The paper presents a quantitative approach based on detailed statistics of senses within the concepts of the analysed verbs. In order to achieve a maximum accuracy, the two preterite-present verbs were analysed in all attested contexts from *The Toronto Corpus*.

2. The semantics of OE *witan*

Following the online Etymological Dictionary (s.v. *witan*), *witan* is related to Lat. *videō* 'see', Lith. *veizdmi* 'see', ON. *vita* 'know', and Goth. *witan* 'know'. The etymological form of the Old English verb *witan*, namely **weid*, denoted both 'to see' and 'know as a fact'. Thus, from the diachronic perspective the etymon of *witan*, namely *weid-* denoted physical perception. The analysis will show that the sense of physical perception was also the part of the semantic profile of OE *witan*.

The semantic analysis of the infinitive *witan* carried out on texts from *The Toronto Corpus* records the following senses:

senses	tokens
the independent sense of knowing	93, 70%
seemingly tautological expressions	17, 13%
fixed expressions functioning as discourse markers	14, 10.5%
sense of knowing underpinned by perception	6, 4.5%
know followed by direct object	1, 0.75%
know how to	1, 0.75%
senses implying 'make somebody know'	1, 0.75%

2.1. Sense of knowing underpinned by perception

The sense of knowing underpinned by perception can be exemplified by the following contexts:

- (1) *We magon gesēon and witan witodlice þæt gē ure lāreowas wæron* (Belfour 1909, 96-106)
(We may see and truly know that you were our teachers).
- (2) *Hīe magon witan þone hīe gesēoþ.* (Ascension: Text from Clemoes 299- 314)
(They may know what they see).
- (3) *Ac hwylce men hī ne magon þā ungesewenlica þing witan þæt hī ne gesēoþ mid heora lichamlicum ēagum.* (Gregory the Great, Preface and Book 4: Hecht 1900-7, 260-350)
(But some men cannot know the invisible thing they cannot see with their bodily eyes).

The above listed contexts indicate that the underlying meaning of the infinitive is the one of perception; to know as a result of seeing. Here, Traugott's (1989) Tendency I (meaning based in the external described situation > meaning based in the internal situation) can be tested to achieve some interesting results. Such a concept of *witan* can be viewed as a result of the metaphorical extension from the field of perception to the field of conceptualization; to know as a result of seeing whereby the concrete, physical activity (the source) is transferred upon the abstract one (the target), the final stage of which is a higher and a more abstract layer of representation. *Witan*, when underpinned by perception, is not an independent sense referring to a pure, abstract act of conceptualization, but a one with which 'seeing' is imbued. This concept constituted the intermediate link to the development of the independent sense of knowing. At that stage, the emerging sense of knowing becomes linked with the sense of perception. Hence, what took place was not an abrupt shift from the Proto-Indo-European sense of seeing to the Old English sense of knowing, but rather a gradual process. In the analyzed contexts, the associations of vision and intellection are evoked together, hence one knows something only if one has seen it. Radden (2003), while discussing the development of the sense of knowing from the Proto-Indo-European sense denoting perception, refers to the

process of the emerging metaphor, which is a continuum of metonymically related senses due to context-induced interpretations. What takes place in the subjective development of the concept of *witan* is the conceptual relationship between a named and an implicated entity based on contiguity, or metonymy. As Radden maintains (2003: 98), this conceptual relationship between Proto-Indo-European **weid* and Old English *witan* can be viewed as based on 'sequential events', or effect and result. Radden continues that the sequential ordering of the two events gives rise to an element of causality, hence 'I know because I saw'. In other words, the two events are 'temporally and causally linked' (Radden 2003) as to know something entails the concept of having seen it. At that point, Radden claims (2003: 99), that the two domains of 'time' and 'causality' have a common experiential basis, namely PRECEDENCE PLUS CAUSE and SUBSEQUENCE PLUS RESULT, and may be interpreted as CAUSAL PRECEDENCE. In other words, according to Radden (2003), the process accountable for the development of the epistemic sense in *witan* is a metonymy-based metaphor. As he maintains (2003: 95), metaphors which are grounded in metonymy are more basic and natural than those which do not have a metonymic basis. He further adds that with these (metaphors grounded in metonymy), metonymy provides an associative and motivated link between the two conceptual domains involved in metaphor. Moreover, the mechanism metonymy-based metaphor can be explained by the fuzziness of the boundary lines, which make metaphor and metonymy interpenetrate. In his analysis of *witan*, Radden claims that the sequential ordering of the two events I know it because I saw it, hence EFFECT standing for the CAUSE constitutes metaphorical basis for the metaphor in which the two domains of perception and cognition are mapped. Consequently, the distinction between metaphor/metonymy can be viewed as scalar rather than discrete (Barcelona, 2003: 16).

The creation of an implicature between a named and implicated entity takes place in the mind of the conceptualizer. Therefore, Traugott's Tendency I (1989) takes place as a result of metonymically related changes.

The interaction between metaphor and metonymy has also been discussed by Koivisto-Alanko (2000)

Metonymical change is prototypical flexibility producing meanings and meaning nuances, some transient, some resulting in lasting metonymous meaning and some leading to actual metaphorical change. The metaphorical level of change can be observed only in 'the big picture' not on the semantic micro level.

2.1. *Witan* in pragmatic contexts

The Toronto Corpus records 39 instances, where *witan* performs a pragmatic function. It was not until the 11th c. when *witan* started to act as a discourse marker in a fixed phrase constituted by *witan* preceded by *wolde*, i.e., *willan* in the past subjunctive. Moreover, the most frequent pronoun in these pragmatic expressions is *ic* (PDE 'I'), which constitutes 28 instances out of 39 recorded in the analyzed texts. *Wāst*, related to the second person singular

þū (PDE 'you'), is attested in seven contexts. The least frequent juxtaposition is of *witan* and a third person pronoun *hē* (PDE 'he'), which is recorded 4 times. *Witan* in pragmatic contexts can be illustrated with the following examples:

(4) Ic wolde witan hweþer þū for heora lufum þas þing underfōn. (Boethius: The Consolation of Philosophy: Sedgefield 1899, 3-6)

(I would like to know /I was wondering if you would suffer this for her love).

(5) Ac ic wolde witan swā be gōde on mīnre. (St. Augustine, Soliloquies, Book 1: 1922, 65-70)

(I would like to know/I was wondering what is good to me).

Such expressions exemplify Traugott's Tendency II (1989). The above phrases can be viewed as corresponding to the Present-Day English expressions 'I/you/he was/were wondering'. They are automatic, fixed phrases whose function is purely emphatic. Thus, the analysis of *witan* in the above phrases requires a different approach than in the previously analyzed examples.

To begin with, *witan*, when juxtaposed with *willan* in the subjunctive form, has developed into a discourse marker, thus representing a change from a mental state to a speech-act verb meaning. According to Traugott (1989: 43), speech-act verbs are epistemic as they express beliefs about the truth of the proposition. In that sense, they either describe the situation or are constitutive of the discourse situation. She further maintains (1989: 44), that when both non-speech-act verb and speech-act verb meanings coexist, the former precedes the latter.

The problem of the pragmaticalization of discourse markers has also been discussed by Aijmer (1985: 2), for whom discourse markers are pragmaticalized since they involve the speaker's attitude to the hearer. Moreover, as she further emphasizes, pragmaticalization of meaning has often been associated with 'bleaching' or the loss of meaning. However, instead of using the term 'bleaching' as it evokes negative associations, it seems better to talk about the 'replacement' of one, objective dimension by another, which is subjective and epistemic. *Witan*, in the analyzed phrase, is devoid of the sense 'to gain information', commonly linked with the verb 'to know', and permits the extension of meaning involving the speaker's attitude to the message. As a result, one can observe the development from a lexical item into an introductory phrase starting the discourse, hence into a purely pragmaticalized expression.

Additionally, Hopper and Traugott (1993) emphasize the process of specialization and divergence while discussing the problem of pragmaticalization. By specialization, Hopper and Traugott mean that lexical forms and structures become more frequent or the only choice in the whole language (Hopper and Traugott, 1993: 113). That is why *witan* in the above expressions ceases to be equipped with the semantic content 'to get information', where it occurred with all persons, and becomes mostly used with the first person singular (only sporadically with the second and third person singular), and with *willan* in the past subjunctive.

Regarding divergence, Hopper and Traugott (1993: 113), indicate that when a lexical form undergoes pragmaticalization, the original lexical form may remain as an autonomous lexical item. Therefore, *witan*, apart from being a fixed item with a different set of attributes in the

analyzed pragmatic expressions, behaves like an autonomous lexical item when in other contexts.

2.2. *The independent sense of knowing*

The independent sense of knowing can be exemplified by the following contexts:

(6) Witodlice *witan* wē mōton þæt Crist for ūs þafode. (*Antichrist: Bethrum 1957, 116*)

(Certainly we must know that Christ suffered for us).

(7) Wē sceolon *witan* þæt nān sylcwala þæt is āgenslaga ne becymþ to Godes rīce. (*De duodecim abusivis: Morris 1868, 296-304*)

(We should know that no person committing suicide can be accepted to the heavenly kingdom).

(8) Gē sceolon *witan*, gif gē wīsdōm lufiaþ, hwæt sý betwux þam twām gecyþnyssum, þære ealdan æ, ær Cristes tōcyme, and þære nīwan gecyþnysse, under Cristes gife. (*First Old English Letter for Wulfstan: Fehr 1914, 68-145*).

(You should know, if you love wisdom, what was between the two testaments, the Old testament before Christ's birth, and the new one after Christ's birth).

The above mentioned sentences are examples of an *emerging metaphor* (Radden 2003), hence the process of metonymy-based metaphor. They signal the completion of Tendency I (Traugott 1989). The reasons for allocating the above change to the process of metonymy-based metaphor are the following. First of all, the effects of the change are metaphorical, yet the steps that lead to that effect are purely metonymic, where the result stands for the cause, or in other words, SUBSEQUENCE PLUS RESULT stands for PRECEDENCE PLUS CAUSE. However, when viewed from a perspective, the change that takes place bridges the two conceptual domains, where its onset is the socio-physical level while the final stage is the abstract, subjectively coloured sense 'to know', no longer rooted in the concrete domain. Furthermore, the above sentences that mark Traugott's Tendency I, should be viewed as epistemic as they code a speaker's commitment to the utterance rather than focusing on the real-world domain. Hence, one should indicate that the process of an *emerging metaphor* marked a new concept, namely the concept of 'knowing', which even though still concrete in the intermediate phase of Tendency I when underpinned by perception, acquired evaluative, internal meanings by the end of Tendency I.

2.3. *Apparently tautological expressions*

The semantic analysis of *witan* illustrates also contexts, which can be referred to as apparently tautological as they co-occur with prepositional phrases or adverbial phrases that centralize the domain of wisdom and cognition. *Witan* occurred in such phrases as:

- *witan gewiss*
- *witan mid wīslīcum geþeahte*
- *wīslīce witan*

- witan mid fullum gerade
- witan mid dēopancum mōde
- witan mid ingeþance
- geare witan
- wīslīcor witan

Approaching these contexts, one can observe that *witan* is applied here to the act of perceiving rather than knowing, otherwise the above contexts would be tautological. The above senses should be localized as an intermediate stage within Tendency I.

3. The analysis of the profile of OE *cunnan*

According to the OED (s.v. *can*), Old English *cunnan* is related to ON *kunna*, and Du *kunnan*, whose meaning was 'to be able to'. The common Indo-European base for all these cognates is **genē-*, **genō-*, which meant 'to be able mentally', 'to know'. As illustrated in the OED (s.v. *can*), the Proto-Germanic sense of the verb was 'to know', 'to know how to', 'to be mentally and intellectually able', 'to be physically able', 'have power'. As further illustrated, the present was formally a preterite, hence the meaning of 'I know' must have derived from that of 'I have learnt', 'I have attained to knowledge'.

As the OED indicates (s.v. *can*), OE *cunnan* could be analyzed along the following senses: 'to know', hence 'to know or be acquainted with (a person)', 'to know or have learned (a thing), to have practical knowledge of'. Other senses related to *cunnan* in the OED are: 'to know how to, to have learned, to be intellectually able, to be able; to have power, ability or capacity' (said of physical as well as mental and of natural as well as acquired ability).

The analysis of the semantics of Old English *cunnan* based on *The Toronto Corpus* records the following contexts:

	10 th c.	10/11 th c.	11 th c.
know	5 38%	1 4%	3 3%
be familiar with	4 31%	13 46%	48 43%
know how to	4 31%	14 50%	60 54%

3.1. The semantic analysis of *cunnan* in the 10th c.

The analysis of *cunnan* in the 10th c. leads to the following conclusions. To begin with, the sense of 'to know' is the most frequent one and covers 5 cases (38%)

- (1) Ic sceal hraþe *cunnan* hwæt to ūs duguþum gedōn wille. (*Andreas: Krapp 1932a, 3-51*)
(I shall swiftly know what you are willing to do for splendor).
- (2) Nū gē geare *cunnan* hwæt se hlāford is. (*Christ: Krapp and Dobbie 1936, 3-49*)
(Now you completely know what Lord is).

Regarding the sense 'be familiar with', the analyzed texts point only to four instances of the above sense. Moreover, three of these forms have an intellectual/perceptual basis, while the other is related only to the concept of familiarity:

intellectual/perceptual basis

(3) *Đū canst ealra gehygdō, and mōd. (Andreas: Krapp 1932a, 3-51)*

(You alone know every thought and soul).

(4) *Đū canst worda wīslīc andgīt. (Andreas: Krapp 1932a, 3-51)*

(You know the proper understanding/sense).

(5) *Nū gē gefyrhsefan and mōdgeþanc mīnne cunnon. (Elene: Krapp 1932a, 66-102)*

(Now you are familiar with my thoughts).

familiarity

(6) *Ðæt sceal æglāwra mann of moldan findan, þæt fram fruman cunne eall þā earfode hē mid ealne ādrēah. (Andreas: Krapp 1932a, 3-51)*

(That shall man of law on Earth find out, that knew from the beginning all the hardships he boldly suffered).

The usage of the sense 'know how to' rendered by *cunnan* points to 4 cases (31%). At that point, the use of *cunnan* in the sense 'know how to' was marginal, yet the verb collocated with verbs not connected with the intellectual sphere. *The Toronto Corpus* records four verbs used with *cunnan*: *findan*, *reccan*, *secgan*, *wyrcean*. This sense can be exemplified by the following contexts:

(7) *Ðe wræstlicost wyrcean cūþon, on þām stedewange gōdes tempel. (Elene: Krapp 1932a, 66-102)*

(They could work most gloriously a God's temple on the open land).

(8) *Ænige ne fugleare cūþen sweotole gesecgan be þām sigeþeacne. (Elene: Krapp 1932a, 66-102)*

(Nobody could quite well speak of the Cross).

The analysis indicates that *cunnan* in the 10th c. denoted cognition (senses- 'to know' and 'be familiar with') as well the general ability type ('know how to'). Yet, the use of *cunnan* in the sense 'know how to' was marginal in the 10th c. and the verb was mainly used to convey either the sense 'to know' or 'be familiar with', the latter one having mainly intellectual/perceptual basis.

3.1. The semantic analysis of cunnan in the 10/11th c.

The analysis of *cunnan* carried out on the texts from 10/11th c. from *the Toronto Corpus* illustrates a general increase in the use of the above verb. Moreover, the number of contexts in which *cunnan* collocates with verbs of intellectual ability increases. Hence, one should emphasize that the contexts where *cunnan* stands in harmonic relationship with verbs of intellectual ability are frequent. Yet, at that stage the verb ceases to be used in the sense

denoting 'to know'. *The Toronto Corpus* records only one instance of this sense in 10/11th c. recorded in *The Rushworth Gospels*:

(9) Gē ne cunnon hwilc dæg oþþe hwilc tīd drihten cymþ. (*The Rushwoth Gospels*: Skeat 1871-87, 13-187)

(You know neither the day nor the time when God comes).

As for the sense 'be familiar with', its frequency is higher when compared with the 10th c. This sense constitutes 46% of contexts in the 10/11th c., while in the 10th c. it amounted to 31%. Moreover, in the transition stage (10/11th c.), the concept of the intellectual/perceptual sphere is highly recorded in the sense 'be familiar with'. Hence, the contexts of *cunnan* denoting 'be familiar with' could be rendered as 'be intellectually or perceptually familiar with'. In other words, at that time, there are no contexts where *cunnan* could denote a general, every-day sense of knowing without implying deeper intellectual/perceptual cognizing, which could be exemplified by the following contexts:

(10) Hī *cunnon* rihtlice heora fulluhtes gescēad. (*Baptism: Bethurum 1957, 172-4*)

(They are rightly familiar with the sense behind baptism).

(11) Sē þe *cann* þæra stafa gescēad, hē herað heora fægenysse, and ræt þā stafas, and undrestent hwæt hī gemenað.

(The one who is familiar with the sense behind letters, he praises their beauty and reads the letters and understands the message they convey).

Regarding the sense 'know how to', the texts in *the Toronto Corpus* contain 14 cases (50%) in the transition stage (10/11th c.). The sense 'know how to' could denote either 'mentally know how to', or 'know how to' reflecting the general everyday use. The recordings from the 10/11th c. indicate 50% of senses denoting intellectual/perceptual ability and the every-day sense of ability, respectively, which could be exemplified by the contexts:

intellectual/perceptual ability

(12) Sē sācerd scolde and git sceal simlesmēalice geþencan þæt hē *cunne* gōd and yfel tōscēadan. (*Gregory The Great, Pastoral Care: Sweet 1871, 24-467*)

(The priest should, and yet shall not very carefully consider, that he would be able to distinguish between good and evil);

every-day sense of ability

(13) Sume men singan ne cunnon. (*Purification: Text from Clemoes 1955-6, 133-47*)

(Some men cannot sing)

One should emphasize that the semantic attribute of intellect/intuition was highly present in the 10/11th c. It set the conceptual framework to the sense of familiarity, and constituted 50% of contexts of the sense 'know how to'.

3.2. *The semantic analysis of cunnan in the 11th c.*

The analysis of contexts in the 11th c. indicates that the verb continued to be used more in the sense of general rather than purely intellectual/perceptual ability. To begin with, *the Toronto Corpus* records only three cases (3%) of *cunnan* denoting 'know'. Moreover, these instances seem tautological. Hence, *cunnan* occurs in phrases

- a) followed by an adverb denoting the process of cognition - *cunnan gemindlice*
- b) followed by a proposition and noun denoting cognition- *cunnan mid gescēade*
- c) supported by a verb denoting cognition - *Ic cann oþþe ic wāt*

(14) Gē sceolon *cunnan* gemindelice and mannum eac secgan þā tynænlican word þē Gōd tæhte.
(*Second Old English Letter for Wulfstan: Fehr 1914, 196-221*)

(You shall wisely know and tell people the edifying words God taught).

(15) Ðæt is to witanne þam þe þis wyle *cunnan* mid gescēade þæt synd twelf tæcna.

(*Byrthferth's Manual: Crawford 1929, 2-238*)

(That is to know to those who would be willing to know with wisdom that there were twelve signs).

Regarding the sense 'be familiar with', the statistics indicate 37% (18 cases) of contexts denoting the every-day use of familiarity out of 48 cases. Even though the number of contexts denoting intellectual/perceptual familiarity is still higher, one should notice the extension of the field of familiarity, when compared with the 10/11th c., hence the weakening of the semantic attribute of intellect/perception in the concept:

(intellectual/perceptual sphere)

(16) Ælc Cristen man sceal *cunnan* his pater noster and his crēdan (*Ash Wednesday: Skeat 1881-1900, I, 260-82*)

(Each Christian should be familiar with our Father and Creed);

every-day use, familiarity

(17) Cunne gē ænigne ma þe hatte Nicolaus? (Saint Nicholas: Dictionary of Old English transcript, edited from Cambridge, Corpus Christi College, MS 303)

(Do you know a man that is called Nicolaus?)

(18) Gif gē cunnon þe þā yfele synd. (In letania maiore: Text from Clemoes 1955-6, 249-63)

(If you know those who are evil).

As for the sense 'know how to', it has been recorded in 54% of contexts (60 cases), yet the sense 'to be generally able to', without implying intellectual undertones, is recorded in 37 cases (61%), thereby outnumbering the sense 'to be mentally able to';

to be intellectually able

(19) Wē cunnon tocnāwan þā dēopnyse his swycolan mōdes. (Addition to Lives of Saints no. 17: De auguris: Pope 1967-8, 790-96).

(We are able to perceive the depth of his deceitful soul).

to be generally able

(20) Ne *cann* ic his mōd āwendan tō þām upplicum þingum. (*Assumption of the Virgin: Godden 1979, 255-9*)

(I am not able to turn his soul to religious matters).

(21) To twegen gebrōþra þe wel þēnnian *cunnon*. (*Benedict, Rule: Schroumler 1885-8, 1-133*)

(To two brothers who are able to serve well).

Approaching the general development of *cunnan*, one can notice that it belonged to the class of verbs in which epistemicity did not emerge. It should be emphasized that factors that account for the lack of subjective undertones in *cunnan* lie in the semantic path of the verb. Namely, *cunnan* indicates only internal ability; either intellectual or general. Moreover, in all of the analysed contexts *cunnan* denoted only intrinsic ability, thus the ability that resided within the subject (Gamon, 1993: 126). The concept of ability profiled by *cunnan* did not rely on external factors that could enable the agent complete the action or the event. In other words, contingent ability was not projected by *cunnan*. Consequently, the concept of ability could not evolve into the concept of possibility that necessitates the presence of external factors. Hence, illustrating the axis of development in *cunnan*, one can notice that its semantics proceeded from the verb when the sense of ability was the most rooted in the intellectual sphere to the general ability sense which lied within the ability of the subject.

4. Conclusions

To conclude, the present paper has been an attempt to carry out an analysis of two OE verbs of cognition, namely *witan* and *cunnan*.

With regard to *witan*, the study showed that the semantic attributes of the Proto-Indo-European **weid-* are still well present in the OE concept of *witan*, though to a different extent in different morphological forms, which also finds a semantic justification.

In addition, while the morphological reanalysis can be approached synchronically, from a definite point in time, the semantic reanalysis started much earlier, and extended well into Old English, thus after the morphological reanalysis had been completed. Therefore, the morphological reanalysis did not mark a new concept but only the onset of the evolution of the new concept, as the early link of Old English preterite-present verb with its Proto-Indo-European ancestor is undeniable. The paper has thus illustrated the semantic analysis of *witan* as a continuing rather than a complete process, pointing to the intermediate instances where the sense of knowing was underpinned by perception.

Besides, the aim of the paper has been also to show that the subjectification process that affected OE verb *witan* marked the shift from the concrete to the abstract, logical and subjective, thereby setting the ground for the further grammaticalization that took place in Middle and Early Modern English. To put in Traugott's terminology, the completion of

Tendencies I and II in Old English resulted in the creation of new meanings, based in the internal situation, and in the textual one, which in turn facilitated the further grammaticalization marked by Tendency III in the course of the development of the verb in the subsequent periods.

As for OE *cunnan*, the analysis showed that this verb was not affected by the process of epistemicity. Approaching the development of *cunnan*, one can notice that it belongs to the class of verbs in which epistemicity did not emerge. Moreover, its prototypical sense in the 10th c. was 'to know', which gradually was becoming replaced by the sense of ability. It should be indicated that the factors that account for the lack of subjective undertones in *cunnan* lie in the semantic path of the verb. Namely, *cunnan* indicates only intrinsic ability; either intellectual or general, but there is no split into intrinsic/contingent ability. Therefore, *cunnan* did not evolve into the sense of probability. Illustrating the semantic path in *cunnan*, one can notice that it proceeded from the sense of ability mostly rooted in the intellectual sphere to the general ability sense which lied within the ability of the subject.

5. References

- Aijmer, K. (1985). The semantic development of will. In J. Fisiak (ed.), *Historical semantics and historical word formation* (pp. 11-21). Berlin: Mouton de Gruyter.
- Barcelona, A. (2003). *Metaphor and metonymy at the cross-roads. A cognitive perspective*. Berlin: Mouton de Gruyter.
- Gamon, D. (1993). On the development of epistemicity in the German modal verbs *mögen* and *müssen*. *Folia Linguistica Historica* 14, 125-176.
- Healey, A. di Paolo. (ed.). (1986). *Dictionary of Old English Corpus*. Toronto: The University of Toronto Press.
- Hopper, P., & Traugott, E. C. (1993). *Grammaticalization*. Cambridge: Cambridge University Press.
- Koivisto, A. (2000). Patterns of semantic change in abstract nouns. The case of *wit*. In M. Rissanen, T. Nevalainen, P. Päivi and I. Taavitsainen (Eds.), *History of Englishes* (pp. 313-327). Berlin: Mouton de Gruyter.
- Online Etymological Dictionary*. Available at: [http://www. etymonline.com/](http://www.etymonline.com/)date of access: August 2019.
- Radden, G. (2003). How metonymic are metaphors. In A. Barcelona (ed.), *Metaphor and metonymy at the cross-roads. A cognitive perspective* (pp. 92-108). Berlin: Mouton de Gruyter.
- Traugott, E. (1989). On the rise of epistemic meaning in English. *Language*, 65, 31-55.

AHOZKO TESTUEN PROSODIAREN BILAKAERA JATORRIZKO INGURUNETIK KANPO. EXTREMADURAKO GAZTELANIAREN KASUA BIZKAIAN (BASAURIN)¹²

Irati de Pablo¹, Iñaki Gaminde² eta Naia Eguskiza³

^{1,2,3} Euskal Herriko Unibertsitatea (UPV/EHU)

¹irati.depablo@ehu.eus, ²inaki.gaminde@gmail.com

³naia.egusquiza@ehu.eus

Laburpena

Lan honen helburua da aztertzea familia bereko hiru belaunalditan zenbat aldatu diren ahozko testuen ezaugarri prosodiko batzuk. Horretarako, silaba motak, talde prosodikoen luzera, silaben iraupena eta silaben iraupenaren eta hizketa abiadaren arteko erlazioak aztertuko dira. Helburu hori abiapuntutzat hartuta, familia bereko 12 informatzaile aukeratu dira: batzuk Extremaduran jaiotakoak dira, nagusitan Euskal Herrira etorriak; beste batzuk han jaio eta umetan hona etorri dira eta, azkenik, beste zenbait bertan sortu dira. Ikerketa honetan erabili den metodologiari dagokionez, informatzaile bakoitzari komiki bera erakutsi zaio denbora batez uler dezan eta, gero, bertan ikusitakoa azaltzea eskatu zaion. Berbaldi hori audio-grabazioen bidez erregistratu eta azterketarako corpusa osatzeko erabili da. Emaitzen artean alde estatistiko esanguratsuak aurkitu dira silaba motetan eta silaben hasieretan ere bai belaunaldi eta jaioterriaren arabera. Talde prosodikoen luzerak, artikulazio-abiadak eta silaben iraupenak alde esanguratsua dute generoaren arabera.

Hitz gakoak: aldakortasuna, soziolinguistika, prosodia.

Abstract

The purpose of this work is to observe how do some prosodic characteristics of the oral texts change along three different generations, all from the same family. For that, it has been analysed the types of syllables, length of the prosodic groups, duration of the syllables and the relations between the duration of the syllables and speed of speech will be analysed. With this goal in mind, 12 informants from the same family have been chosen; some born in Extremadura and who have come to live in Biscay as a senior citizen; others were born there and have come to Biscay when they were young to Biscay and, finally, others have been born in Biscay. Regarding the methodology used in this investigation, all the informants have been shown the same comic and they have been

¹² Ikerketa hau burutu da IT1028/19 (Eusko Jaurlaritza) eta GIU 16/22 (Euskal Herriko Unibertsitatea) proiektuen barnean eta UPV/EHUK doktoretza aurreko formakuntzan dagoen pertsonala kontratatzeke (PIF) emandako laguntza ekonomikoari esker.

given some time to understand it and, later, they have been asked to explain what they have seen. The speech of them was recorded through audio recordings and have constituted the corpus to be analysed. Among the results, there are significant differences in the types of syllables and the beginning of them depending on the generation and place of birth. In the case of the length of the prosodic groups, the speed of articulation and the duration of the syllables show significant differences according to gender.

Keywords: variation, sociolinguistics, prosody.

1. Sarrera

Gauza jakina da, edo izan beharko litzateke, pertsonak migratzean euren kulturak, hizkuntzak eta hizkuntza aldakiak ere migratzen direna. Behin helmugara helduta, kulturak eta hizkuntzak, bertakoekin nahastean, aldatu eta garatu egiten dira.

Hizkuntza guztiak denboran zehar aldatzen diren arren, ez dira berdin aldatzen jatorrizko ingurunean edo joaneko lekuan. Hori ezaguna da euskararen euskalkien bilakaerari dagokionez, izan ere, leku bateko gurasoek seme-alabei euren jatorrizko herriko euskalkia transmititu arren, euskalki hori ez da berdin garatzen atzean utzitako herrian edo Bilbo bezalako hiri handi batean.

Egun, Euskal Herria kolore eta soinu anitzez janzten ari denean, gizarte askotarikoa eta integratua egin gura denean, jatorrizko kultura eta hizkuntza horien guztien lekua eta funtzioa birpentsatzeko tenorea heldu dela uste dugu. Euskararen beraren funtzioa zein den ez da beti argi azaltzen, euskara bera gizarte koloretan horren koherentziarako tresna izan beharko litzateke, hori praktikan nola gauzatu ez dakigula aitortuta.

Migrazioak ez dira gauza berria gurean; etorkin berriak ikustean, agian gogotara ekarri beharko genuke, hainbeste urte ez direla, gutako batzuk eta gure aurreko asko joakinak izan direla. Batzuk itzuli ziren, baina beste asko bertan geratu ziren; horrexegatik dago hainbeste Euskal Etxe munduan barreiatuta. Ezin ahantzizkoak dira oraintsuago, erbesteratze ekonomikoak eraginda, kanpora joan diren gazte horiek guztiak. Denbora luze ez dela, asko dira Espainiatik gure artera etorri direnak bizimodu hobearen bila. Askok euren herrietara itzultzea dute amets, hori guztiz zilegi eta normala da, baina beste asko hemen geratu dira eta gure artekoak dira.

Testuinguru honetan aurkezten dugu gure lantxo hau, beronen helburua ez da handia; urrats txikiekin hasten da bidea urratzen. Hala bada, gure helburua da aztertzea familia bereko hiru belaunalditan zenbat aldatu diren ahozko testuen ezaugarri batzuk.

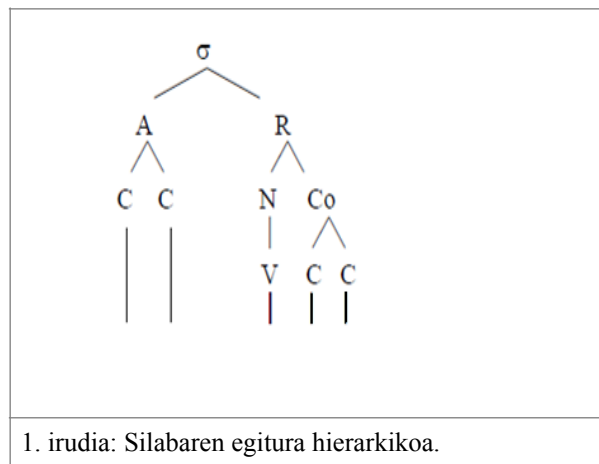
Lan honetan silaben erabilerari lotuko gatzazkio; silaba motak, talde prosodikoen luzera, silaben iraupena eta silaben iraupenaren eta hizketa abiadaren arteko erlazioak aztertuko ditugu. Gure helburua betetzeko familia bereko 12 informatzaile aukeratu ditugu (ikusi xehekiago 2. atalean), batzuk Extremaduran jaio ziren eta nagusitan emigratu zuten; beste batzuk han jaio eta umetan hona etorri ziren eta, azkenik, beste zenbait bertan sortu dira.

Gorago esaneko moduan, aukeratu ditugun ezaugarrietan alderik dagoen ikusi nahi da hiru aldagai soziolinguistikoren arabera: adina, generoa eta jaioterria.

Silaben gainean egin diren azterketetan (Moreno Cabrera, 1994) adierazi den bezala, eurotan osagai batzuk bereizi dira eta nolabaiteko egitura hierarkikoa proposatu da. Silabak oinarrizko hiru osagai dauzka: gunea, eta marjinetako hasiera eta koda. Hiru hauetatik euskaraz gunea baino ez da ezinbestekoa, beste biak aukeran daude.

Guneak segmenturik ozenena biltzen du, ozentasuna segmentuen propietate abstraktutzat hartzen da silabaren barruan gertatzen diren simetriak eta asimetriak biltzeko; kontzeptu honen arabera ozentasun hierarkia bat (1) ezarri da (Cutillas, 2003; Zec, 2007).

Gaztelaniaz gunean beti egoten da bokal bat gutxienez (V) eta alboetan, berriz, kontsonanteak (C). C eta V konbinatuaz, silaben oinarrizko tipologia bat egin da: CVC, CV, VC eta V. Silaben azterketetan aurkitu diren asimetriak hobeto harrapatzeko, gunea eta koda biltzen dituen tarteko maila bat proposatu da, hots, errima. Silaben egitura hierarkikoa 1. irudian erakusten dena da.



Talde prosodikoen luzera aztertzeko talde prosodikoa zer den argitu behar da. Esan dezakegu talde prosodikoa dela eten biren artean dagoen hizketa zatia; etenak isilunedunak (%) edo isilune bakoak (\$) izan daitezke (Gaminde, 2010). Talde prosodikoa ondoko era honetara irudikatuko litzateke:

$$\% \text{ TP } \% (\$)$$

Alta, badakigu txertaketak (tx) ere gertatzen ahal direla talde prosodikoaren aurrean zein ostean; beraz, definizioa egokia izanda ere, zehaztu beharko genuke ondoko era honetara:

$$\% (\text{tx}) \$ \text{ TP } \$ (\text{tx}) \%$$

Bestalde, luzera eta iraupena bereiziko ditugu. Luzera unitatetzat silaba hartuko dugu eta, hortaz, talde prosodikoen luzera biltzen dituzten silaba kopuruen arabera emango dugu. Iraupena, berriz, talde prosodikoen denborari dagokio eta milisegundotan (ms.) emango da.

Atal honetan testuetan informatzaileek erabiltzen duten hizketa abiadaz jardungo dugu. Hizketa abiadatzat jotzen dugu tempoa, hau da, hiztunak daukan abiadura testua egitean. Gil-ek (2007) abiaduraz berba egiterakoan honako definizio hau ematen du:

“Velocidad de elocución: Número de elementos fónicos (sonidos y pausas) que se pronuncian en una unidad de tiempo determinada” (548. or.).

Abiadura adierazten da denbora unitateko (segundoko gure kasuan) ekoizten diren hitz, silaba edo soinu kopuruaren bidez. Sarritan minutuko erabiltzen diren berba kopuruaz berba egiten da eta horren arabera arin edo astiro egiten dela esaten izan da. Hitza oso kontzeptu lausoa denez gero, hemen silabak erabiltzea erabaki dugu. Segundoko egiten diren silaben abiadura adierazteko aurreko lan batzuetan erabili dugun irizpidea hartu dugu (Gaminde et. al., 2012; 2017).

Abiadura ez da beti bat izaten eta alda daiteke faktore askok eraginda, besteak beste, hiztunaren ezaugarri pertsonalak eta testua gauzatzen den egoeraren baldintzak, Llisterri-k (2013) adierazten duen moduan:

“La velocidad de elocución puede ser variable en un mismo locutor, ya que depende, entre otros factores, de la relevancia informativa de los elementos que configuran el discurso. La velocidad de elocución también puede reflejar estados emotivos del hablante.”

Horrela, hizkera abiada geldoa mezuari eman nahi zaion garrantziarekin edo tonu arranditsuarekin lotzen da; haatik, hizkera bizia emozioekin, jarrerekin edo urduri egotearekin lotzen da. Tempoaren oso ezaugarri garrantzitsua izan daiteke mezuaren esangura eta hiztunaren intentzio komunikatiboa aztertzeke.

Arrunta izaten da, bestalde, hizketa abiada orokorra eta artikulazio-abiada bereiztea (Gaminde, 2010; Gaminde et. al., 2012; Llisterri, 2013). Hizketa abiada orokorra aztertzean testu osoaren denbora kontuan hartzen da, hau da, berbaldia eta etenaldiak. Artikulazio-abiada aztertzean etenaldien denbora ez da kontuan hartzen. Hemen artikulazio-abiada aztertuko dugu.

Aurreko lan batean erakutsi genuen bezala (Gaminde, 2016), silabaren iraupena ez da beti bat eta bera, informatzaile berak iraupen desberdina egin dezake hizkuntzaren arabera, testu motaren arabera edo kanpoko eta barneko beste eragile batzuen arabera. Hori dela eta, interesgarri iritzi diogu hau aztertzeari.

Lana lau ataletan banatuta aurkezten dugu: sarrera honen osteko bigarren ataletan gure azterketa egin ahal izateko erabili dugun corpusa deskribatuko da eta baita bera jasotzeko, prozesatzeko eta aztertzeke erabili den metodologia ere. Hirugarren ataletan datuen azterketa burutuko da. Azkenik, laugarren ataletan, ondorio nagusiak laburbilduko dira.

2. Corpusa eta Metodologia

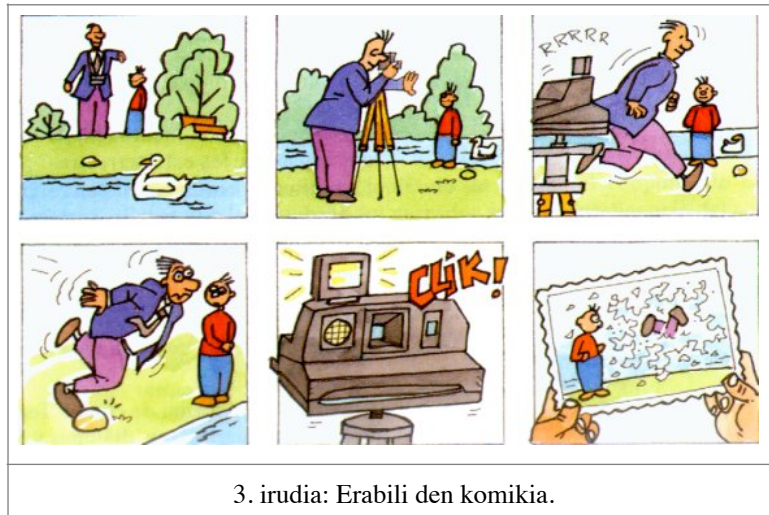
Ikerketa honetan familia bereko 10 informatzailek hartu dute parte. Informatzaile horiek sailkatu dira hiru parametroren arabera (ikusi 2. irudiko eskema):

- Belaunaldiaren arabera: 1950. urtean edo lehenago jaiotakoak (1. belaunaldia), 1951 eta 2000 urte bitartean jaiotakoak (2. belaunaldia), eta 2000. urtetik aurrera jaiotakoak (3. belaunaldia).
- Sexuka: lehen belaunaldiaren barruan 3 andrazko eta gizonezko bat daude, bigarren aldia 3 andrazko eta 3 gizonezko eta, hirugarren belaunaldian andrazko bat eta gizonezko bat.
- Jaioterriaren arabera: lehen belaunaldiko parte-hartzaile guztiak Extremaduran jaiotakoak dira eta helduaroan Bizkaira etorri dira bizitzen, bigarren belaunaldiko andrazko bat extremaduran jaio da eta gaztetan etorri da Bizkaira bizitzen eta beste bi Bizkaian jaiotakoak dira; bigarren belaunaldiko gizonezkoek dagokienez, hirurak Bizkaian jaioak dira.

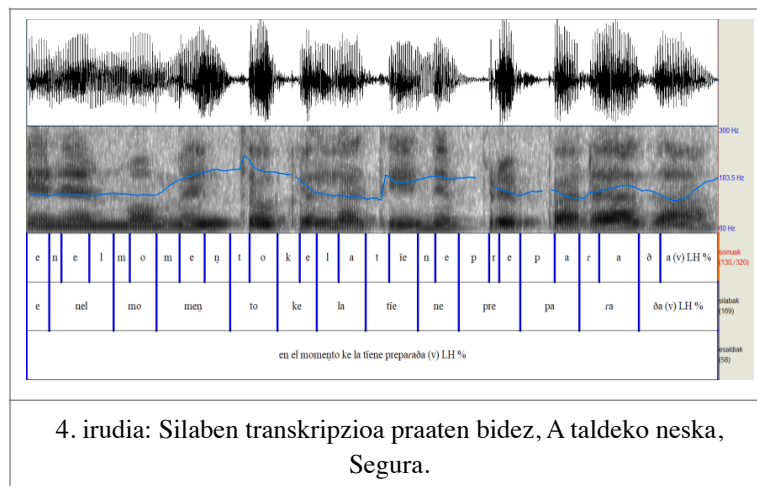
Hirugarren belaunaldiko parte-hartzaileak Bizkaian jaio dira. Eta guztien ama hizkuntza gaztelania da.

12 informatzaile	1. belaunaldia (4)	3 andrazko	(3) Extremadura
		1 gizonezko	(1) Extremadura
	2. belaunaldia (6)	3 andrazko	(1) Extremadura (2) Bizkaia
		3 gizonezko	(3) Bizkaia
	3. belaunaldia (2)	1 andrazko	(1) Bizkaia
		1 gizonezko	(1) Bizkaia
2. irudia: Informatzaileen inguruko informazioa.			

Erabilitako metodologiari dagokionez, informatzaileei 3. irudiko komikia erakutsi zaie eta minutu batzuen buruan komikian bertan gertatzen den istorioa azaltzeko eskatu zaie euren ama hizkuntzan. Ahozko emisioak grabatu dira, ondoren grabazioak transkribatu eta etiketatzeko.



Silaba guztiak transkribatu eta etiketatu egin dira “Praat” programaren bidez (Boersma & Weenink, 2018), 4. irudian erakusten den bezala.



Silaben iraupenaren azterketa gehien erabiltzen den CV silaba motarekin burutu dugu. Gure azterketa burutzeko ez ditugu denak neurtu; esaldiaren amaieran agertzen direnak baztertu egin ditugu; izan ere, hauek luzeagoak izan daitezke testuaren egokitzapen pragmatikorako luzatu egiten direlako. Era berean, talde prosodikoaren hasierako gunean agertzen direnak baztertu egin ditugu, herskariaren eta afrikatuen kasuan ezin neur daitezkeelako osorik.

Hizketa abiadaren aldakortasuna erkatu ahal izateko testuen talde prosodikoaren batezbestekoak eta desbideratzeak erabilia aldakuntza koefizienteak kalkulatu ditugu, hau da:

$$Ak = \frac{sd}{x^-} * 100$$

Aldakuntza koefizientearen emaitzak dira gero erabili ditugunak, informatzaileen emaitzak aldagaien arabera alderatzeko.

3. Datuen azterketa

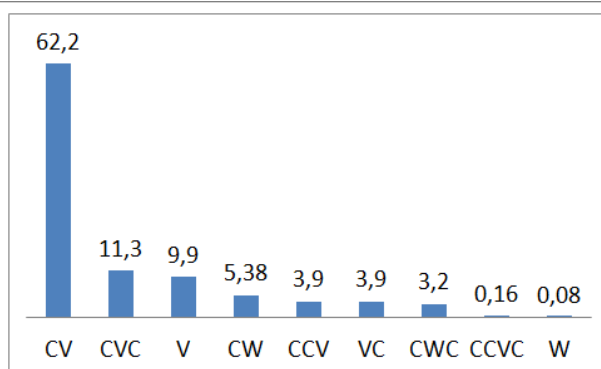
Gorago esaneko moduan, lan honetarako jaso diren testuen prosodiaren azterketaren arlo batzuk aukeratu dira: silaba motak, talde prosodikoen luzera, hizketa abiada eta silaben iraupena, hain zuzen ere. Atal honetan datuen azterketa aurkeztuko dugu segida horretan.

Testuetan jaso ditugun silabak 1.283 izan dira denetara. 1. taulan silaba moten kopuruak eta ehunekoak ematen ditugu euron osagaien arabera sailkatuta¹³ (5. irudiko grafikoa).

¹³ C kontsonantea, V bokala eta W diptongoa adierazteko erabiliko dira.

Mota	Kopurua	%
CV	798	62,20
CVC	145	11,30
V	127	9,90
CW	69	5,38
CCV	50	3,90
VC	50	3,90
CWC	41	3,20
CCVC	2	0,16
W	1	0,08
Guztira	1283	100

1. taula: Silaba moten kopuruak eta ehunekoak.



5. irudia: Silaba motak kopuruen arabera sailkatuta.

2. taulan mota nagusien kopuruak eta ehunekoak ematen ditugu informatzaileen belaunaldien arabera sailkatuta. Euren arteko aldea estatistikoki esanguratsua da ($\chi^2_{(a.m.: 6)} = 23,765$; $p = 0,001$).

Belaunak	CV		CVC		V		W	
	Kopurua	%	Kopurua	%	Kopurua	%	Kopurua	%
1	182	64,08	28	9,86	50	17,61	24	8,45
2	501	72,29	94	13,56	59	8,51	39	5,63
3	115	70,99	23	14,20	18	11,11	6	3,70

2. taula: Silaba mota nagusien kopuruak eta ehunekoak.

3. taulan datuak ematen dira informatzaileen generoaren arabera sailkatuta. Bertan agertzen den aldea ez da estatistikoki eragingarria.

Belaunal	CV		CVC		V		W	
	Kopurua	%	Kopurua	%	Kopurua	%	Kopurua	%
1	479	68,33	94	13,41	87	12,41	41	5,85
2	319	72,83	51	11,64	40	9,13	28	6,39

3. taula: Silaba motak informatzaileen generoaren arabera.

Azkenik, 4. taulan datuak ematen dira informatzaileen jaioterriaren arabera sailkatuta. Bertan agertzen den aldea estatistikoki esanguratsua da ($\chi^2_{(a.m.: 3)} = 18,957$; $p = 0,000$).

Belaunal	CV		CVC		V		W	
	Kopurua	%	Kopurua	%	Kopurua	%	Kopurua	%
1	269	66,09	45	11,06	67	16,46	26	6,39
2	529	72,27	100	13,66	60	8,20	43	5,87

4. taula: Silaba motak informatzaileen jaioterriaren arabera.

Silaba hasiera hutsa izan daiteke (178, % 13,87) edo kontsonante bat (1053, % 82,07) edo bi (178, % 4,05) egon daitezke. Hasierarik aruntena kontsonante batekoa izatea da. 5. taulan hasiera bakoitzeko kopuruak eta ehunekoak ematen ditugu belaunaldien arabera. Aldea estatistikoki esanguratsua da ($\chi^2_{(a.m.: 4)} = 10,713$; $p = 0,03$).

Belaunaldia	C		CC		Ø	
	Kopurua	%	Kopurua	%	Kopurua	%
1	242	77,07	12	3,82	60	19,11
2	658	84,25	30	3,84	93	11,91
3	153	81,38	10	5,32	25	13,30

5. taula: Silaben hasiera informatzaileen belaunaldiaren arabera.

6. taulan datuak ematen dira informatzaileen generoaren arabera sailkatuta, bertan agertzen diren aldeak ez dira estatistikoki esanguratsuak.

	C		CC		Ø	
Generoa	Kopurua	%	Kopurua	%	Kopurua	%
Andrazkoak	644	81,11	29	3,65	121	15,24
Gizonezkoak	409	83,64	23	4,70	57	11,66

6. taula: Silaben hasiera informatzaileen generoaren arabera.

Azkenik, 7. taulan datuak aurkezten dira informatzaileen jaioterrien arabera sailkatuta. Aldea estatistikoki esanguratsua da ($\chi^2_{(a.m.: 2)} = 11,453$; $p = 0,003$).

	C		CC		Ø	
Jaioterria	Kopurua	%	Kopurua	%	Kopurua	%
Oliva	353	77,75	18	3,96	83	18,28
Bizkaia	700	84,44	34	4,10	95	11,46

7. taula: Silaben hasiera informatzaileen jaioterriaren arabera.

Silaba guneetan bokal bakar bat egotea arruntena da (1172, % 91,35), diptongoak askozaz gutxiagotan erabiltzen dira (111, % 8,65). Diptongo hauetako 31 (% 27,93) beheranzkoak dira eta 80 (% 72,07) goranzkoak. 8. taulan datuak ematen ditugu informatzaileen belaunaldien arabera sailkatuta. Bertan agertzen diren aldeak ez dira estatistikoki esanguratsuak.

	V		W	
Belaunaldia	Kopurua	%	Kopurua	%
1	281	89,49	33	10,51
2	718	91,93	63	8,07
3	173	92,02	15	7,98

8. taula: Silaba guneak informatzaileen belaunaldiaren arabera.

9. taulan datuak erakusten dira informatzaileen generoaren arabera sailkatuta. Aldea ez da estatistikoki esanguratsua.

	V		W	
Generoa	Kopurua	%	Kopurua	%
Andrazkoak	722	90,93	72	9,07
Gizonezkoak	450	92,02	39	7,98

9. taula: Silaba guneak informatzaileen generoaren arabera.

10. taulan datuak informatzaileen jaioterrien arabera sailkatuta aurkezten dira. Kasu honetan ere aldea ez da estatistikoki esanguratsua.

	V		W	
Jaioterria	Kopurua	%	Kopurua	%
Oliva	414	91,19	40	8,81
Bizkaia	758	91,44	71	8,56

10. taula: Silaba guneak informatzaileen jaioterriaren arabera.

Azkenik, kodak hutsak izan daitezke (1045, % 81,45) edo, gure corpusean, kontsonante bakarrekoak (238, % 18,55). 11. taulan datuak ematen ditugu informatzaileen belaunaldien arabera sailkatuta. Bertan agertzen diren aldeak ez dira estatistikoki esanguratsuak.

	C		Ø	
Belaunaldia	Kopurua	%	Kopurua	%
1	46	14,65	268	85,35
2	153	19,59	628	80,41
3	39	20,74	149	79,26

11. taula: Silaba kodak informatzaileen belaunaldiaren arabera.

12. taulan datuak erakusten dira informatzaileen generoaren arabera sailkatuta. Aldea ez da estatistikoki esanguratsua.

	C		Ø	
Generoa	Kopurua	%	Kopurua	%
Andrazkoak	158	19,90	636	80,10
Gizonezkoak	80	16,36	409	83,64

12. taula: Silaba kodak informatzaileen generoaren arabera.

13. taulan datuak informatzaileen jaioterrien arabera sailkatuta aurkezten dira. Kasu honetan ere aldea ez da estatistikoki esanguratsua.

	C	Ø
--	---	---

Jaioterria	Kopurua	%	Kopurua	%
Oliva	74	16,30	380	83,70
Bizkaia	164	19,78	665	80,22

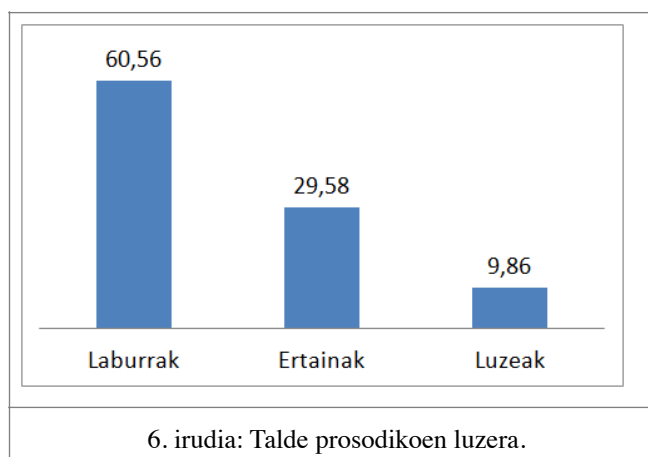
13. taula: Silaba kodak informatzaileen jaioterriaren arabera.

Denetara 213 talde prosodiko jaso dugu, euron luzerak biltzen dituzten silaba kopuruen arabera kalkulatuta. 14. taulan ematen ditugu talde prosodiko mota bakoitzeko kopuruak eta ehunekoak adierazita.

Silabak	Kopurua	%	Silabak	Kopurua	%
1	24	11,27	11	4	1,88
2	22	10,33	12	4	1,88
3	27	12,68	13	8	3,76
4	23	10,80	14	2	0,94
5	20	9,39	15	1	0,47
6	13	6,10	16	3	1,41
7	24	11,27	17	3	1,41
8	13	6,10	18	2	0,94
9	7	3,29	19	2	0,94
10	11	5,16			

14. taula: Talde prosodiko mota bakoitzeko kopuruak eta ehunekoak.

Talde prosodikoen luzeraren arabera hiru multzo egin ditzakegu: laburrak, 1 eta 6 silaba bitartekoak (129); ertainak, 7 eta 12 silaba bitartekoak (63); eta luzeak, 13 silabatik gorakoak (21). 6. irudiko grafikoan bakoitzaren ehunekoak ikus daitezke.



Banaketa honen arabera sailkatzen baditugu datuak informatzaileen belaunaldiak kontuan hartuta 15. taulako emaitzak lortzen dira. Aldeak ez dira estatistikoki esanguratsuak.

Belaunaldia	Laburra		Ertaina		Luzea	
	Kopurua	%	Kopurua	%	Kopurua	%
1	35	62,50	20	35,71	1	1,79
2	80	62,02	35	27,13	14	10,85
3	14	50,00	8	28,57	6	21,43

15. taula: Talde prosodikoen luzera informatzaileen belaunaldiaren arabera.

16. taulan datuak ematen dira informatzaileen generoaren arabera sailkatuta. Bertan agertzen den aldea estatistikoki esanguratsua da ($\chi^2_{(a.m.: 2)} = 9,549$; $p = 0,008$).

Generoa	Laburra		Ertaina		Luzea	
	Kopurua	%	Kopurua	%	Kopurua	%
Andrazkoak	93	64,58	43	29,86	8	5,56
Gizonezkoak	36	52,17	20	28,99	13	18,84

16. taula: Talde prosodikoen luzera informatzaileen generoaren arabera.

17. taulan datuak aurkezten dira informatzaileen jaioterriaren arabera sailkatuta. Bertan agertzen den aldea estatistikoki esanguratsua da ($\chi^2_{(a.m.: 2)} = 9,025$; $p = 0,011$).

Jaioterria	Laburra		Ertaina		Luzea	
	Kopurua	%	Kopurua	%	Kopurua	%
Oliva	55	64,71	28	32,94	2	2,35
Bizkaia	74	57,81	35	27,34	19	14,84

17. taula: Talde prosodikoen luzera informatzaileen jaioterriaren arabera.

Artikulazio-abiadari dagokionez, orokorrean berorren batezbestekoa 5,41 silaba segundoko da; desbideratzea 1,61; minimoa 0,88; maximoa 9,09; eta heina 8,21. Datuak informatzaileen belaunaldien arabera aztertuta, 1. belaunaldikoen batezbestekoa 5,29 da (sd= 1,64), 2. belaunaldikoena 5,5 (sd= 1,62) eta 3. belaunaldikoena 5,25 (sd= 1,53), aldeak ez dira estatistikoki esanguratsuak. Datuak informatzaileen generoaren arabera aztertzen ditugunean, ikusten dugu andrazkoen batezbestekoa 5,17 (sd= 1,57) dela eta gizonezkoena 5,92 (sd= 1,59), alde hori bada estatistikoki esanguratsua ($T = (a.m.: 211) -3,235$; $p = 0,001$). Azkenik, datuak informatzaileen jaioterriaren arabera aztertuta, Olivan jaiokoen batezbestekoa 5,3 (sd= 1,42) dela ikusten dugu eta Bizkaian jaiokoena 5,49 (sd= 1,73), alde hori ez da estatistikoki esanguratsua.

18. taulan informatzaile guztien artikulazio-abiadaren batezbestekoa, desbideratzea eta aldakuntza koefizienteak ematen dira; 7. irudiko grafikoan batezbestekoen eta aldakuntza koefizienteen bidez egineko dispersio grafikoa erakusten da. Bertan agertzen den bezala, gizonezkoak gaineko aldean kokatzen dira eta andrazkoak berriz beheko aldean.

Informatzaileak	X	Sd	AK
A1937	5,05	1,86	36,87
A1940	5,70	1,21	21,26
A1947	4,17	1,38	33,00
A1962	5,32	0,90	16,99
A1970	5,38	2,02	37,61
A1973	5,05	1,40	27,68
A2005	5,00	1,96	39,14
G1935	6,04	1,54	25,55
G1965	6,32	1,28	20,24
G1973	6,23	1,86	29,91
G1977	5,47	1,96	35,89
G2006	5,55	0,80	14,49
18. taula: Artikulazio-abiada.			

7. irudia: Informatzaileen sailkapena artikulazio-abiadaren arabera.

Denetara 601 silabaren iraupena neurtu dugu, euren batezbestekoa 140,14 ms.koa da, desbideratzea 43,30, minimoa 40, maximoa 320 eta heina 280. Datuak informatzaileen belaunaldien arabera aztertzean, 1. belaunaldikoen batezbestekoa 146,23 da (sd=41,95), 2. belaunaldikoena 136,29 (sd=40,7) eta 3. belaunaldikoena 147,34 (sd=52,88); 1. eta 3. belaunaldikoen artean ez dago alderik eta 2. belaunaldikoek silabak laburrago egiten dituzten arren, aldea ez da estatistikoki esanguratsua. Datuak informatzaileen generoaren arabera aztertzen ditugunean, ikusten dugu andrazkoen batezbestekoa 147,52 (sd=43,12) dela eta gizonezkoena 130,12 (sd=41,57), alde hau bada estatistikoki esanguratsua ($T = (a.m.: 599) 4,962; p = 0,000$). Azkenik, datuak informatzaileen jaioterriaren arabera aztertuta, Olivan jaiokoen batezbestekoa 152,07 (sd=42,8) dela ikusten dugu eta Bizkaian jaiokoena 134,75 (sd=42,45), alde hau estatistikoki esanguratsua da ($T = (a.m.: 599) 4,625; p = 0,000$).

Amaitzeko interesgarri iritzi diogu silaben iraupenaren eta hizketa abiadaren artean korrelaziorik dagoen aztertzeari; 19. taulan informatzaile bakoitzaren CV silaben iraupenaren batezbestekoak eta hizketa abiada ematen dira batera. Pearsonen korrelazioa kalkulatu

berori -0,765ekoa dela azaltzen da, korrelazioa nahikoa altua da eta esan dezakegu zenbat eta abiada handiagoa eduki orduan silabak laburragoak izan daitezkeela, zein bestalde logikoa baita.

Informatzailea	Iraupena	Abiada
A1937	144,71	5,05
A1940	147,07	5,70
A1947	190	4,17
A1962	165,17	5,32
A1970	124,25	5,38
A1973	160	5,05
A2005	142,41	5,00
G1935	128,72	6,04
G1965	133,66	6,32
G1973	119,38	6,23
G1977	129,74	5,47
G2006	154	5,55
19. taula: Silaben iraupena eta hizketa abiada.		

4. Ondorioak

Atal honetan ondorioak eta hurrengo lanak aurkeztuko dira.

Silaben frekuentziak aztertuta, ondorioztatu dugu gehien erabiltzen den silaba mota CV dela. Onartzen diren silaba motak, euron erabileren ehunekoen bidez sailkatu ditugu eta gehien erabiltzen diren bost motekin ondoko rankinga osatu dugu:

$$CV > CVC > V > VC > CW$$

Bestetik, azertu diren tasun eta aldagaietatik, alde estatistikoki esanguratsuak ikusi dira hurrengoetan:

- (1) Silaba motetan, belaunaldiaren eta jaioterriaren arabera.
- (2) Silaben hasieretan, belaunaldiaren eta jaioterriaren arabera.
- (3) Talde prosodikoen luzeretan, genero eta jaioterriaren arabera.
- (4) Artikulazio-abiadan, generoaren arabera.

(5) Silaben iraupenetan, genero eta jaioterriaren arabera. Silaben gunetan eta kodetan ez da alde esanguratsurik ikusi.

Hurrengo lanei dagokienez, interesgarria litzateke hemen aztertu diren tasunak aztertzen jarraitzea eta Extremaduran bizi diren gainerako familiakideekin eta lortutako emaitzak ikerketa honetan egindakoekin konparatzea.

5. Erreferentziak

- Boersma, P. & Weenink, D. (2018). *Praat: doing phonetics by computer, Version 5.1*. [Computer program]. Hemen berreskuratuta: <http://www.praat.org>.
- Cutillas, J.A. (2003). *Teoría Lingüística De La Optimidad. Fonología, Morfología y Aprendizaje*. Murtzia: Murtziako unibertsitatea.
- Gaminde, I. (2010). *Bizkaiko gazteen prosodiaz: euskaraz eta gaztelaniaz*. Bilbo: Mendebalde Kultura Alkartea eta BFA.
- Gaminde, I. (2016). Hizkuntzaren eragina prosodiaren tasun batzuetan. *Euskalingua*, 29, 6-14.
- Gaminde, I.; Romero, A. & Legarra, H. (2012). *Gramatika eta Hizkuntz Bariazioa Bermeon*. Bermeo: Bermeoko Udala eta Campos Hegaluzea.
- Gaminde, I.; Olalde, A.; Etxebarria, A.; Eguskiza, N. & Gaminde, U. (2017). *Hizkuntza Aldakortasuna Larrabetzun*. Bilbo: Larrabetzuko Udala.
- Gil, J. (2007). *Fonética para profesores de español: De la teoría a la práctica*. Madrid: Arco/ Libros.
- Llisterri, J. (2013). *Experimental phonetics*. Eskuragarri: <http://liceu.uab.es/~joaquim/>.
- Moreno Cabrera, J.C. (1994). *Curso Universitario de lingüística general. Tomo II: Semántica, pragmática, morfología y fonología*. Madril: Síntesis
- Zec, D. (2007). The Syllable. In P. de Lacy (arg.), *The Cambridge Handbook of Phonology* (161-194). Cambridge: Cambridge University Press.

Juan Abasolo, Aitor Iglesias, Asier Romero eta Lorea Unamuno
UPV/EHU

Juan.abasolo@ehu.eus, aitor.iglesias@ehu.eus, a.romero@ehu.eus, lorea.unamuno@ehu.eus.

Laburpena

Lan honen helburua baliatutako transkripzio motaren arabera, distantzia linguistikoak kalkulatzeko zein neurri erabili behar den eztabaidatzea da. Era honetako corpusekin lan egitean dauden bost une kritikoak identifikatu eta aztertu dira. Emaitzetan distantzia ezberdinen erabilpenaren eragina zein izan daitekeen ikusteko lau distantzia, RIV, Levenshtein, Q-gram eta ALINE, aplikatu dira Arratiako aditza jasotzen duen corpusean.

Hitz gakoak: dialektometria, distantzia linguistikoak, ALINE.

Abstract

The aim of this paper is to discuss which linguistic distances should be used in an oral corpus according to the type of transcription used. The five critical aspects that exit when working with this type of corpus are identified and examined. Four linguistic distances, RIV, Levenshtein, Q-gram and ALINE, have been used to analyse the effect on the result of each of them in the verbal corpus of the Arratia valley.

Keywords: dialectometry, linguistic distance, ALINE.

1. Sarrera

Lan honen bidez dialektometriari distantziak neurtzeko erabiltzen diren metodoez jardun eta eztabaidatu nahi dugu. Artikulu honen helburua da ondo azaltzea zein metodo aukeratu behar den corpusaren transkripzioaren arabera. Gaminderen (1984) corpusaren zati batez baliatuz, distantziak neurtzeko ohikoak diren *RIV* eta *Levenshtein* neurriak zelan aplikatzen diren ikusiko dugu eta, horietaz gain, orain arte, behintzat, Euskal Herrian egin diren lanetan erabili ez diren beste neurri bi aurkeztuko ditugu, *q-gram* (van der Loo, 2014) eta *ALINE* (Downey et. al, 2017).

Lan honek, sarreraz gain, sei atal izango ditu. Bigarren atalean erabilitako corpusaren ezaugarrien berri emango da; ostean, marko teorikoaz jardungo du, bertan zein transkripzio mota egin daitekeen aztertuko da eta baita dialektometriari egin diren ekarpenik

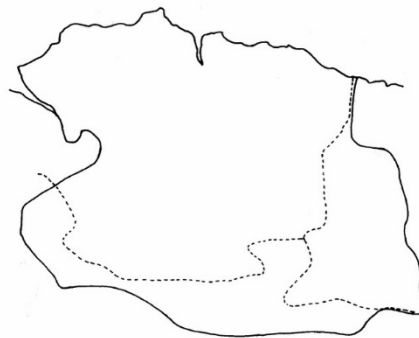
¹⁴ Ikerketa hau burutu da IT1028/19 (Eusko Jaurlaritza) eta GIU 16/22 (Euskal Herriko Unibertsitatea) proiektuen barnean eta UPV/EHUk doktoretza aurreko formakuntzan dagoen pertsonala kontratatzeke (PIF) emandako laguntza ekonomikoari esker.

esanguratsuenetan erabili izan ohi diren neurriak distantzia linguistikoak neurtzeko. Marko teorikoaren ondoren, lan honetan aipatutako corpusean distantzia linguistikoak neurtzeko erabili den metodologia azalduko da. Bosgarren puntuan lortutako emaitzak aurkeztuko dira eta, azkenik, distantziak kalkulatzeko modu ezberdinek dituzten ondorioak zeintzuk diren azalduko dira. Amaitzeko, erabilitako bibliografiaren berri dugu.

2. Corpora

Lanaren alderdi teoriko eta metodologikoarekin hasi baino lehen, erabili dugun corpusaren ezaugarriak aipatuko ditugu. Gamindek (1984) idatzitako *Aditza bizkaieraz* liburutik corpus txiki bat atera dugu. Liburu honetan Gamindek 117 herritako aditza batu zuen, herri bakoitzean 217 adizki jasoz. Guk 117 herrietatik Arratiako herrietan batutako adizkiekin corpus txiki bat lortu dugu: 8 herri (Bedia, Lemoa, Igorre, Dima, Zeanuri, Artea, Areatza eta Zeberio) eta esandako moduan, 217 adizki herri bakoitzeko; beraz, guztira 1736 adizki izan ditugu aztergai. Adizki hauen artean Nor, Nor-Nori, Nor-Nork eta Nor-Nori-Nork adizki laguntzaile gehienak batu ziren eta baita *etorri*, *eduki*, *jakin*, *ekarri* eta *eroan* aditzen hainbat adizki trinko ere. Liburuan adizki hauek transkripzio fonologiko batez izan dira transkribatuak.

ADITZA BIZKAIERAZ I. TOMOA



IEI udako
euska
unibertsitatea

IRUÑEA 1984

1. irudia: Gaminde (1984:azala)

Corpus hau on-line dago eskuragarri eta esan behar da oso homogeneoa dela, dialektologo berak batu duelako metodologia berari jarraituz.

3. Marko teorikoa

Lehen aipatu dugun legez, lan honekin hausnartu eta eztabaidatu nahi dugu jakiteko zein distantzia erabili behar dugun daukagun transkripzio motaren arabera. Horretarako, lehenik eta behin azaldu behar da zein transkripzio mota dauden. Orokorrean transkripzioak bi eratara egin daitezke: fonetikoki eta fonologikoki; eta transkripzio fonetikoak, zabalak edo estuak izan daitezke, transkribatzaileak jartzen du muga (ikus. 1. taula). Transkripzio fonetiko estua ahalik eta informazio gehien jasotzen ahalegintzen da; zabalak, ordea, ezaugarri esanguratsuenak baino ez ditu ematen; eta, beraz, fonologikora hurbiltzen da. Alfabeto fonetiko batzuk dauden arren, euskaraz gehienbat IPA erabili izan da. Euskarazko lanetan batez ere transkripzio fonologikoak erabili dira eta, hortaz, corpus linguistiko ugari dugu transkripzio mota honekin.

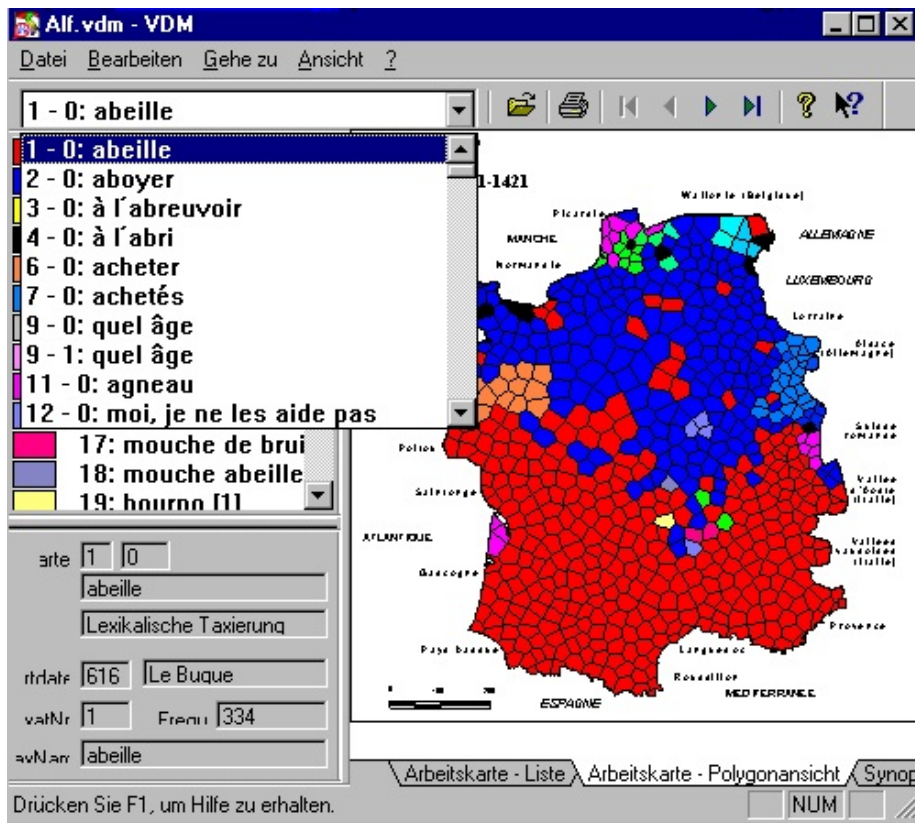
Transkripzio motak		
Fonetikoak		Fonologikoak
Zabalak	Estuak	
oro	oro	oro
	oro	
	oro	
	...	

1. taula: transkripzio motak

1. taulan *oro* (todo / tout) hitzaren hainbat transkripzio ikus dezakegu. Alde batetik, fonologikoa dugu, Euskal Herriko ohiko dialektologian autore gehienek erabili izan dutena (Arretxe, 1994; Carrera & García, 2005; Etxebarria, 2016; Gaminde, 1994; Zuazo, 2014 eta abar); eta, beste aldetik, fonetikoak dugu. Transkripzio mota hau Euskal Herri Hizkeren Atlasean topatu dezakegu edota Etchebest-en (2017) tesian; izan ere honen moduko lanetan zehazteko beharra dago, horrela, adbididez, Etchebestek bere lanean zortzi dardarkari mota ezberdin bereizten ditu zubererarako. Beraz, ikusi dugunez, hainbat transkripzio mota daude eta horien arabera hizkuntzalariak jakin behar du zein azterketa dialektometriko den egokiena.

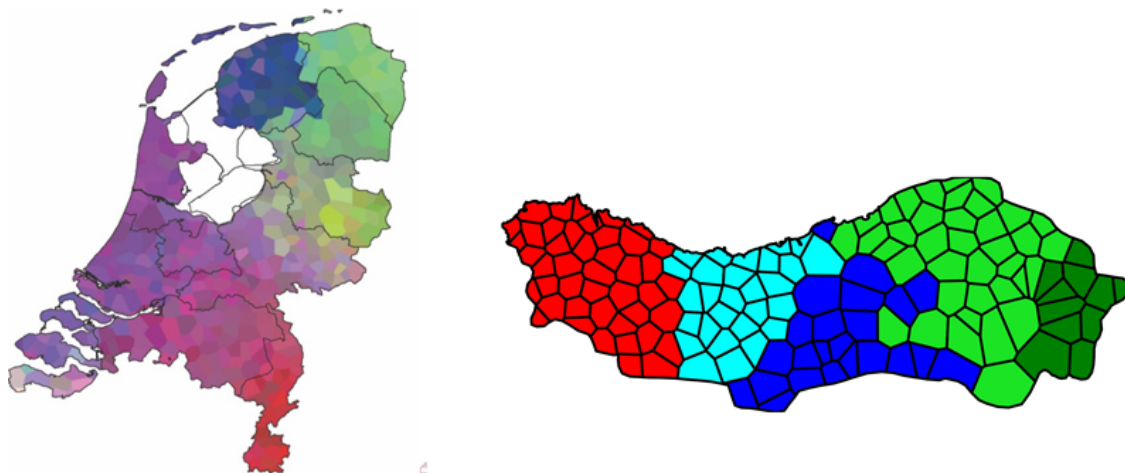
Dialektometria distantzia linguistikoak neurtzeaz arduratzen den diziplina da eta honi buruz historia apur bat egiten badugu, esan daiteke Hans Goebel dela dialektometriaren aita. Berak Salzburgoko eskola sortu zuen eta bertan VDM (*Visual dialectometry*) programa garatu zuten, lehenengoa arlo honetan. Euskal Herrian ezaguna da batez ere Bourciezen corpusarekin egindako lanengatik (Aurrekoetxea et al., 2004 eta 2005). VDM programak datu-matrizeekin hainbat ebaluazio-algoritmo eskaintzen ditu. RIW, GIW eta DEM antzekotasun neurriek herrien arteko antzekotasuna kalkulatzeko dute datu-matrizean. MINMWMAX, MWMAX eta MED segmentu-algoritmoekin ere lan egiteko aukera ematen du eta hainbat sinopsi-kalkulu (minimoa, maximoa, desbideratze estandarra, asimetria, eta abar) ere egin daitezke. Programa

honek aurkeztu zuten berrikuntzarik handiena izan zen datuen emaitzak islatzea Voronoi diagramekin mapa batean, hortik *visual* hitza.



2. irudia: VDM programako webgunetik

Gaur egun, Goeblen ostean, esan daiteke beste eskola bi daudela: Groningengoa, John Nerbonne buru duela; eta Euskal Herrikoa, Gotzon Aurrekoetxea buru. Nerbonnen taldeak *Gabmap* web-aplikazioa garatu du, eta Aurrekoetxearen EUDIA Ikerketa-taldeak (UPV/EHU) *Diatech* softwarea.



3. irudia: Gabmap eta Diatech programen euren webguneetatik hartuak

Gabmapek Levenshtein neurria erabiltzen du transkripzio fonetikoetan oinarritutako distantzia linguistikoak neurtzeko. Diatechek Levenshteinez gain, RIV (Relative Index Value) eta WIV (Weighted Identity Value) distantziak ere erabiltzeko aukera ematen du. Dialektologia kuantitatiboan RIV eta Levenshtein dira maizen erabiltzen diren neurriak. Labur-labur azalduz, RIV berdintasun neurria da, erabateko berdintasuna edo ezberdintasuna markatzen duena. Adibidez, hiru erantzun hauek baditugu (*abadea*, *abadia* eta *apaiza*), RIVek esango digu hiruren arteko distantzia bera dela, erabatekoa; izan ere, neurri honentzat hiru erantzunak ezberdinak dira. Levenshtein neurriak, ordea, item bateko kodeketatik bestera pasatzeko zenbat aldaketa dauden neurtzen du. Adibidez *abadia* itemetik *abadea* itemera pasatzeko aldaketa bakarra egin behar da, horregatik neurri hau kontuan hartzen badugu, item bi hauen arteko distantzia nahiko laburra da. Honegatik guztiagatik, hizkuntzalariak ondo aztertu behar du zer erantzun eta transkripzio mota duen edota behar den azterketa egiteko. Aurrekoetxeak (Aurrekoetxea et al. 2015) proposatzen duen moduan, RIV neurria egokia da estandarizatutako edo lematizatutako datuak erkatzeko; izan ere, similaritatea edo antzekotasun maila adierazten du, alderatzen dituen bi itemak berdinak edo desberdinak diren. Gure adibideari berriro helduta, RIV distantzia erabiltzeko modu honetan lematizatu beharko genuke :

Erantzunak	Lema
<i>apaiza</i>	apaiz
<i>abadia</i>	abade
<i>abadea</i>	abade

2. taula: lematizazioaren adibidea

Horrela, RIV neurria *abade* eta *apaiz* lemen artean baino ez litzateke aplikatuko eta, beraz, *abadia* eta *abadea* erantzunen artean ez litzateke distantziarik egongo, bai ordea, *apaiza*-rekin. Levenshtein distantzia, ordea, egokiagoa da itemak euren artean erkatzeko, lema bereko transkripzio fonetikoak alderatzeko, hain zuzen.

4. Metodologia

Gaminderen (1984) Arratiako corpusari RIV eta Levenshtein distantzietan gain, beste bi aplikatzea proposatuko dugu: *q-gram* (van der Loo, 2014) eta *ALINE* (Downey et. al, 2008). *q-gram* distantziak, Jaccarden teorian oinarritua, transkripzio batetik beste batera duten elementuen multzokatzeen antzekotasuna neurtzen du. *ALINE* distantziak, berriz, itemen arteko distantzia fonetikoak kalkulatu du, itemak IPAko transkripzio fonetikoan idatzita edota erabiltzaileak egoki deritzon kodeketaren arabera izan.

Distantzia motak aztertu baino lehen, era honetako azterketa egiten denean, bost une kritikoak zeintzuk diren aztertuko dugu:

- a. Datuak batzea
- b. Datuen kudeaketa

- c. Distantziak kalkulatzeko
- d. Multzokatze prozesua
- e. Taldekatzeko aukerak

Lehenengo une kritikoa datuak batzea da. Datu-bilketa era askotakoa izan daiteke: elkarriketa naturaletik jasotako datuak (Ormaetxea, 2005; Makazaga, 2011); idatzizko testuetatik batutakoa (Aurrekoetxea et al., 2004 eta 2005); eta, ohikoena, González Salgadok aipatzen duen moduan, galdetegia. Euskararen inguruan galdetegi asko egin izan dira. Lehenengoetakoa arlo honetan Azkueren *Erizkizundi Irukoitza* izan zen (Etxaide, 1984), 1924-25ean jasotakoa; orain arte egindako handiena *Euskal Herriko Hizkuntz Atlasa* (Aurrekoetxea & Videgain, 1993) da eta horren gainean EHUKo EUDIA ikerketa-taldeak moldatu zuen EAS (Euskararen atlas soziolinguistiko-dialektologikoa) galdetegia (Aurrekoetxea & Ormaetxea, 2006). Talde honek EDAK (Euskara Dialektalaren Ahozko Corpora) galdetegia ere prestatu zuen azenturako eta intonaziorako (Aurrekoetxea et al., 2009); eta, jakina, beste autore batzuek ere galdetegiak erabili dituzte bere lanak osatzeko. Dena den, galdetegiak datu batuketa errazten duen arren, ezin dira albo batera utzi dituen mugak eta eragozpenak (Dörnyei 2007). Adibidez, galdetegiaren sinpletasunak ez du ikerketan sakontzen uzten, lekukoak txarto interpreta dezake galdera (eta batzuetan ikertzaileak ere txarto galde dezake), ikertzaileak lekukoarengan duen eraginak ere emaitzak baldintza ditzake; izan ere, batzuetan, lekukoek erantzuten dute inkestagileak espero duena eta ez benetan pentsatzen dutena. Maiz, lekukoek orokortzeko joera ere badute, Laboven *behatzailearen paradoxa*. Egon daitezkeen arazo hauek guztiak kontuan hartuta, ikertzaileak galdetegia egin behar du azterketaren helburua kontuan hartuta eta datuak batzen hasi baino lehen, galdetegia probatu eta balidatu.

Bigarren une kritikoa datuen kudeaketa da. Lehenik eta behin corpora jaso behar da. Gaur egun hori egiteko modu bi daude: audio grabazioen bidez, edo bideo grabazioen bidez. Ahozko corpus bat jasotzen dugunean datuak kodifikatu egin behar dira, erabaki behar da zein sistema erabili datuak kodifikatzeko. Arestian aipatu dugunez, transkribatzeko mota batzuk ditugu, baina guztietan informazioa galtzen da (Linell, 2005) transkribatzailearen interesen arabera egiten delako; horregatik, oso garrantzitsua da argi izatea zein den azterketaren xedea, transkribatzeko sistema ikerketaren helburuaren menpe egon behar delako ezinbestean (Briz et al. 2002). Hidalgo eta Sanmartín (2005) arabera joera bi daude datuen kodeketarako: etiketan oinarritutako sistema bat, unibertsala izan beharko litzatekeena eta datuen konparazioak eta analisi estatistikoak egitea errazten duena; eta bestea, etiketa bako sistema bat irakurketa errazten duena. Hala eta guztiz ere, nahiz eta transkripziorik zehatzena egin eta nahi beste etiketa jarri, ez dugu inoiz izango audio artxibo batek duen informazio guztia.

Hirugarren une kritikoa distantziak kalkulatzeko da. Lan honetan batez ere puntu honi erreparatu diogu. Arestian aipatu dugun moduan, distantziak kalkulatzeko neurriak datuen transkripzioen arabera egin behar dira edo, bestela, datuek trataeraren bat beharko dute. Marko teorikoan aipatutako bi distantzietatik, RIV eta Levenshtein, lan honetan beste bi ere aztertuko ditugu: q-gram eta ALINE. Arestian aipatu dugun legez, RIV berdintasun neurria

da, erabateko berdintasuna neurtzen duena eta hau aplikatzeko komeni da datuak lematizatuta egotea. Levenshtein distantziak aldiz, kate edo item batetik bestera pasatzeko eman behar diren urrats minimoak kalkulatu ditu. Bere aldaera batean, Damerau-Levenshtein distantzian, elkarren ondoan dauden sinboloen trukaketa aldaketa bakartzat onartzen da. q-gram, batzuetan *k-mers*, *w-words* edo *k-tuples* ere deitua, n-gram bi kateetako bektoreen arteko diferentzia absolutuen batura da. ALINE distantzia transkripzio fonetikoekin erabiltzeko garatu zen. Metodo horrek funtsezko esanahi bertsua duten bi hizkeretako hitz pare bat kontuan hartzen ditu. Bi hitzak lerrokatuta daudela aldaketa fonetikoak antzekotasun puntuazioa sortzeko erabiltzen dira (Downey et al., 2008; 2017).

Laugarren une kritikoa multzokatze prozesua da. Puntu honetan elementu guztiak multzoka batzeko prozesua egiten da, algoritmo baten bitartez. Ward, Average eta complete dira dialektometriari erabilienak, Diatchek aukeran ematen dituenak. Ward algoritmoak barne bariantza txikiena aukeratzen du multzoak sortzeko (Ward, 1963) eta hori da, hain zuzen ere, gure azterketan erabiliko duguna eskaintzen duen egonkortasunagatik. Baina, esan bezala, badaude beste algoritmo batzuk, hala nola, *single*, *complete*, *average*, *median*, *centroid* eta abar, zeinak erabil daitezke betiere kontuan hartuta zein den lanaren helburua. Multzokatze hori dendrograma izeneko grafiko bidez irudikatu ohi da. Horrelakoetan multzokatze prozesu osoa antzeman daiteke, prozesuaren zuhaitz itxuraren bitartez.

Eta, azkenik, bosgarren une kritikoa zenbat talde egin erabakitzea da. Behin dendrograma daukagunean, erabaki behar dugu zenbat talde egin, hau da, multzoak erakusten dituen zuhaitza nondik moztu. Dendrogramak, clusterrak nola osatzen diren eta antzekotasun maila kontutan hartuta, datuen egitura hobeto irudikatzen duten taldeen kopurua erabakitzen lagundu dezake. Azterketa honetan, dendrogramak ikusita, eta desberdintasunak ondo markatzeko, lau talde egitea erabaki dugu, sailkapen guztiak taldekatze egonkorrekin alderatzeko.

5. Emaidak

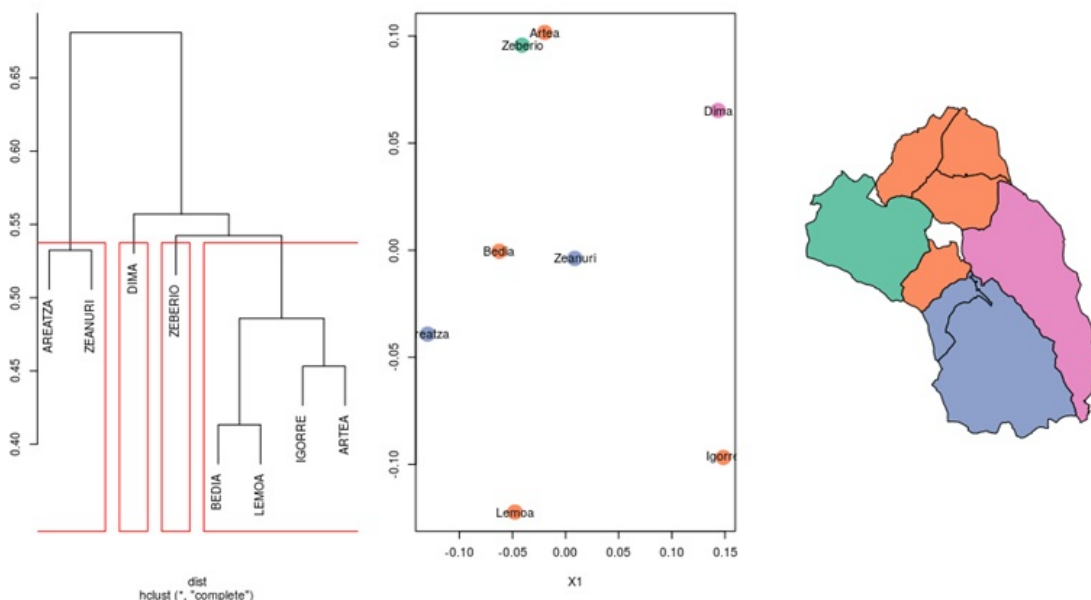
Zortzi herrik eta 1736 adizkik osatzen duten datu-basearekin lau azterketa mota egin ditugu: RIV, Levenshtein-Damerau, q-gram eta Aline. Distantzia sailkatzeko erabili den algoritmoa *Ward* izan da eta datuak tratatu barik daude, beren-beregi jaso ziren moduan erabili dira lau azterketetan.



4. irudia: Arratiako herriak

Lau azterketa hauetan clusterrek sortutako dendrograma, bi dimentsioko MDS irudia (Multidimensional Scaling) eta mapa linguistikoa eskaintzen dira.

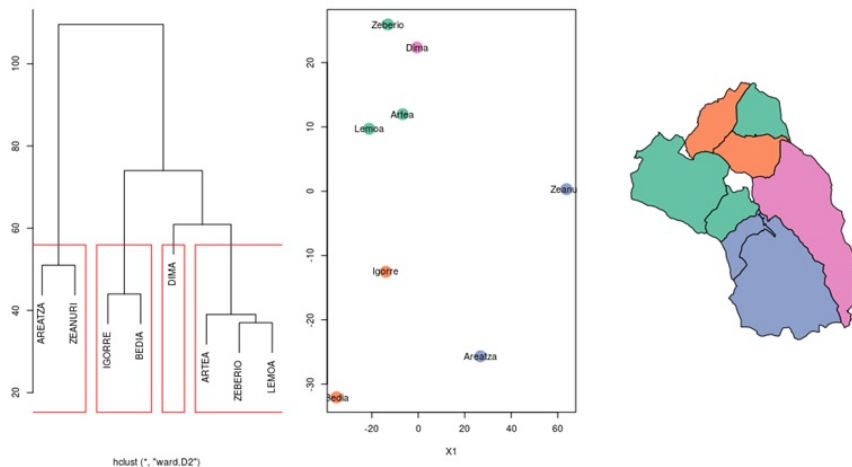
Lehenengo azterketan, RIV distantzia erabilia, 5. irudian dauden emaitzak lortzen dira.



5. irudia: RIV distantzia

5. irudiko dendrogramari erreparatuz ikusten da herri guztien arteko distantziak oso txikiak direla. Cluster honen zuhaitzean mozketa bat egin beharko bagenu talde bi bereziko genituzke: alde batetik Areatza eta Zeanuri eta, bestetik, beste guztiak. Arestian esan dugun moduan, ezberdintasunak markatzeko, lau taldetan bereizi nahi izan dugu. Horrela, dendrograman markatu dugun legez, alde batetik Areatza eta Zeanuri daude, mapan eta MDSn urdinez; Dima beste aldetik, arrosaz; Zeberio, berdez; eta azkenik lau herri, Bedia, Lemoa, Igorre eta Artea, laranja.

Bigarrenean, Levenshtein distantzia Damerau-ren egokitzapenarekin erabilia, 6. irudian dauden emaitzak lortzen dira.



6. irudia: Levenshtein distantzia

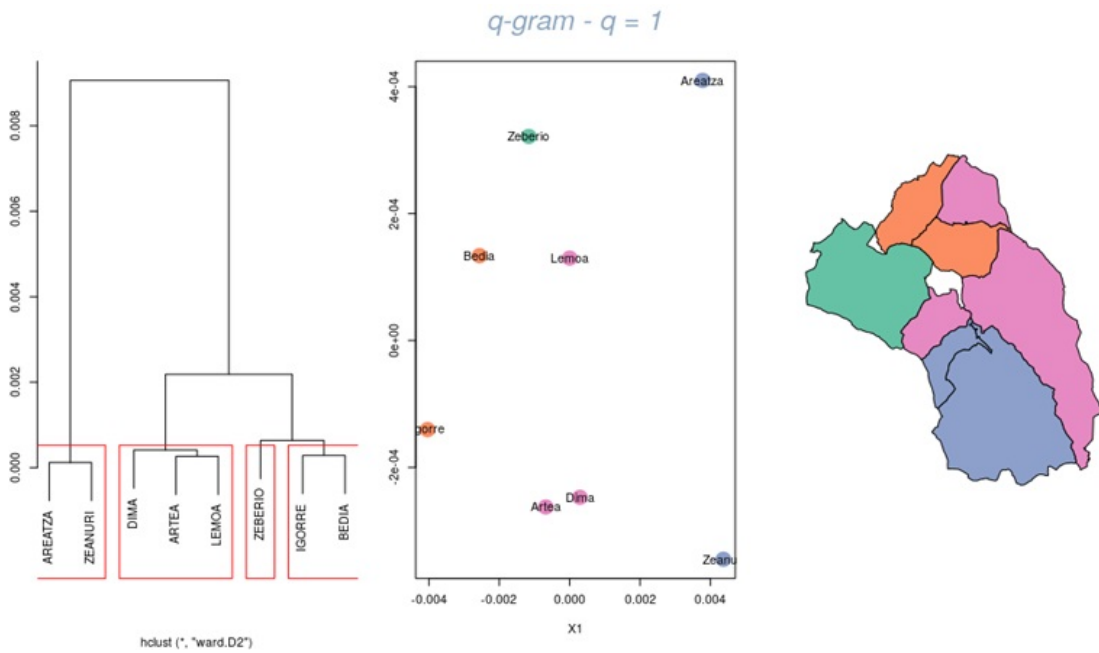
RIV distantziarekin gertatzen zen moduan, Levenshteinekin ere, clusterreko dendrogramak (6. irudia) talde bi egitea iradokitzen digu: Areatza eta Zeanuri alde batetik eta, bestetik, beste guztiak. Baina mozketak beherago egiten badugu lau talde egiteko, taldekatzeak ezberdinak dira. Areatzak eta Zenurik jarraitzen dute elkarrekin, Dimak bakarrik, arrosaz; Igorrek eta Bediak, laranjaz; eta, azkenik, Arteak Zeberio eta Lemoak, berdez.

Hirugarren azterketan q-gram (n-gram) distantzia erabili dugu. n-grama bat sekuentzia baten barne dagoen eta n elementuk osatzen duten azpisekuentzia da. 1 neurriko n-gramei unigrama deritze; 2 neurrikoei bigrama; 3 neurrikoei trigrama; eta 4tik gorakoei n-grama. Gure azterketan unigrama, bigrama eta trigrama aurkeztuko ditugu.

$d(q=1) \{d,a,r,o,i,e,s\}$	<pre> daroardies daroardies darouries daroardies 0.000 darouries 0.125 0.125 darues 0.375 0.375 0.250 </pre>
$d(q=2) \{da,ar,ro,oa,ar,rd,di,ie,es\}$	<pre> daroardies daroardies darouries daroardies 0.222 darouries 0.545 0.545 darues 0.700 0.700 0.700 </pre>
$d(q=3) \{dar,aro,roa,oar,ard,rdi,die,ies\}$	<pre> daroardies daroardies darouries daroardies 0.500 darouries 0.727 0.750 darues 0.900 0.909 0.900 </pre>

7. irudia: n-gram-en adibidea

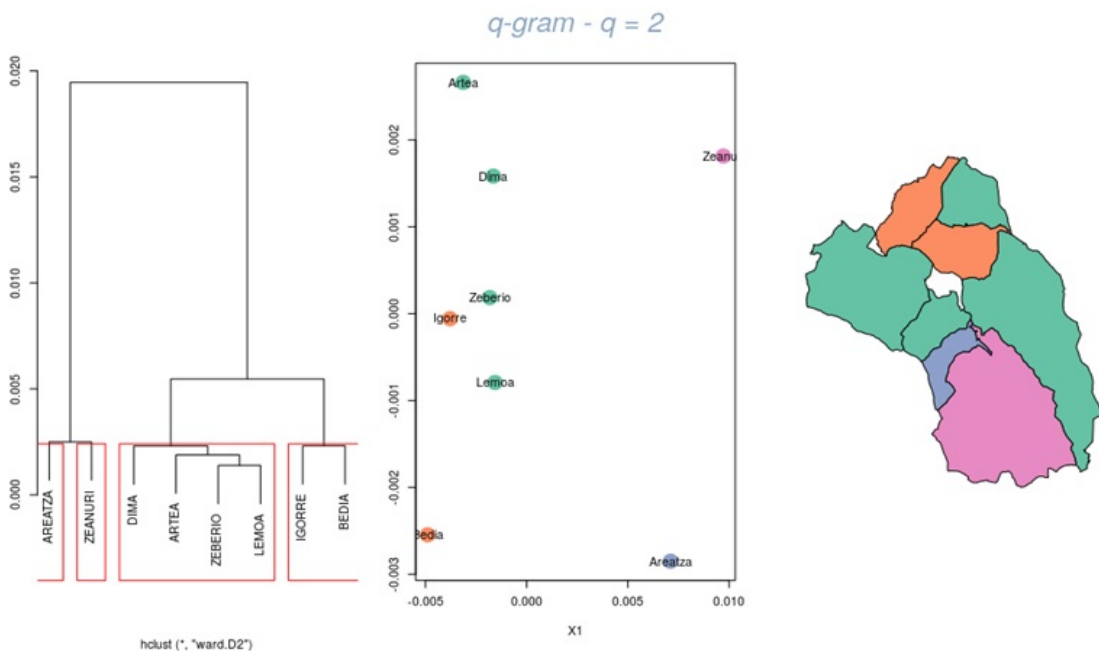
Unigramak ematen dituen emaitzak 8. irudian aurkeztu dira.



8. irudia: q-gram – q = 1

Unigramak erakusten duen dendrogramak (8. irudia) ere, multzo nagusi bi erakusten ditu. RIVek eta Levenshteinek legez, alde batetik Areatza eta Zeanuri batzen ditu eta, bestetik, beste guztiak. Ostera, lau talde egiten baditugu, emaitzak ezberdinak dira. Lehenengo taldean segitzen dute Areatzak eta Zeanurik, urdinez; bigarren talde batean arrosaz, Dimak, Arteak eta Lemoak; Zeberio, berdez; eta, azkenik, Igorrek eta Bediak, laranja.

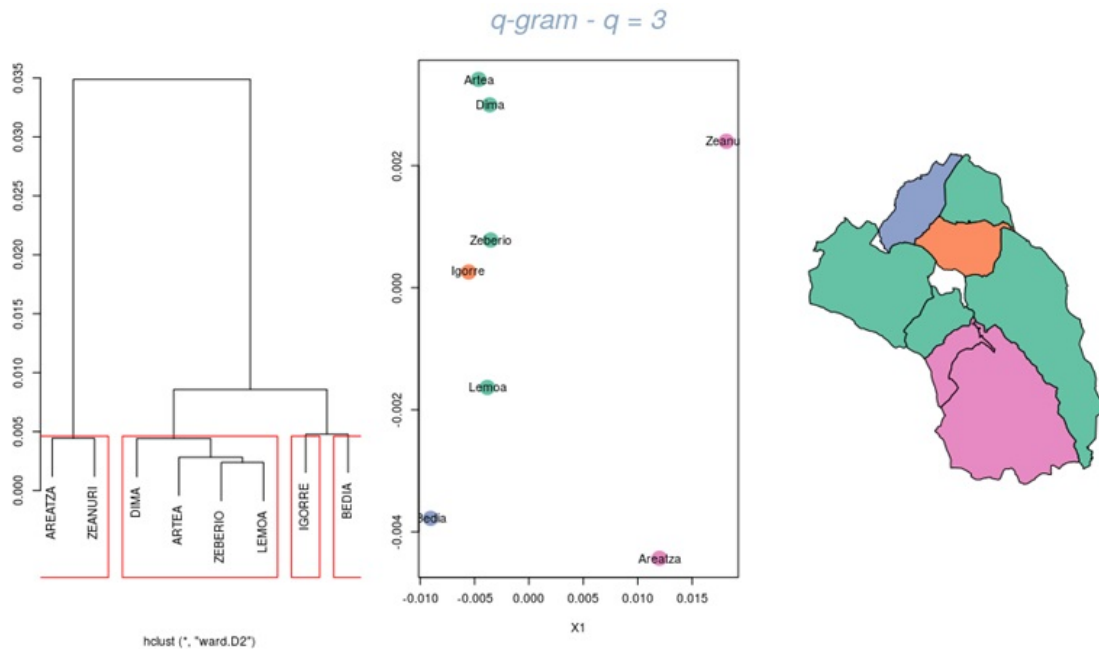
Bigramak ematen dituen emaitzak 9. irudian ditugu.



9. irudia: q-gram – q = 2

Bigramak ere (9. irudia), lehenengo mozqueta batean Areatza eta Zeanuri batzen ditu, baina, atzera ere, lau talde eginez gero, emaitzak ezberdinak dira: Zeanuri, arrosaz; Areatza, urdinez; Igorre eta Bedia, laranja; eta, azkenik, Dima, Artea, Zeberio eta Lemoa, berdez.

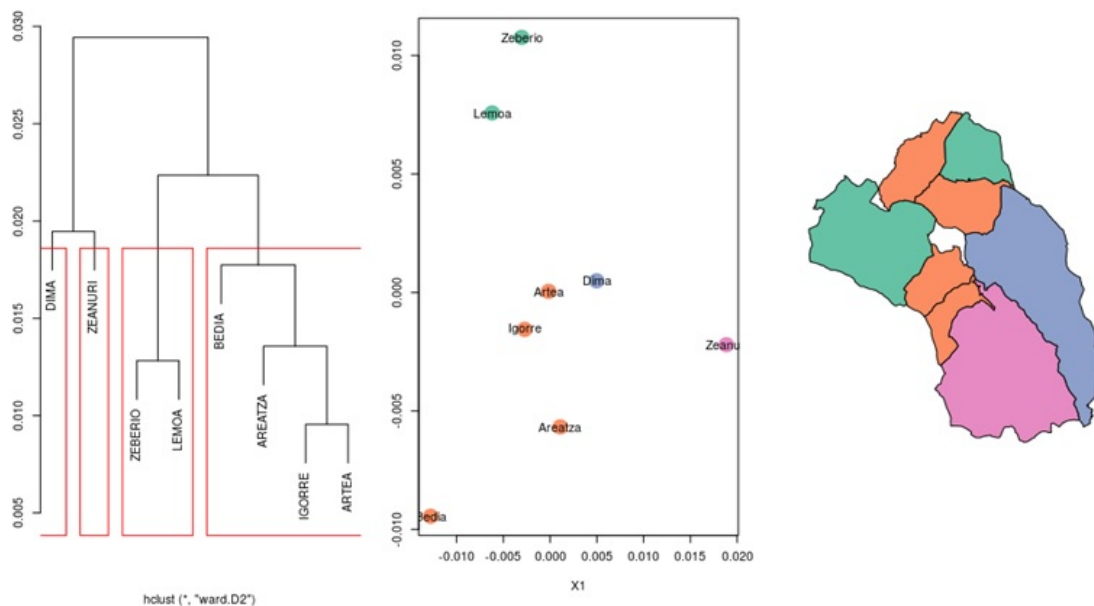
Trigramak erakusten dituen emaitzak 10. irudian ditugu.



10. irudia: q-gram – q = 3

Trigrama distantziarekin lortutako dendrogramak (10. irudia) ere, hasierako mozqueta bera iradokitzen du: Areatza eta Zeanuri multzo batean. Baina, aurrekoetan bezala, lau talde bereiziz gero, emaitzak ezberdinak dira. Horrela, Areatza eta Zeanuri ditu multzo batean, arrosaz; berdez, Dima, Artea, Zeberio eta Lemoa; laranja, Igorre; eta urdinez, Bedia.

Azkenik, laugarren azterketan, ALINE distantzia erabilia, 11. irudian dauden emaitzak lortzen dira.



11. irudia: ALINE distantzia

Areatza eta Zeanuri batzen ez dituen distantzia bakarra ALINE da (11. irudia). Lehenengo mozketan batean, dendrogramak iradokitzen du alde batetik Dima eta Zeanuri egon behar direla eta, beste alde batetik, beste herri guztiak. Mozketan beherago egiten badugu, hurrengo talde hauek agertzen dira: Dima, urdinez; Zeanuri, arrosaz; Zeberio eta Lemoa, berdez; eta, azkenik, Bedia, Areatza, Igorre eta Artea, laranja.

6. Ondorioak

Dialektometriari distantziak neurtzeko erabiltzen diren metodoez eztabaidatzea izan du helburu lan honek. Azaldu nahi izan dugu corpusaren transkripzioaren arabera zein metodo aukeratu behar den. Horretarako Gaminderen (1984) corpusaren zati baten distantziak neurtzeko lau neurri erabili ditugu: RIV, Levenshtein, q-gram eta Aline. Lehenik eta behin, corpus batekin lan egiterakoan dauden bost une kritikoak aztertu ditugu eta ostean, distantziak kalkulatzeari, erreparatu diogu. Lan honetan corpusa bere horretan, argitaratu zen moduan, erabili dugu aipatutako lau distantzia hauek aplikatuz, eta aurreko atalean ikusi dugunez, emaitza guztiak ezberdinak izan dira. Egia esan, RIV, Levenshtein eta q-gram distantzien clusterreko dendrogramek hasierako mozketan bera iradokitzen dute. Multzo bi bereizten dituzte: alde batetik, Areatza eta Zeanuri herriak; eta beste aldetik, beste herri guztiak. ALINE neurria multzokatze honetatik urruntzen da, distantzia honekin Dima eta Zeanuri multzokatzen dira alde batetik, eta bestetik, beste herri guztiak. Edozelan ere, ezberdintasunak agerian utzi nahi izan ditugunez, mozketan zehar apur bat beherago aztertu dugu lau talde egiteko. Mozketan honi erreparatu badiogu, emaitzak ezberdinak dira distantzia guztiekin. Honek agerian uzten du distantzia neurria aukeratzeko duen garrantzia; izan ere, aipatu dugun moduan, multzokatzeak ezberdinak izango dira distantzia neurriaren arabera.

7. Erreferentziak

- Arretxe, J. (1994): *Basauriko euskara*. Basauriko Udala
- Aurrekoetxea, G., Iglesias, A., Santander, G. & Usobiaga, I. (2015): ‘Diatech’ muga dialektalak zehazteko tresna. In Aurrekoetxea, G., Romero, A. & Etxebarria, A. (arg.). *Linguistic variation in the basque and education I*. Bilbo: UPV/EHU, 132-144.
- Aurrekoetxea, G. & Ormaetxea, J.L. (2006): “Euskararen Atlas Sozio-Geolinguistikoa” ikerketa-proiektua. *Euskalingua* 9, 157-163.
- Aurrekoetxea, G., Sánchez, J. & Odriozola, I. (2009): “EDAK: A Corpus to Analyse Linguistic Variation”. In Cantos, P. & Sánchez, A. *A Survey on Corpus-based Research Panorama de investigaciones basadas en corpus*. Murtzia: Asociación Española de Lingüística del Corpus, 489-503.
- Aurrekoetxea, G., & Videgain, X. (1993). Euskal Herriko Hizkuntz Atlasa: Galdesorta. *Euskera: Euskaltzaindiaren lan eta agiriak*, 38(3), 529–648.
- Aurrekoetxea, G., Videgain, X. & Iglesias, A. (2004): "Bourciez" bildumako euskal atlasa (BBEA): I. Lexikoa. ASJU XXXVIII-2.
- Aurrekoetxea, G., Videgain, X. & Iglesias, A. (2005): "Bourciez" bildumako euskal atlasa (BBEA): II. Gramatika. ASJU XXXIX-1.
- Briz, A. & grupo Val.Es.Co (2002): *Corpus de conversaciones coloquiales*. Madrid: Arco/ Libros.
- Carrera, I. M. & García, I. (2005): *Legutiano aldeko euskara*. Gasteiz: Arabako Foru Aldundia.
- Dörnyei, Z. (2007): *Research methods in Applied Linguistics*. Oxford: Oxford University Press.
- Downey, S. S., Guowei S. & Norquest, P. (2017): AlineR: An R Package for Optimizing Feature-Weighted Alignments and Linguistic Distances. *The R Journal* 9 (1): 138-52.
- Downey, S. S., Hallmark, B., Cox, M. P., Norquest, P., & Lansing, J. S. (2008). Computational Feature-Sensitive Reconstruction of Language Relationships: Developing the ALINE Distance for Comparative Historical Linguistic Reconstruction. *Journal of Quantitative Linguistics*, 15(4), 340–369. DOI:10/bf4f89
- Etchebest, X. (2017): *Zuberotar Dardarkarien Bariazio Sozio-Fonetikoa*. Doktorego-tesia. UPPA & UPV/EHU.
- Etxaide, A. M. (arg.) (1984): *Erizkizundi Irukoitza*. Euskaltzaindia, Iker 3.
- Etxebarria, M. (2016): *Gorbeiak etno-ipuina eta esaundak II*. Bilbo: Euskaltzaindia & Deustuko Unibertsitatea.
- Gaminde, I. (1984): *Aditza Bizkaieraz*. Iruñea: UEU.
- Gaminde, I. (1994): *Bilboko Larraskituko euskaraz: Lexikoa eta Testuak*. Bilbo: Bilboko Udala.
- Hidalgo, A. & Sanmartín, J. (2005). Los sistemas de transcripción de la lengua hablada, *Oralia*, 8, 13-36.
- Linell, P. (2005): *The written language bias in linguistics. Its nature, origins and transformations*. Londres: Routledge.
- Makazaga, J. M. (2011): *Elgoibarko ahozko euskara*. Bilbo: Euskaltzaindia.
- Ormaetxea, J. L. (2005): *Aramaioko euskara*. Doktorego tesia. UPV/EHU.
- van der Loo, M. P. J. (2014): The stringdist package for approximate string matching. *The R Journal*, 6(1), 111-122.
- Ward, J. H., Jr. (1963): Hierarchical grouping to optimize an objective function, *Journal of the American Statistical Association*, 58, 236–244.
- Zuazo, K. (2014): *Euskalkiak*. Donostia: Elkar.

UMEEI ETA NAGUSIEI EGINEKO TESTUEN PROSODIAREN ALDEEZ¹⁵

Naia Eguskiza, Irati De Pablo eta Iñaki Gaminde

^{1,2,3} Euskal Herriko Unibertsitatea (UPV/EHU)

naia.egusquiza@ehu.eus, irati.depablo@ehu.eus, inaki.gaminde@gmail.com

Laburpena

Lan honen helburua da egoera bitan informatzaile batek ekoiztutako testuen adierazle prosodiko batzuk erkatzea, ikusteko testu motak zenbaterainoko eragina daukan. Horretarako, 60 urteko informatzaile baten ahozko testu bi jaso dira. Informatzaileari eskatu zaio "Txanogorritxo" izeneko ipuin ezaguna kontatzeko batetik umeei egingo liekeen bezala eta bestetik, nagusiei egingo liekeen bezala.

Testu moten arteko aldeak ikertzeko aztergai ezberdinak aukeratu dira, lau multzotan banatuta: F0, energia, iraupena eta bokalen tasunak. F0ri dagokionez, talde prosodikoen batezbestekoak, heinak eta muga-tonuen erabilerak erkatu dira, jakiteko testu motaren eragina zenbatekoa den. Talde prosodikoen energiaren batezbestekoak eta heinak erkatu dira testu motaren eragina ikertzeko. Iraupenari dagokionez, denbora osoa, isiluneena, talde prosodikoa, eten beteena, txertaketa, lapsusena eta silaba nagusia kalkulatu da. Soinu mailan testu motak bokalen tasunei eragiten dien ala ez jakiteko, bokalen lehen formakin biak aztertu dira. Azterketak egin ostean agerian gelditu dira gure corpusean oinarritutako testuen arteko bereizkuntzan daragiten aldagaiak zeintzuk diren.

Hitz-gakoak: prosodia, ahozko testuak

Abstract

The aim of this work is to compare some prosodic indicators of texts produced by an informant in two situations, in order to see the influence of the type of text. For this purpose, two oral texts were received from a 60-year-old informant. The informant was asked to tell a popular story called "Little Red Riding Hood" as if she was telling it to children and the other to adults.

Different characteristics have been chosen to investigate the differences between types of texts, divided into four groups: F0, energy, duration and vowel rates. In terms of F0, the averages, ranks, and boundary tones of prosodic groups were compared, to know how much does the text type influence. The energy averages and ranks of prosodic groups were compared to investigate the effect of text type. In terms of duration, total time, silence, prosodic group, discontinuation, insertion, lapse, and major syllable were

¹⁵ Ikerketa hau burutu da IT1028/19 (Eusko Jaurlaritza) eta GIU 16/22 (Euskal Herriko Unibertsitatea) proiektuen barnean eta UPV/EHUK doktoretza aurreko formakuntzan dagoen pertsonala kontratatzeke (PIF) emandako laguntza ekonomikoari esker.

calculated. To determine whether text types affect sound rates, the first two vowel formats were analyzed. After performing the analysis, it was revealed which variables are involved in the differentiation of texts based on our corpus.

Key-words: prosody, oral texts

1. Sarrera

Lan honen helburua da egoera bitan informatzaile batek ekoitzitako testuaren adierazle prosodiko batzuk erkatzea, ikusteko testu motak zenbaterainoko eragina daukan. Informatzaileari eskatu zitzaion “Txanogorritxo” izeneko ipuin ezaguna kontatzeko batetik umeei egingo liekeen bezala eta, bestetik, nagusiei egingo liekeen bezala.

Informatzaileen ahotik harako testuen ebakeraren aldakortasunaren azterketa egin nahi dugunean, kontuan hartu behar dugu ahozko testu guztiak ez direla berdinak eta continuum batean azaltzen zaizkigula. Ñabardurak ñabardura, continuum horren mutur batean bat-bateko testuak edukiko genituzke; argi dago testu motak ere zelan edo halan eragin diezaiokeena testuaren ebakerari eta, interesgarria izan liteke, hizkuntzaren didaktikan esate baterako, hain erabiliak diren testu tipologietan agertzen diren motek zenbaterainoko eragina daukaten ikertzea.

Continuumaren beste muturrean testu irakurriak egongo lirateke. Bien arteko alderik nabarmenena honelakoetan testua egin behar ez izatea da; testua ez da eraiki behar, eginda dago eta. Testu idatziak ez dira beti bat, zailtasun maila desberdinak eduki ditzakete: ez da gauza bera poema bat irakurtzea edo laburdurez eta zenbakiz beterik dagoen testu bat irakurtzea, adibide bi ematearren.

Bada testu irakurrietan nabarmendu daitekeen berezitasun bat: testu batzuk isilpean irakurtzeko idatzi dira eta beste batzuk, berriz, ahoz gora irakurtzeko idatzi dira edo moldatu dira. Arazo honezaz, irakurketa ozenaren didaktikaren ikuspuntutik ohar batzuk egin izan dira orain artean ere (Gaminde & Goikoetxea, 2005; Gaminde et al., 2006).

Muturreko egoera hauen artean xehetasun asko egin daitezke; horrela bada, badaude buruz ikasten diren testuak eta gero errezipitatu egiten direnak. Testu mota hau irakurrietatik hurbilago dago bat-batekoetatik baino.

Badaude, era berean, aldez aurretik jende aurrean esateko prestatzen diren testu idatziak. Adibide franko aipatu dezakegu, hala nola, biltzarretako hitzaldi, ponentzia eta komunikazioetan egileek jaulkitzen dituztenak; politiko profesional batzuek prentsaurrekoetan erabiltzen dituztenak; irakasleek ikasleen aurrean azalpenak ematean bulegoan aldez aurretik prestatu dituztenak; esatari profesional askok kameraren aurrean “naturaltasunez” egiten dituztenak, eta abar. Testu hauek guztiak ez dira guztiz bat-batekoak eta maila bateko edo besteko antzezpena daukate; hauek, jakina, hurbilago leudeke bat-batekoetatik irakurrietatik baino (honako aukeratu dugun testu mota hemen kokatuko litzateke

betetzat hartu ditugu. Hortaz, eten beteak isilune biren artean dauden soinu zatiak dira eta lotura egiten duten arren, ezin jo daitezke hizkuntzaren beraren soinutzat; hau da, ez daukate esangura lexiko zehatzik.

Batzuetan soinu txertaketa batzuk gertatzen dira; gehienetan “e” eta “a” bokalak izaten dira. Hauek ere ez daukate esangura lexiko zehatzik eta eten beteen antzekoak direla esan dezakegu. Txertaketak hizketa zatiaren hasieran, erdian eta amaieran egon daitezke.

Ager dakigukeen beste soinu multzo berezi bat lapsusek osatzen dute. Hauek hizkuntzaren soinutzat jo daitezkeen arren, ez daude osaturik eta gertatzen dira hiztunek euren burua zuzentzen dutenean, trabatzen direnean, gaia, aurrekoa amaitu barik, aldatzen dutenean eta antzeko egoeretan.

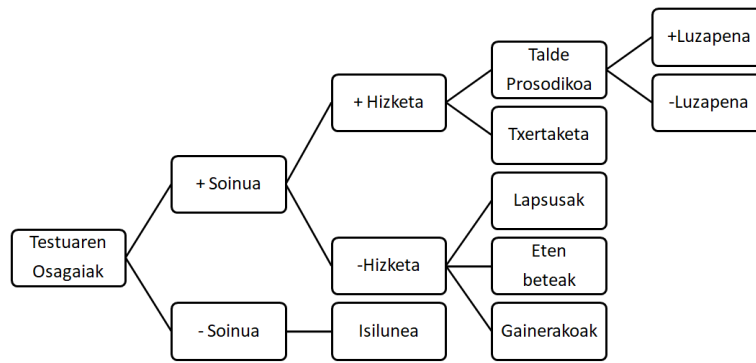
Aztertuko ditugun hizketa zatien mugetan txertaketak (tx) edo isiluneak (%) egon daitezke, aukerak lau izanik (1).

- (1) (tx) __ (tx)
- (tx) __ (%)
- (%) __ (tx)
- (%) __ (%)

Hizketa zati horiek, era berean, banatuta ager daitezke edo ez. Zenbait kasutan, barne txertaketek banatzen dute hizketa zatia, beste batzuetan soinu luzapenek mugatzen dute eta beste zenbaitetan F0 kurbaren aldaketa bortitzek sorraraz dezakete etenaldiaren efektua.

Beraz, hizketa zatien barruko multzo horiek osatuko lukete azterketa prosodikoaren oinarria. Zati horiek etenaldi biren artean dauden zatiak, alegia, talde prosodikoak (TP) dira; beste zenbaitetan (Gaminde, 2010) adierazi den bezala, euren mugak eten motek zehazten dituzte.

3. irudian testuaren osagai guztiak adierazten dira grafikoki. Osagaiak sailkatzeko erabili diren irizpideak hiru izan dira: ± isilunea, ± hizketa eta ± luzapena. Soinua dagoenean - isilunea erabiliko da eta isilunea adierazteko + isilunea. + hizketa irizpidearen bidez talde prosodikoak eta eurokin batera agertzen diren txertaketak adierazten dira; - hizketa lapsusei, eten beteei eta gainerako hizkuntzarentzat ezin har daitezkeen soinuei dagokie. Azkenik, talde prosodikoekin batera + luzapena azken silabaren, gehienetan azken silabaren bokalaren, luzapenari dagokio eta – luzapena berori gertatzen ez denean adierazten da.



3. irudia: Testuaren osagaiak.

Orain arteko aztergaiok ebakeraren aldakortasuna aztertzeko adierazle onak ez badira ere, hizkuntzaren ikaskuntzaren eta jabeakuntzaren ikuspuntu didaktikotik adierazle onak izan daitezke testuaren kalitate maila zelanbait neurtzeko. Horrela, bada, txertaketa, eten bete eta lapsus asko biltzen dituen testua kalitate eskasagokoa izango litzateke. Era berean, isiluneen iraupena egokitu behar zaie testuaren ezaugarriari eta testuaren bidez entzuleen baitan lortu nahi diren efektuari eta testuinguruari. Edozelan ere, egokitasun maila neurtu beharko litzateke pertzepzio testen bidez.

Hurrengo azterketa-maila talde prosodikoena da. Talde prosodikoa azterkizun askoren gunea izanik, lehenengo eta behin berau nola definitzen den zehaztu behar dugu. Esan dezakegu talde prosodikoa dela eten biren artean dagoen hizketa zatia. 6. atalean etenak beren-beregi landuko ditugu; oraingoz, esan dezagun etenak isilunedunak (%) eta isilune bakoak (\$) izan daitezkeela. Orduan, talde prosodikoa ondoko era honetara irudikatuko litzateke:

$$\% \text{ TP } \% (\$)$$

Alta, badakigu txertaketak (tx) ere gertatzen ahal direla talde prosodikoaren aurrean zein ostean; beraz, definizioa egokia izanda ere, zehaztu beharko genuke ondoko era honetara:

$$\% (\text{tx}) \$ \text{ TP } \$ (\text{tx}) \%$$

Horrenbestez, egin daitezkeen konbinazioak, 4. irudian erakusten dira. Atal honen helburuetarako xehetasun posible guztiak ez ditugu bereiziko eta definizio orokorra oinarritzat hartuta, eten biren artean dauden talde prosodikoen tasun batzuk oro har aztertuko ditugu.

% (tx)	\$ TP	\$ (tx)	%
%	TP	\$ (tx)	%
% (tx)	\$ TP		%
%	TP		%

4. irudia: Etenen, txertaketen eta talde prosodikoen artean egin daitezkeen konbinazioak.

Etenak eta etenaldiak bereizi beharko lirateke; izan ere, etenalditzat jo daiteke soinu zatia hausten denean gertatzen dena, hau da, isilunea dena. Etena, berriz, gure honetan da talde prosodikoa mugatzen duena eta, ondorioz, ez da beti egoten isilune baten aurrean. Etenek testuaren osagaien egituraren eta antolaketaren berri ematen digute eta testuaren esangura ulertzen laguntzen digute. Etenak faktore batzuek eragiten dituzte; horietako batzuk fisiologikoak (arnasa hartu beharra), semantikoak (esangura interpretatzen laguntzen dutenak), logikoak (kateatzeari dagozkionak), psikologikoak (entzulearen arretari eusten laguntzen dutenak) eta pragmatikoak (estrategia komunikatiboekin lotzen direnak) izan daitezke.

Eten motak sailkatzeko oso irizpide eta maila desberdinak proposatu dira bibliografian; horrela bada, maila bitik bostera aurkitzen ditugu. Maila bi erabiltzen dituztenek eten luzeak eta laburrak bereizten dituzte (adibidez Garcia-Arriola & Iruskieta, 2016), hiru maila bereizten dituzten batzuek laburrak, ertainak eta luzeak bereizten dituzte; ToBIren erabiltzaileen artean lau maila bereizten dituztenak (Jun & Elordieta, 1997) eta bost bereizten dituztenak daude. Argi dago, bestalde, hauen arteko aldea oinarri teorikoetan eurretan datzana; horrela bada, ToBI sistema Hierarkia Prosodikoaren Teoriarekin erlazionatzen da eta aurrekoek ez dute argi uzten zein teoria datxekion sailkapenari.

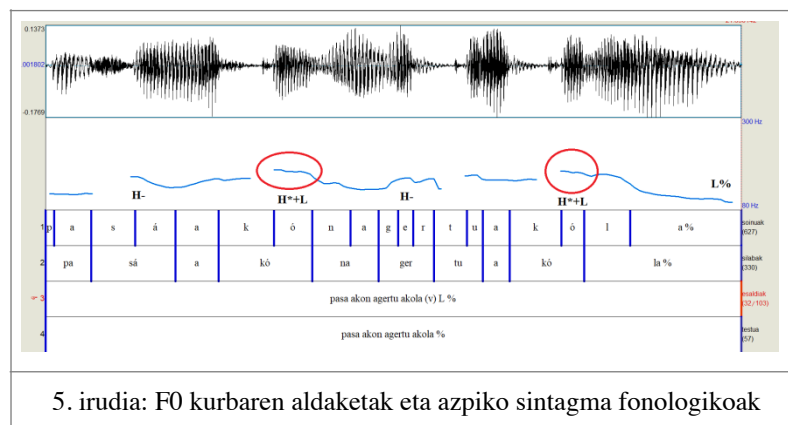
Behin honezkero hainbat lekutan esan dugun bezala (Gaminde 2004, 2006, 2010; Gaminde et al., 2012, 2014), etenak isilunearen bidez egin daitezke edo isilunea egin barik. Horren arabera batzuk isilunedunak izango dira (%) eta beste batzuk hain bakoak (\$). Zenbaitetan kurba zatien arteko loturak egiteko edo pentsatzeko denbora hartzeko bokalak luzatu egiten dira isilunearen aurrean (honetan $\pm v$ erabiliko dugu). Luzapenaren kasuan, luzatzen den bokala hainbat luzeagoa izaten da bokal arrunten batezbestekoaren aldean. Beraz, irizpide bi erabiliko ditugu etenak sailkatzeko: \pm isilunea eta \pm luzapena. Denak konbinatuta lau aukera gertatzen dira:

- (a) + isilunedun etenak / - bokal luzapena (%)
- (b) + isilunedun etenak / + bokal luzapena (v%)
- (c) - isilunedun etenak / - bokal luzapena (\$)
- (d) - isilunedun etenak / + bokal luzapena (v\$)

Sintagma fonologikoak talde prosodikoaren azpiko mailan dauden osagaiak dira eta, bigarren atalean adierazi dugun bezala, gailur batek bere menpean hartzen duen zatia da; bestalde, euren azpian silabak daude. Honek esan nahi du euren barruan talde klitikoak ere kokatzen direla eta gure lan honetan berba fonologikoak ez ditugula kontuan hartuko. Sintagma fonologikoek prosodiaren eta sintaxiaren osagaien arteko erlazioaren berri ematen digute.

Talde prosodiko batek sintagma fonologiko bakarra eduki dezake eta, horrenbestez, gailur bakarra egongo litzateke bertan. Talde prosodikoan gailur bat baino gehiago egonez gero, orduan sintagma fonologiko batzuek osatua izango litzateke; hau da, zenbat gailur, beste

horrenbeste sintagma fonologiko. Aztertzeko metodologiari dagokionez, lehen urratsean F0 kurban zenbat aldaketa bortitz dagoen ikusi behar da, 5. irudian bezala. Behin gailurrak identifikatuz gero, testuaren transkripzioari begiratu behar zaio jakiteko zer den gailurraren azpian kokatzen den talde prosodikoaren zatia; honetan, sintaxiaren osagaiek lagundu diezagukete erabakia hartzen, hau da, sintagma osoak izango lirateke gailurraren eremuak. Arazo nagusia da sintagma muga tonuen eta gailurren artean bereiztea. Sintagma fonologikoak beti ez datoz bat sintaxiaren sintagmekin; izan ere, lokailuak sintagmekin batera, sintaxiaren sintagma bi baino gehiago, aditz trinkoak klitizatuta, sintaxiaren sintagma batzuk hautsita, eta abar ager daitezke. Azterketaren etapa honetan azentuarekin lotzen diren azentu-tonudunak, sintagma tonuak eta muga-tonuak txerta daitezke.



Silaben azterketari ekiteko silabifikazioaren arazoez ohar metodologiko batzuk egin behar ditugu. Gure lan honen oinarria informatzaileak jaulkitako seinalea izanik, esan behar dugu bertan ez dagoela silabaren inongo adierazlerik edo lortatzik eurak identifikatzeko eta markatu ahal izateko; beraz, gure silaba banaketa ez da oinarritzen irizpide akustikoetan, fonologikoetan baino.

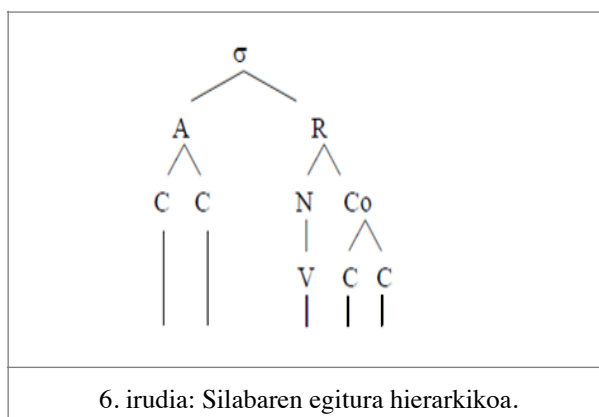
Silabak etiketatzeko eta gero aztertu ahal izateko, birsilabifikazioa hartu dugu oinarritzat, hau da, talde prosodiko osoa hartu dugu bersilabifikazioaren eremutzat. Prozesu honetan ikertzailearen intuizioa erabili da markak ipintzeko; bersilabifikazioa ezker-eskuin egin da eta, eremu horren barruko, kodak ondoko silabaren guneekin kontatu ditugu hasierak hutsak izan direneta; nahiz berori beste egoera batzuetan kodan silabifika daitekeen. 1. taulan irizpide hauen adibide batzuk erakusten dira.

Arauk	[bost] [lagun]	[bost] [ume]
Silabifikazioa	[bost] [la.gun]	[bost] [u.me]
Bersilabifikazioa	[bos.la.gun]	[bos.tu.me]

1. taula: Silabifikazioaren eta bersilabifikazioaren adibideak.

Euskaraz silabaren gunean beti egoten da bokal bat gutxienez (V) eta alboetan, berriz, kontsonanteak (C). C eta V konbinatuaz, silaben oinarritzko tipologia bat egin da: CVC, CV, VC eta V. Silaben azterketetan aurkitu diren asimetriak hobeto harrapatzeko, gunea eta koda

biltzen dituen tarteko maila bat proposatu da. Maila hori errima da; orduan, silabek 6. irudian erakusten den egitura hierarkikoa daukate.



Azkenengo edo beheengo maila soinu-maila da; arlo honetan, soinuen tasunetan testu motak daukan eragina azter daiteke. Honenbestez, ahozko testuen prosodia aztertzeko sei maila bereizten dira, 2. taulan erakusten den bezala.

Maila	Osagaiak
Testua	Hizketa zatia __isiluneak
Hizketa zatia	Talde prosodikoak, eten beteak, txertaketak, lapsusak
Talde prosodikoak	Etenak, Muga-tonuak
Sintagma fonologikoak	Gailurrak
Silabak	Hasierak, guneak, errimak, kodak
Soinuak	Tasunak

2. taula: Testua aztertzeko mailak eta mailen osagaiak

Lana lau ataletan banatuta aurkezten dugu: sarrera honen osteko bigarren atalean metodologiaz jardungo dugu, hirugarren atalean datuen azterketa aurkeztuko da eta azkenik, laugarren atalean, ondorio nagusiak laburbilduko dira.

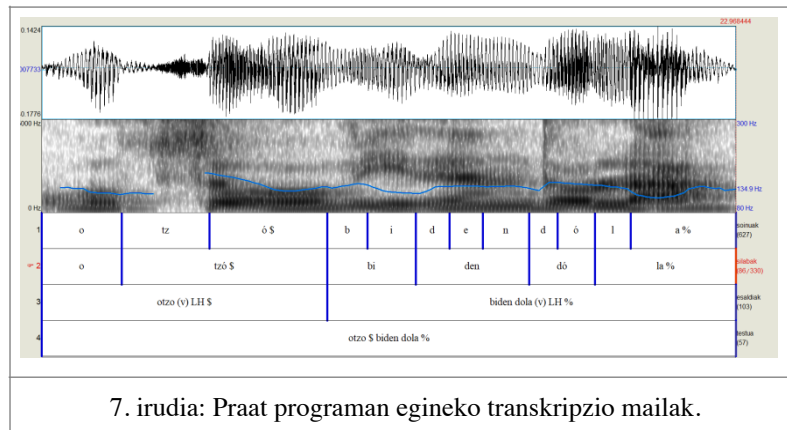
2. Metodologia

Atal honetan lana burutzeko egin ditugun hautu metodologikoak azaltzen dira, datuen jasotzeari, prozesatzeari eta aztertzeari dagokienez.

Gorago aipatu den moduan, testuak jasotzeko informatzaile baten laguntza izan dugu. Informatzailea 60 urteko andrazkoa izan da eta testu biak egiteko bere Gatikako jatorrizko barietatea erabili du, nahiz euskaraz alfabetatuta dagoen eta estandarra ere dakien.

Informatzaileari eskatu zaio “Txanogorritxo” izeneko ipuina kontatzeko umeei eta nagusiei egingo liekeen moduan. Grabazioak egiteko telefono bat erabili da “wav” formatuan grabazioa ahalbidetzen duen aplikazio bat erabilia. Testuak etiketatu eta transkribatzeko

“Praat” (Boersma & Weenink 2018) programa erabili da eta lau tira egin dira 7. irudian ikus daitekeen bezala.



7. irudia: Praat programan egineko transkripzio mailak.

Informazioa jaulkitzeko Praaterako egin diren script batzuk erabili dira eta gero izari bakoitzaren neurriekin aztertzeko beharrezko matrizeak egin dira. Sarreran aipatu diren testuak aztertzeko sei mailetatik bost mailetako adierazle batzuk landu dira, 3. taulan maila horietako bakoitzean landu diren adierazleak erakusten dira.

	Adierazleak			
Maila	F0	Energia	Iraupena	Tasunak
1. Testua			Testu osoa Isiluneak	
2. Hizketa zatia			Talde prosodikoak Txertaketa Eten beteak Lapsusak	
3. Talde prosodikoak	Batezbestekoak Heinak Muga-Tonuak	Batezbestekoak Heinak	TPen luzea Abiada	
4. Silabak			CV silabak Erritmoa	
5. Soinuak				Bokalen F1, F2

3. taula: Maila bakoitzeko landu diren adierazleak.

F0ri dagokionez, talde prosodikoen batezbestekoak eta heinak hertzetan neurtu dira. Bestalde, talde prosodikoen muga-tonuak seinaleei erreparatuta transkribatu dira eta lau muga-tonu berezi dira: H erabili da F0ren igoera talde prosodikoaren azken silaba osoan zehar gertatzen denean, L erabili da azken silaban zehar F0ren jaitsiera gertatzen denean, HL erabili da azken silabaren hasieran tonu gorakada gertatzen denean eta, amaieran beherakada gertatzen denean, LH kontrako egoeretan erabili da; hau da, azken silabaren hasieran beherakada dagoenean eta silabaren amaieran gorakada.

Energiari dagokionez, talde prosodikoen energia osoaren batezbestekoa eta heinak neurtu eta erkatu dira.

Iraupenari dagozkion azterketak maila guztietan egin dira. Hizketa mailan talde prosodikoen, eten beteen, txertaketen eta lapsusen iraupenak neurtu dira. Talde prosodikoen mailan eten motak aztertu dira eta talde prosodikoen luzera; era berean, luzerarekin lotuta dauden artikulazio-abiada eta abiada osoa neurtu dira. Silaba mailan gehien erabiltzen den CV silabaren iraupenak neurtu eta aztertu dira egongunearen arabera eta erdiko egongunean baino ez. Azkenik, soinu mailan erritmoa ere aztertu da.

Prosodiaren ezaugarri garrantzitsuenetako bat erritmoa da; berau hizketaren denbora-egiturari lotuta agertzen zaigu. Erritmoa patroï baten errepikapenak osatua da eta patroï hori ezinbestekoa da pertzepzioan. Berau hizkuntzaren fonologiaren propietatea da; tempoa, ostera, hiztunaren ezaugarriei eta hizketa egiten den baldintzei lotuta agertzen zaigu. Erritmoa aztertzeke proposatu diren metriketan, Ramus et al., (1999) lehenak izan ziren erritmoaren korrelatu akustiko behagarriak proposatzen eta metrika bat sortu zuten. Euren metrikan hiru dira neurtu eta koerlazionatzen diren parametroak:

ΔV : bokal tarteen desbiderapena

ΔC : kontsonante tarteen desbiderapena

%V: bokalen portzentajea

Indize horien arabera hizkuntza isosilabikoetan emaitza txikiagoak lortzen dira ΔV_n eta ΔC_n isoazentualetan baino (Toledo, 2010).

Tempoaren efektuak apaltzeko varco-C eta varco-V izeneko indizeak proposatu dira (Dellwo, 2006); horretarako, ondoko formulen arabera, ΔV_a edo ΔC_a bider ehun egiten dira eta gero bokalen edo kontsonanteen iraupenen batezbestekoekin zatitzen dira:

$$\text{varco}\Delta C = \frac{\Delta C * 100}{\text{mean}C}; C = \text{duration of consonantal intervals}$$

Grabe-k eta Lowk (2002) proposatutako indize bi dauzkagu, bata nPVI da eta bestea rPVI. Eurak kalkulatzeko ondoko formulak erabiltzen dira.

$$nPVI = 100 \times \left[\sum_{k=1}^{m-1} \left| \frac{d_k - d_{k+1}}{(d_k + d_{k+1})/2} \right| / (m - 1) \right]$$

$$rPVI = \left[\sum_{k=1}^{m-1} |d_k - d_{k+1}| / (m - 1) \right]$$

Azkenik, Bertinettok eta Bertinik (2008) proposatutako metrika kalkulatzeko ondoko formula erabiltzen da:

$$CCI = \frac{100}{m - 1} \sum_{k=1}^{m-1} \left| \frac{d_k}{n_k} - \frac{d_{k+1}}{n_{k+1}} \right|$$

Metrika bakoitzaren emaitzen interpretazioetan konbinazio asko dauden arren (Gibbon & Gut 2001; Prieto & Vanrell 2010), gure datuak aztertzeko orokorrenak erabiliko ditugu.

Kalkuluak eta grafikoak egin ahal izateko “Correlatore” izeneko programa erabili dugu; programa honek “Praat” programak sorturiko testu artxiboak erabiltzen ditu eta sano errazten du kalkulu prozesu osoa.

Soinuen tasunei dagokienez, bokalen formakinak baino ez ditugu landu testu bietan bakoitzaren bokalen triangeluak ateratzeko eta erkatu ahal izateko.

3. Datuen azterketa

Atal honetan testu moten arteko aldeak aztertzeko aukeratu ditugun aztergaiak lau multzotan banatuta aurkeztuko ditugu: F0, energia, iraupena eta bokalen tasunak.

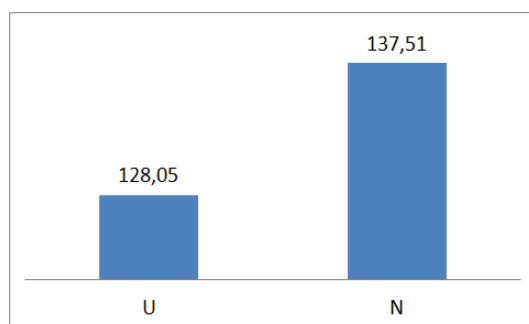
3.1. F0

F0ri dagokionez, talde prosodikoen batezbestekoak, heinak eta muga-tonuen erabilera erkatuko ditugu, jakiteko testu motaren eragina zenbatekoa den.

4. taulan talde prosodikoen F0ren batezbestekoak (\bar{x}) desbideratzeak (sd), minimoak (min), maximoak (max) eta heinak ematen dira, bai umeei (U) egineko testuan, bai nagusiei (N) eginekoan). Bertan ikusten den moduan, nagusiei egiten zaienean F0ren batezbestekoa altuagoa da umeei egiten zaienean baino (8. irudiko grafikoa); alde hau estatistikoki esanguratsua da T probaren arabera ($T = -2,509$ (a.m.: 130), $p = 0,013$).

	\bar{x}	sd	min	max	heina
U	128,05	21,68	96,56	199,80	103,24
N	137,51	21,06	97,33	199,84	102,51

4. taula: Talde prosodikoen F0ren datu orokorrak testu bakoitzeko.



8. irudia: F0ren batezbestekoa testu bakoitzeko.

5. taulan talde prosodikoen F0ren heinei dagozkien datuak ematen ditugu. Bertan erakusten den bezala, hein handixeagoa erabiltzen da nagusien testuan umeen testuan baino; hala ere, alde hau ez da estatistikoki esanguratsua.

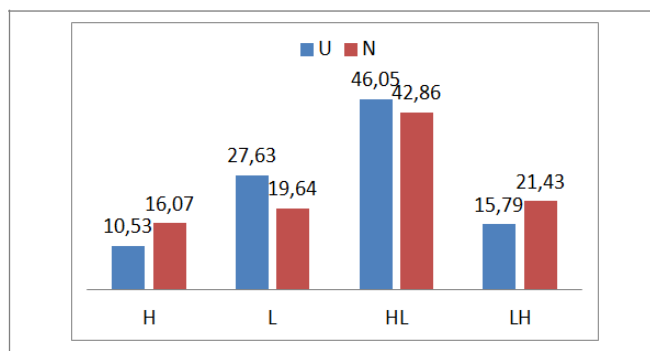
	\bar{x}	sd	min	max	heina
U	60,24	29,44	11,00	136,62	125,62
N	65,93	32,78	12,43	158,34	145,91

5. taula: Talde prosodikoen F0ren heinen emaitzak testu bakoitzeko.

Muga-tonuei dagokienez, 6. taulan agertu zaizkigunen kopuruak eta ehunekoak ematen ditugu testu mota bakoitzeko (9. irudiko grafikoa). Bertan agertzen diren aldeak estatistikoki esanguratsuak izan ez arren, alderik handienak L eta LH muga-tonuen erabileraren ehunekoetan gertatzen direla argi agertzen da.

	U		N	
	Kopurua	%	Kopurua	%
H	8	10,53	9	16,07
L	21	27,63	11	19,64
HL	35	46,05	24	42,86
LH	12	15,79	12	21,43
Orotara	76	100	56	100

6. taula: Muga-tonuen kopuruak eta ehunekoak testu mota bakoitzeko.



9. irudia: Muga-tonuen ehunekoak testu mota bakoitzeko.

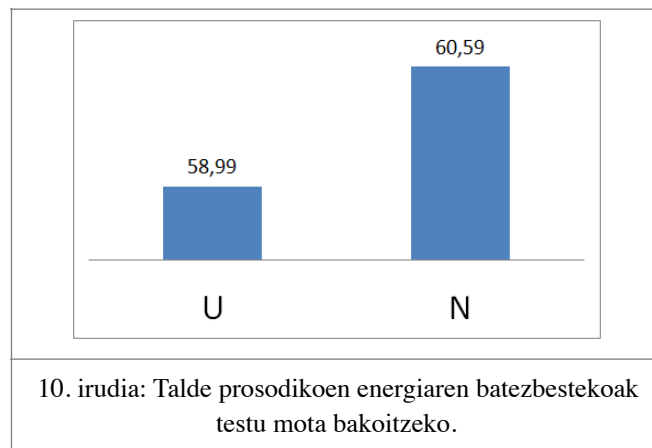
3.2. Energia

Aurreko azpiatalean F0rekin egin dugun bezala, oraingo honetan talde prosodikoen energiaren batezbestekoak eta heinak erkatuko ditugu testu motaren eragina ikertzeko.

7. taulan talde prosodikoen energiaren gaineko datu orokorrak ematen dira. Bertan ikusten den bezala, nagusiei egineko testuaren talde prosodikoen energia altuagoa da umeei eginekoaren talde prosodikoena baino (10. irudiko grafikoa); alde hau estatistikoki esanguratsua da ($T = -2,348$ (a.m.: 130), $p = 0,020$).

	\bar{x}	sd	min	max	heina
U	58,99	3,63	51,43	67,72	16,29
N	60,59	4,19	50,40	71,39	20,99

7. taula: Talde prosodikoen energiaren gaineko datu orokorrak.



8. taulan talde prosodikoen energiaren heinei dagozkien datuak ematen ditugu. Bertan erakusten den bezala, hein handixeagoa erabiltzen da umeen testuan nagusienean baino; hala ere, alde hau ez da estatistikoki esanguratsua.

	\bar{x}	sd	min	max	heina
U	25,02	8,22	4,99	38,87	33,88
N	24,47	7,34	6,04	36,96	30,92

8. taula: Talde prosodikoen energiaren heinei dagozkien datuak.

3.3. Iraupena

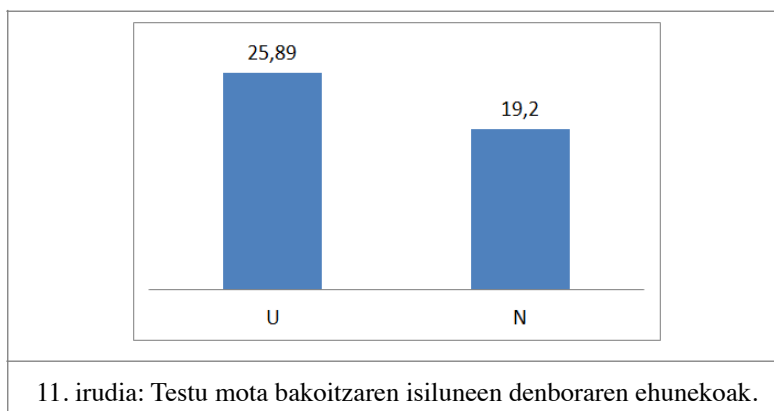
Azpiatal honetan iraupenaren datuen azterketa aurkezten dugu maila desberdinetan eginda: denbora osoa, isiluneena, talde prosodikoena, eten beteena, txertaketena, lapsusena eta silaba nagusiena. Bestetik hauekin erlazionatutako eten motak, talde prosodikoen luzera, abiada eta erritmoa ere aztertuko dira.

Denbora osoari dagokionez, umeei zuzendutako testua (102,940 s.) luzeagoa da nagusiei zuzendutakoa baino (73,030 s.).

Testu bakoitzaren isilune kopurua eta isiluneen denbora 9. taulan erakusten dira. Bertan ikusten den bezala, umeei zuzendutako testuan isilune kopuru handiagoa egiten da eta isilune guztien denboraren ehuneko handiagoa da (11. irudiko grafikoa).

	Kopurua	Denbora	Denbora Osoa	%
U	45	26650	102940	25,89
N	27	14020	73030	19,20

9. taula: testu bakoitzaren isilune kopurua eta isiluneen denbora.



11. irudia: Testu mota bakoitzaren isiluneen denboraren ehunekoak.

Isiluneen iraupena erkatzen dugunean (10. taula), ikus dezakegu isiluneak oso desberdinak direla euron iraupenari dagokionez; aldeak estatistikoki esanguratsuak ez diren arren, umeei egiten zaien testuan agertzen den heina askozaz handiagoa da nagusiei eginekoan baino.

	\bar{x}	sd	min	max	heina
U	595,11	499,74	70	2490	2420
N	519,26	433,44	80	1530	1450

10. taula: Isiluneen iraupenaren emaitzak testu mota bakoitzeko.

Hizketa zatiaren osagaien (talde prosodikoak (TP), txertaketak (tx), eten beteak (EB) eta lapsusak (lap)) kopuruak eta denborak 11. taulan erakusten direnak dira. Bertan agertzen den bezala, aldeak oso txikiak dira kasu guztietan.

	U			N		
	Kopurua	Denbora	%	Kopurua	Denbora	%
TP	76	69280	67,30	56	50570	69,25
tx	13	5360	5,21	13	4850	6,64
EB	2	1200	1,17	4	2560	3,51
lap	1	320	0,31	1	1030	1,41

11. taula: Hizketa zatiaren osagaiak kopuruak eta denborak.

12. taulan talde prosodikoen iraupenari dagozkion datuak erakusten dira; batezbestekoetan agertzen diren aldeak oso txikiak dira eta ez dira estatistikoki esanguratsuak. Hala ere, aldakuntza koefizienteak oso desberdinak dira, umeei egineko testuan % 47,04 eta nagusiei eginekoan % 72,36.

	\bar{x}	sd	min	max	heina
U	911,58	428,80	330,00	2250,00	1920,00
N	903,04	653,45	240,00	4250,00	4010,00

12. taula: Talde prosodikoen iraupenaren emaitzak testu mota bakoitzeko.

Talde prosodikoen luzera aztertzean 13. taulako emaitzak lortzen ditugu; batezbestekoetan agertzen diren aldeak estatistikoki esanguratsuak izan ez arren, aldakuntza koefizienteak oso desberdinak dira % 63,83 umeei egineko testuan eta % 93,17 nagusiei eginekoan.

	\bar{x}	sd	min	max	heina
U	4,68	2,99	1,00	16,00	15,00
N	5,13	4,78	1,00	29,00	28,00

13. taula: Talde prosodikoen luzeraren emaitzak testu mota bakoitzeko.

Talde prosodikoen silaba kopurua eta iraupena korrelazionatzean ikusten dugu umeei egineko testuan korrelazioa positiboa eta handia dela ($r=0,874$, $p = 0,000$) baina, nagusiei egineko testuan korrelazioa handiagoa da ($r=0,97$, $p = 0,000$).

Korrelazioa ez da erabatekoa silabaren iraupena egongunearen arabera aldatu egiten delako eta testuaren arabera ere, geroxeago ikusiko dugun bezala. 14. taulan testu bietako CV silaben iraupenak ematen dira egongunearen arabera. Bertan ikusten den bezala, alderik nabarmenena amaierako silaben batezbestekoetan gertatzen da. Kruskal-Wallis estatistikoak adierazten du aldea estatistikoki esanguratsua dela ($\chi^2_{(a.m.:2)} = 28,923$; $p = 0,000$). Hala ere, binan-binan erkaketak egitean, ikusten dugu hasierako eta erdiko silaben batezbestekoetan dagoen aldea ez dela estatistikoki esanguratsua; bai, ostera, hasierakoen eta amaierakoen artekoa ($z = -4,706$; $p = 0,000$) eta erdikoen eta amaierakoen artekoa ($z = -5,344$; $p = 0,000$).

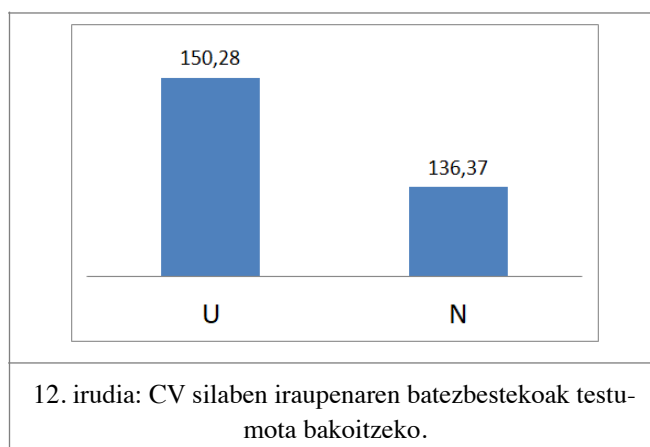
	\bar{x}	sd	min	max	heina
hasieran	142,81	45,31	70,00	310,00	240,00
erdian	144,05	51,55	70,00	490,00	420,00
amaieran	334,17	109,00	160,00	570,00	410,00

14. taula: Silaben iraupenaren emaitzak egongunearen arabera.

Gehien erabiltzen den silaba mota CV da. Testu bietako silaba mota hauen iraupena neurtu dugu erdiko gunean eta C baten aurrean gertatzen denean, hau da, CV.C ingurunean. 15. taulan datuak erakusten dira. Bertan agertzen den bezala, umeei egineko testuan silabak luzeagoak dira (12. irudiko grafikoa); alde hau estatistikoki esanguratsua da ($T = 2,218$ (a.m.: 242), $p = 0,028$).

	\bar{x}	sd	min	max	heina
U	150,28	52,59	70,00	490,00	420,00
N	136,37	41,64	40,00	390,00	350,00

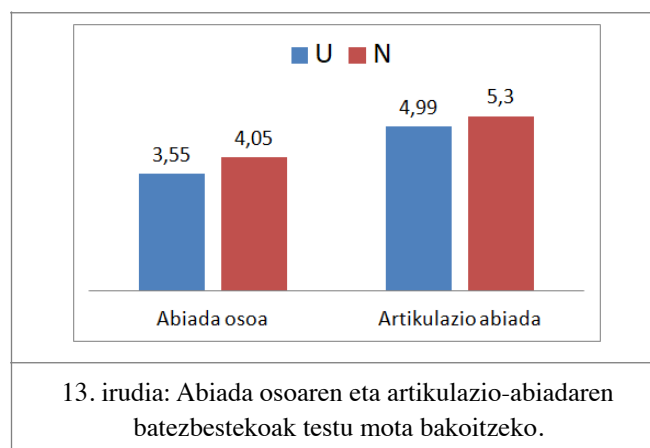
15. taula: CV silaben iraupenaren emaitzak testu mota bakoitzeko.



Gorago esan dugun moduan, abiada osoa eta artikulazio-abiada bereiz azter daitezke. Abiada osoari dagokionez, gure informatzaileak umeei egineko testuan 3,55 silaba segundoko egiten ditu eta 4,05 nagusiei egineko testuan. Artikulazio-abiadaren datuak 16. taulan agertzen dira. Bertan ikusten den bezala, nagusiei egineko testuan, abiada osoan bezalaxe, abiada altuagoa da umeei egineko testuan baino; hala ere, aldea ez da estatistikoki esanguratsua. 13. irudiko grafikoan abiada mota bien emaitzak erka daitezke.

	\bar{x}	sd	min	max	heina
U	4,99	1,66	0,97	8,33	7,36
N	5,30	1,68	1,52	9,76	8,24

16. taula: Artikulazio-abiadaren emaitzak testu mota bakoitzeko.



17. taulan txertaketen iraupenen gaineko datuak ematen ditugu. Umeei egineko testuaren txertaketen iraupenaren batezbestekoa altuagoa da nagusiei egineko testuan baino; aldea, baina, ez da estatistikoki esanguratsua.

	\bar{x}	sd	min	max	heina
U	412,31	150,89	240,00	770,00	530,00
N	373,08	243,22	60,00	940,00	880,00

17. taula: Txertaketen iraupenaren emaitzak testu mota bakoitzeko.

Eten motak aztertzeko sarreran aipatu ditugun eurok sailkatzeko irizpideak erabiliko ditugu. Orduan, \pm isilune eta $\pm v$ etenen hedadura ikusiko dugu lehenengo eta, gero, irizpide bien konbinazioen hedadura.

18. taulan \pm isilunedun etenen kopuruak eta ehunekoak erakusten dira testu motaren arabera; umeei egineko testuetan isilunedun eten gehiago erabiltzen da eta nagusiei egineko testuan, ostera, gehienak isilune bakoak dira; hala ere, aldea ez da estatistikoki esanguratsua.

	+isilunea		-isilunea	
	Kopurua	%	Kopurua	%
U	40	52,63	36	47,37
N	21	37,5	35	62,5

18. taula: Isilunedun eta isilune bako eten kopuruak testuko.

Bokal luzapenei dagokienez, 19. taulan ikusten den moduan, esan dezakegu ez dagoela alde handirik, nahiz testu mota bietan nagusitzen diren etenak bokal luzapena daukatenak izan; aldea oso txikia da eta berori ez da estatistikoki esanguratsua.

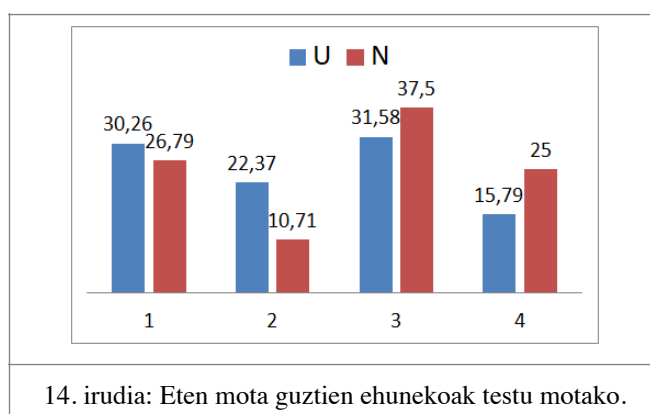
	+v		-v	
	Kopurua	%	Kopurua	%
U	47	61,84	29	38,16
N	36	64,29	20	35,71

19. taula: Bokal luzedun eta luzapen bako etenen kopuruak eta ehunekoak testuko.

Aurreko tasun biak (\pm isilunea eta \pm bokala) konbinatuta lau aukera agertzen dira: 1) +isil, +bokala, 2) +isil, -bokala, 3) -isil, +bokala eta 4) -isil, -bokala. 20. taulan aukera bakoitzaren kopuruak eta ehunekoak ematen ditugu (14. irudiko grafikoa); aukera bakoitzaren ehunekoan arabera umeei egineko testuan honela agertzen dira erreskadan: $3 > 1 > 2 > 4$ eta nagusiei eginekoan, berriz, $3 > 1 > 4 > 2$. Aldea ez da estatistikoki esanguratsua.

	+isil, +bokala		+isil, -bokala		-isil,+bokala		-isil, -bokala	
	Kopurua	%	Kopurua	%	Kopurua	%	Kopurua	%
U	23	30,26	17	22,37	24	31,58	12	15,79
N	15	26,79	6	10,71	21	37,50	14	25,00

20. taula: Eten mota guztien kopuruak eta ehunekoak testu motako.

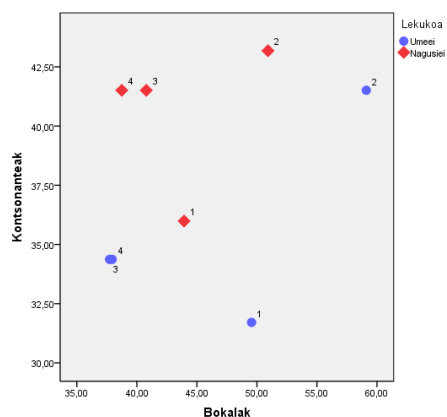


Erritmoari dagokionez, gorago esan dugun bezala, proposatu diren lau metrika aplikatzearen emaitzak lortu ditugu eta horien arabera egin ditugu erkaketak eta dispersio grafikoak. 21. taulan bokalen (V) eta kontsonanteen (C) emaitzak ematen ditugu metrika bakoitzeko (1. metrika $\%V$ eta ΔC , 2. metrika $\text{varco}V$ eta $\text{varco} C$, 3. metrika V_{npvi} eta C_{rpvi} eta 4. metrika V_{cci} eta C_{cci}). 15. irudiko dispersio grafikoan erakusten den bezala nagusiei egiten zaien testuaren erritmoaren emaitzak gorago agertzen dira umeei egiten zaienean baino.

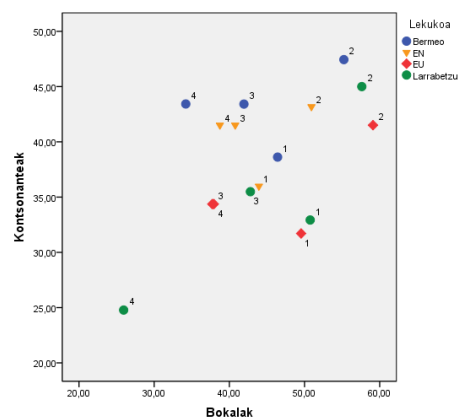
Metrika	Umeei		Nagusiei	
	V	C	V	C
1	49,55	31,71	43,93	35,99
2	59,11	41,51	50,92	43,18
3	37,73	34,37	40,78	41,51
4	37,94	34,37	38,74	41,51

21. taula: Bokalen eta kontsonanteen emaitzak metrika bakoitzeko.

Alde hauek adierazgarriak diren edo ez jakiteko, gure hemengo emaitzak erkatu ditugu Bermeon (Gaminde et al., 2012) eta Larrabetzun (Gaminde et al., 2017) lortutakoekin. 16. irudiko grafikoan ikus daitekeen bezala, nagusiei egiten zaien testuaren erritmoaren emaitzak Bermeon lortutakoetatik hurbilago agertzen dira eta, oster, umeei egineko testuarenarenak Larrabetzutik hurbilago agertzen dira.



15. irudia: Metrika bakoitzaren dispersioak.



16. irudia: Metrika bakoitzaren emaitzak testuko Bermeoko eta Larrabetzuko emaitzen aldean.

3.4. Bokalen tasunak

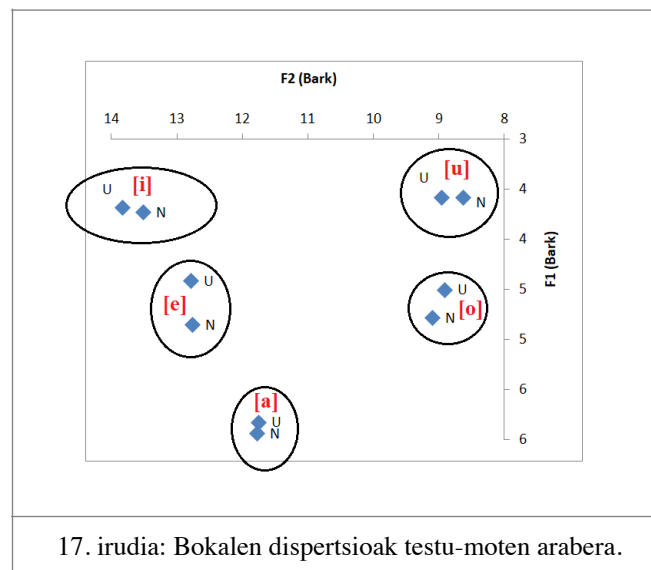
Soinu mailan testu motak bokalen tasunei eragiten dien ala ez jakiteko, bokalen lehen formakin biak aztertu ditugu. 22. taulan bokal bakoitzaren formakinen batezbestekoak eta desbideratzeak aurkezten dira; aldeak testu motaren arabera estatistikoki esanguratsuak diren jakiteko Mann-Whitney U proba estatistikoa erabili da eta taulan beronen “z” eta “p” balioak ere paratu ditugu.

		U		N		z	p
		\bar{x}	sd	\bar{x}	sd		
[i]	F1	381,78	73,66	386,53	73,22	-0,585	0,558
	F2	2246,05	208,60	2136,96	253,45	-1,529	0,216
[e]	F1	462,77	50,49	513,18	119,11	-2,054	*0,040
	F2	1910,37	184,10	1902,30	188,51	-0,819	0,413
[a]	F1	629,19	96,82	643,20	75,80	-1,201	0,230
	F2	1631,34	175,87	1638,83	150,10	-0,161	0,872
[o]	F1	472,70	54,11	505,45	64,11	-2,056	*0,040
	F2	1065,85	105,10	1097,45	118,32	-0,947	0,344
[u]	F1	370,88	39,30	370,59	37,44	-0,103	0,918
	F2	1074,94	186,24	1021,06	150,25	-0,999	0,318

22. taula: Bokalen formakinen batezbestekoak, desbideratzeak eta esanguratasuna.

Taulan ikusten den bezala [u]ren kasuan izan ezik, bokal guztiak irekiagoak dira nagusiei egineko testuan. F2ren arabera ikusten dugu goiko bokalak aurreratuago egiten direla umei egineko testuan nagusiei eginekoan baino. [o]rekin alderantziz gertatzen da, aldea txikia

izanik eta [e] eta [a] bokalekin aldea oso txikia da (17. irudiko grafikoa). Edozelan ere, aldeak [e] eta [o] bokalen F1ekin baino ez dira estatistikoki esanguratsuak.



4. Ondorioak

Atal honetan datuen azterketan jaso ditugun ondoriorik garrantzitsuenak laburbilduko ditugu. Erabili dugun corpusa eta berori jasotzeko metodologia gogotan, argi dago hemengo ondorioak behin-behinekotzat baizik ezin har daitezkeena. Hurrengo lanetarako corpusa handitu ez eze, era naturalean ere, hots, umeein batera jaso beharko litzateke. Nolanahi ere den, testuen artean badaude alde interesgarri batzuk, jarraian biltzen ditugunak atalik atal.

F0ri dagokionez, talde prosodikoen batezbestekoei begira ikusi dugu umeei egiten zaien kasuan baxuagoa dela eta aldea estatistikoki esanguratsua dela. Heina ere baxuagoa da umeei eginekoan, baina oraingo honetan aldea ez da estatistikoki esanguratsua. Testu bietan mugatona berberak erabiltzen dira eta, alderik handienak L eta LH tonuen erabileran gertatzen den arren, aldea ez da estatistikoki esanguratsua izan.

Energia aztertu ondoren ikusi dugu talde prosodikoen energiaren batezbestekoa altuagoa dela nagusiei egineko testuetan eta aldea estatistikoki esanguratsua dela. Energiaren heinari dagokionez, umeei egineko testuan handixeagoa izan arren, aldea oso txikia da.

Iraupenari dagokionez, umeei egineko testua luzeagoa da, umeei egineko testuan isilune gehiago eta isiluneen denbora handiagoa da, hizketa zatiaren osagaien denbora absolutuetan ikusi dugu ez dagoela alde handirik, nagusiei egineko testuaren talde prosodikoak luzeagoak dira, hau da, silaba gehiago biltzen dute. CV silabak luzeagoak dira umeei egineko testuetan eta aldea estatistikoki esanguratsua da. Abiadari dagokionez, bai abiada osoa, bai artikulazio-abiada handiagoak dira nagusiei egineko testuan, aldea estatistikoki esanguratsua izan ez arren. Eten mota berberak agertzen dira testu bietan eta agertzen diren aldeak ez dira oso handiak; edozelan ere, umeei egineko testuetan isilunedun etenek ehuneko handiagoa daukate nagusiei egineko testuan baino. Azkenik, erritmoa aztertzean agertzen diren aldeak oso

handiak izan ez arren, beste herri batzuetako emaitzekin erkatzean ikusi dugu nagusiei egineko testuaren emaitzak Bermeon lortutakoetatik hurbilago daudela eta umeei egineko testuarenak, berriz, Larrabetzun lortutakoetatik.

Bokalen tasunak aztertzean ikusi dugu nagusiei egineko testuaren bokalak irekiagoak direla oro har eta goiko bokalak aurreratuago egiten direla umeei egineko testuan.

Azkenik, laburbilduz, testu bien artean agertzen diren alde estatistikoki esanguratsuak hiru dira: F0ren batezbestekoa, energiaren batezbestekoa eta CV silaben iraupena. Hortaz, esan dezakegu umeei zuzendutako testuaren ezaugarri nagusiak baxuagoak, apalagoak eta silaba luzeagokoak direla.

5. Erreferentziak

- Bertinetto, P.M., & Bertini, Ch. (2008). On modeling the rhythm of natural languages, *Proceedings of Speech Prosody*. Campinas.
- Boersma, P., & Weenink, D. (2018). *Praat doing phonetics by computer*. [Computer Program] Retrieved from: <http://www.fon.hum.uva.nl/praat/>
- Dellwo, V. (2006). “Rhythm and speech rate: a variation coefficient for delta C” in P. Karnowski & I. Szigeti (arg.), *Sprache und Sprachverarbeitung: Akten des 38. Linguistischen Kolloquiums in Pilsen 2003*. <http://maincc.hufs.ac.kr/~tae/past/2010_2/phonetics_g/10_delwo06.pdf>
- Gaminde, I. (2004). Doinuak eta etenak Gatikako intonazioan. *FLV*, 97, 519-536.
- Gaminde, I. (2006). “Intonazio kurben etenez” in *Studies in Basque and Historical Linguistics in Memory of R.L. Trask*. ASJU XL:1-2 (351-376).
- Gaminde, I. (2010). *Bizkaiko Gazteen Prosodiaz: Euskaraz eta Gaztelaniaz*. Bilbo: Mendebalde Kultura Alkartea eta Bizkaiko Foru Aldundia.
- Gaminde, I., Bilbao, B., Perez, K., Goikoetxea, J. L., Goikoetxea, U., Barrutia, E., Izagirre, E. & Unamuno, M. (2006). *Irakurketa Ozena Bizkaian lantzeko materialak: Gida Didaktikoa*. Bilbo: Mendebalde Kultura Alkartea.
- Gaminde, I. & Goikoetxea, U. (2005). *Irakurketa Ebaluatzen Irizpideak Bizkaian*. Bilbo: Mendebalde Kultura Alkartea.
- Gaminde, I., Etxebarria, A., Garay, U. & Romero, A. (2012). Esaldi barruko eten moten pertzepzioaz. *Ikastorratza*, 8.
- Gaminde, I.; Etxebarria, A.; Garay, U. & Romero, A. (2014). Análisis fonológico de las emociones básicas simuladas en las variedades orientales de la lengua vasca. *Dialectologia*-12; 67-87 or.
- Gaminde, I., Olalde, A., Etxebarria, A., Eguskiza, N., & Gaminde, U. (2017). *Hizkuntza Aldakortasuna Larrabetzun*. Bilbo: Larrabetzuko Udala.
- García-Arriola, E., & Iruskieta, M. (2016). Intenziozko irakurketa ozena: lehen urratsak. *TANTAK*, 28 (2), 2016, 85-110 orr.
- Gibbon, D., & Gut, U. (2001). Measuring Speech Rhythm. *Proceedings of Eurospeech 2001*.
- Grabe, E., & Low, E. L. (2002). Durational variability in speech and the rhythm class hypothesis. *Papers in laboratory phonology*, 7(515-546).
- Jun, S. A. & Elordieta, G. (1997). “Intonational structure of Lekeitio Basque”, in A. Botinis; G. Kouroupetroglou, eta G. Carayiannis, (eds), *Intonation: Theory, Models and Applications*. Proceedings of an ESCA Workshop, Athens, Grezia, 193-196.
- Mairano, P. (2009). *Correlatore*. Retrieved from: <http://www.lfsag.unito.it/correlatore/>

- Prieto, P., & Vanrell, M.M. (2010). Speech rhythm as durational marking of prosodic heads and edges. <<http://prosodia.upf.edu/activitats/prosodicdevelopment/presentacions/prieto.pdf>>
- Ramus, F., Nespors, M. & Mehler, J. (1999). Correlates of linguistic rhythm in the speech signal. *Cognition*, 72, 1-28.
- Toledo, G. (2010). Métricas rítmicas en microdiscursos. *Onomázein* 21. <http://www.onomazein.net/21/03-TOLEDO.pdf>

Zabalduz

Jardunaldi, kongresu, sinposio, hitzaldi
eta omenaldien argitalpenak

Publicaciones de jornadas, congresos,
simposiums, conferencias y homenajes

INFORMAZIOA ETA ESKARIAK • INFORMACIÓN Y PEDIDOS

UPV/EHUko Argitalpen Zerbitzua • Servicio Editorial de la UPV/EHU
argialetxea@ehu.eus • editorial@ehu.eus
1397 Posta Kutxatila - 48080 Bilbo • Apartado 1397 - 48080 Bilbao
Tfn.: 94 601 2227 • www.ehu.eus/argitalpenak

eman ta zabal zazu



Universidad
del País Vasco

Euskal Herriko
Unibertsitatea