

Bilboko euskaldun gazteen lexikoaren erabileraz

Iñaki Gaminde

Euskal Herriko Unibertsitatea
inaki.gaminde@gmail.com

Abstract

This article is a small contribution to the discussion proposed in the Mendebalde Cultural Association on December 15, 2017. The purpose is double; on the one hand, using the lexicon of similar free texts by young speakers from Bilbao, it will analyze significant subgroups according to the mother tongue and gender; on the other hand, objective quantitative methods are also presented. The aim of this paper is to analyze the data that emerge from a sociolinguistic structure through a specific methodology, in line with adaptable dynamic complex systems. According to the results obtained, we can state that there are no significant subgroups, not even among the young speakers' mother tongue or gender, regarding the use of lexical texts that are presented here.

Laburpena

Artikulu honek ekarpen txiki bat egin gura dio Mendebalde Kultura Alkarteak 2017ko abenduaren 15ean antolatutako eztabaidari. Gure helburua hemen bikoitza da. Alde batetik, Bilboko gazte euskaldunen antzeko testu aske baten lexikoaren erabilera aztertuta, informatzaileetan azpi-multzo esanguratsurik egin daitekeen euren ama hizkuntzaren edo generoaren arabera aztertu nahi da. Beste aldetik, hori egiteko metodo kuantitatibo objektiboak ere aurkeztu nahi dira. Gure lantxo honetan datuen azpian egon daitekeen egituraketa soziolinguistikoa azalatu nahi da metodologia zehatz baten bidez, sistema konplexu dinamiko moldagarrien ildotik. Datuen azterketan eta ondorioetan argi azalduko da informatzaileetan ez dagoena inongo azpi multzo esanguratsurik egiterik, ez ama hizkuntzaren arabera, ez eta generoaren arabera ere, hemen aurkeztu diren testuen lexikoaren erabilerari dagokionez, behintzat.

Keywords: Lexical variability, quantitative method, complex systems

Gako hitzak: Lexiko aldakortasuna, metodo kuantitatiboak, sistema konplexuak.

1. Sarrera

Hemen aurkezten dugun lan honen erroak Mendebalde Kultura Alkarteak Euskaltzaindian 2017ko abenduaren 15ean antolatutako jardunaldian bilatu behar dira. Bertan Jacinto Iturbek idatzi zuen txostenaren azalpenaren zuzuan sortu zen geroko eztabaidari zor zaio hau egiteko ideia. Beraz, bihoazkie lehenengo eta behin nire eskerrik beroenak han egon ziren guztiei eta bereziki nirekin, batzuetan sutuki, eztabaidatu zutenei¹.

Batzar hartan, beste askoren artean, lexikoa, gazteak, aldakortasuna, geroa, Bilboren garrantzia eta abar ekarri ziren harira. Horregatik deliberatu dut Bilboko euskaldun gazteen lexikoaren erabileraz orain plazaratzen dudan honi ekitea. Ez da gai honi oratzen diodan lehen aldia (Gaminde, 2010b), baina oraingo honetan, ordukoaren aldean, beste ikuspuntu batetik aditu nahi diot gaiari. Bestalde, berriki hemen aurkezten dugun honen antzeko lan bat Larrabetzun egin dugula aipatzen duta (Gaminde et. al, 2017a), hemengo emaitzak hangoekin erkatuko ditugu, hiriko euskaldunenak eta herri txikikoan alderatuta, egon daitezkeen aldeak agerian jartzekotzat.

Iturbe jaunak bere txostenaren azalpenean aldarrikatu bezala, euskal gizartearen bilakaera nabarmena izan da azken urteetan eta bilakaera hori ezinbestez gomutan eduki behar da gaurko hizkuntza aldakortasuna aztertzean; are gehiago, paradigma aldaketaren beharra ere aldarrikatu zuela esango nuke, azken hau norberaren nahiaren isla errealitatearen interpretazio zuzena bainoago izan badaiteke ere. Nolanahi ere den, paradigma aldaketa horren beharra hainbat lekutan iradoki da aspaldion (Gaminde, 2007, 2010a; besteak beste).

Lan honen helburu nagusia da aztertzea Bilboko gazte euskaldunek antzeko testu aske baten lexikoaren erabilera eta ikustea informatzaileen artean azpi multzo esanguratsurik egin daitekeen euren ama hizkuntzaren edo/eta generoaren arabera, hots, alde batetik, lexikoa bera aztertzen da esparru batean eta, bestetik, populazioaren egituraketa lexiko horren arabera.

Oraintsu egindako lan batean (Eguskiza et. al. 2017) adierazi genuen bezala, hizkuntza aldakortasuna aztertzean hizkerak edo informatzaileak euren ezaugarri linguistikoen arabera sailkatzeko beharrezana sortzen da. Benetako arazoa sailkapen horiek egiteko metodo objektibo bat topatzea izaten da. Hizkera edo informatzaile biren ezaugarriak erkatzean erraza izaten da jakitea zein den euren arteko distantzia linguistikoa aukeratutako ezaugarri horien arabera; alta,

¹ Eskerrik beroenak Irati de Pablori lan honen lehen zirriborroari egindako oharrengatik.

informatzaileak edo hizkerak gehiago direnean distantziak zehaztea konplikatua izaten da eta honetan metodo kuantitatiboek laguntzan handia ekar diezagukete.

Datu linguistikoak tratatzeko metodo kuantitatiboak erabiltzea ez da gauza berria aldakortasuna aztertzean. Euskal Herrian metodo hauek aurrera egin dute aspaldiko urteotan (Aurrekoetxea, 1992, 1995, 2003, 2004, 2005, 2008, 2009, 2016; Aurrekoetxea eta Ormaetxea, 2006, 2010; Aurrekoetxea et al., 2014a, b; Gaminde, 2007, Iglesias, 2014, Ensunza, 2015, Lujanbio, 2016); metodo hauek hizkerak sailkatzeko ez ezik, informatzaileak sailkatzeko ere erabili dira (Gaminde et al. 2012, 2016a).

Gehien bat erabili den sailkatzeko metodoa multzokatze hierarkikoa izan da eta gutxiagotan dimentsio anizkun eskalarena. Azkenaldion metodo hauen gabezia bat azpimarratu da, mugan dauden hizkeren edo hiztunen sailkapenei dagokienez eta beste metodo batzuk proposatu dira logika lausoaren (lógica difusa, fuzzy logic, logique floue) hatsarreak kontuan hartuta (Clua et al., 2017). Guk gure saio honetan multzokatze hierarkikoa, K-batezbestekoen multzokatzea eta dimentsio anizkun eskala erabiliko ditugu.

Metodoak eurak aukeratu baino lehen arazo larria izaten da jakitea zeintzuk diren irizpideak ezaugarri batzuk aukeratzeko eta beste batzuk baztertzeko. Aukeratzaren diren ezaugarriak gura baino sarriagotan homogeneotzat eta absolututzat jotzen dira hiztun komunitate osoarentzat. Lan askotan (Ensunza, 2015; Lujanbio, 2016, besteak beste) erakutsi den bezala, informatzaile guztiek ez dute berdin egiten hiztun komunitate berean ez eta hiztun berak egoera guztietan ere. Hau gramatikaren esparruetan gertatzen bada ere, lexikoaren esparruan arazoa areagotu egiten da. Beste lan batean iradokitzen genuen bezala (Gaminde, 2007), lexikoa ezin aldera daiteke gramatikarekin maila berean, izan ere, kanpo eta barne faktore askoren menpe dago. Aipatu dugun lanean hainbat adibide aurkeztu genuen, baina eman dezagun aspaldion jaso dugun adibide bat: “autobusa”, “busa” eta “kotxea” hitzak gauza bera adierazteko erabiltzen dira. Gehiegi sakondu barik esan dezakegu lehena (“autobusa”) formalagoa dela eta edade batetik gorako hiztunek erabiltzen dutela (esate baterako hau da nik neuk beti esaten dudana), bigarrena (“busa”) nire ikuspuntutik informalagoa da eta nire inguruko gazte gehienek erabiltzen dutena da, azkenik, Bizkaibuseko langileek euren artean (eta batzuetan erabiltzaileekin ere bai) “kotxea” erabiltzen dute, beraz teknolektoaren terminotzat jotzen dute.

Bigarren arazoa datuen zehaztasun ezari dagokio, aldakiak agertzen direnean aldaki guztiak maila berean agertzen dira eta arauen kasuan, esate baterako, araua bezain garrantzitsua dateke adieraztea zein diren berorren murriztapenak (Gaminde, 2002, 2007).

Aurreko arazoez gain, barietateen sailkapenak lortzeko metodologia sarri askotan ez da zehazten, ez

dakigu metodo objektiboren bat erabili den edo ikertzailea gogoeten arabera egin den aurkezten den sailkapena. Batzuetan isoglosen metodoa aipatzen da baina metodo hori nola aplikatzen den ez da esaten esplizituki.

Guk aldakortasuna aztertzeke proposatzen dugun metodologiak, sistema konplexu dinamiko moldagarrien teoriaren ildotik (Léonard et al., 2015; Gaminde et al., 2016a, 2017, Eguskiza et al., 2017), aldi nagusi batzuk dauzka:

1. Datuen azterketa linguistikoa
2. Datuen azterketa kuantitatiboa
3. Informatzaileen sailkapena

Datuen azterketa linguistikoaren xehetasunez Etxebarria et al. (2016); Gaminde et al. (2016a eta 2016b; Eguskiza et al. (2017) ikus daitezke.

Azterketaren bigarren aldiak aldaki bakoitzaren agerpen maiztasunean oinarrituta, bakoitzaren agerpen probabilitateak kalkulatu dira eta probabilitate horiek erabiliko dira datu matrizeak egitean eta ez “0” eta “1”. Beste lan batean (Gaminde et al., 2017b) erakutsi genuen moduan, “0” eta “1” erabiltzeak moldeak maila berean jartzea dakar eta sailkapenen emaitzak desberdinak izaten dira. Bakoitzaren probabilitateak kalkulatzeko emaitzak doitu egiten ditu eta zehaztasun handiagoa ematen du.

Informatzaileen sailkapenaren helburua datuen azpian datzakeen egituraketa geolinguistikoa edo soziolinguistikoa azalartzea da, sistema konplexu dinamiko moldagarrien ildotik.

Lana lau ataletan banatuta aurkezten dugu, sarrera honen osteko bigarren atalean corpusaren ezaugarriak eta berori jasotzeko eta prozesatzeko metodologia aurkezten dira, hirugarren atalean datuen azterketa egiten da, azkenik, laugarren atalean, ondorio nagusiak laburbiltzen dira.

2. Corpusa eta Metodologia

Lan honetan aztertuko dugun corpusa osatzeko Bilboko 55 lekukoren datuak erabiliko dira. Lekuko guztiak bilbotarrak izan dira, 1979-1991 urte bitartean jaiotakoak euren adinaren batezbestekoa gaur egun 29 urtekoa da eta D erduan ikasi dutenak izan dira. Informatzaileen ama hizkuntzari dagokionez, hainbat lekutan adierazi dugun bezala (Gaminde, 2010a, b), hemen ere hiru talde nagusi egingo ditugu: A taldea osatzen dute familiarren bidez euskararen barietate tradizional bat jaso dutenek (ezin ahatz dezakegu hemengo informatzaileen gurasoak herri eta hizkera tradizional desberdinetakoak direla eta euren seme-alabek ez dutela jaso barietate tradizional hori berez garatzen eta bilakatzen den lekuan Bilbon baino), B taldekoek euskara ikasi dute hezkuntzan txikitzen eta C taldekoak euskaldun berrien seme-alaba euskaldun zaharrak dira, hauen kasuan guraso biak euskaldun

berriak dira. Kontuan hartu behar da talde guztietako informatzaileak elebidun goiztiarrak direla. 1. taulan talde bakoitzeko eduki ditugun informatzaile kopuruak generoaren arabera sailkatuta ikus daitezke.

	A	B	C	Orotara
Neskak	10	26	2	38
Mutilak	5	9	3	17
Orotara	15	35	5	55

1. Taula: Informatzaile kopurua ama hizkuntza eta generoaren arabera sailkatuta.

Lexikoa aztertzeko erabili den materiala, 1. irudian agertzen den komikia kontatzeko, informatzaileek erabili dituzten testuetatik atera da. Informatzaileei komikia erakutsi eta istorioa ulertzeko denbora laburra eman ondoren kontatzeko eskatu zitzaizkien zer gertatu den galderari erantzun izan baliote bezala. Audio seinaleak jasotzeko erabili diren tresnak Sony Minidisc, Marantz PMD620 eta ZOOMH4n izan dira. Grabazioen kalitatea bermatzeko kanpo mikrofonoak erabili dira kasu gehienetan eta zuzenean “wav” formatuan jaso dira.



1. Irudia: Bat-bateko testua jasotzeko erabili den komikia.

Testuak, grabatu eta digitalizatu ostean, “praat” (Boersma eta Weenink, 2016) aztergailuaren bidez etiketatuta eta transkribatu dira.

Azterketa egin ahal izateko informatzaileek erabili dituzten berba guztiak ez ditugu kontuan hartu gure corpusean; horrela, lokailuak, erakusleak, izen ordainak, aditz laguntzaileak eta gramatika erlazioak adierazten dituzten berbak baztertu ditugu. Aukeratu ditugunak eduki semantikoa (Roberto et al., 2012) daukatenak izan dira, hau da, izenak, aditzak, adjektiboak, aditzondoak, adberbioak, interjekzioak eta aditz lagunak.

Bestalde, gauza jakina da berbak deklinaturik ager daitezkeena eta aditz partizipioen kasuan aspektu markekin edo beste morfema batzuekin. Esate baterako, “aitak”, “aitaren”, “aitarekin”, e.a. agertzen badira sarrera hauek guztiak izendatzeko aukeratu genukeen

lema bakarra “aita” izango litzateke. Aditz trinkoak erabiltzen direnean partizipioa hartu da lematzat; adibidez “doas”, “doa” eta “noa” adizkien lema “joan” izan da.

Behin lematizazio prozesua burututa, corpusaren datu basea egin da, horretarako berba bakoitzari bere maiztasunaren arabera dagokion agerpen probabilitatea esleitu zaio eta 2. irudian erakusten den moduko matrize bat eraiki da. Hor agertzen diren emaitzak dira erabiliko direnak azterketa guztiak burutzeko.

Leh. egin	egin	eman	erori	esan	gizon	harri	ikusi	izini	izan	jarri
Aa01	0,00000	0,03448	0,00000	0,00000	0,00000	0,03448	0,00000	0,03448	0,00000	0,00000
Aa02	0,00000	0,03846	0,00000	0,00000	0,03846	0,00000	0,03846	0,00000	0,00000	0,00000
Aa03	0,01695	0,03390	0,01695	0,00000	0,03390	0,00000	0,05085	0,01695	0,00000	0,03390
Aa04	0,00000	0,04000	0,00000	0,00000	0,00000	0,00000	0,04000	0,00000	0,04000	0,00000
Aa05	0,00000	0,01923	0,00000	0,00000	0,01923	0,00000	0,01923	0,03846	0,00000	0,03846
Aa06	0,04348	0,06522	0,00000	0,00000	0,02174	0,06522	0,02174	0,02174	0,02174	0,00000
Aa07	0,02174	0,02174	0,00000	0,00000	0,02174	0,00000	0,02174	0,00000	0,00000	0,02174
Aa08	0,02941	0,02941	0,00000	0,00000	0,00000	0,08824	0,02941	0,05882	0,02941	0,00000
Aa09	0,01852	0,03704	0,03704	0,00000	0,00000	0,00000	0,01852	0,01852	0,05556	0,00000
Aa10	0,00000	0,04348	0,00000	0,00000	0,04348	0,00000	0,04348	0,00000	0,00000	0,00000
Ag01	0,02703	0,05405	0,00000	0,00000	0,00000	0,10611	0,02703	0,00000	0,05405	0,00000
Ag02	0,00000	0,06383	0,02128	0,00000	0,04255	0,00000	0,02128	0,02128	0,00000	0,00000
Ag03	0,00000	0,00000	0,00000	0,04878	0,02439	0,00000	0,02439	0,00000	0,00000	0,02439
Ag04	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000	0,03704	0,00000	0,00000	0,00000
Ag05	0,00000	0,00000	0,04545	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000	0,02273
Ba01	0,02500	0,05000	0,00000	0,00000	0,02500	0,07500	0,02500	0,00000	0,00000	0,00000
Ba02	0,01389	0,06944	0,00000	0,00000	0,00000	0,01389	0,00000	0,00000	0,01389	0,02778
Ba03	0,00000	0,10714	0,00000	0,00000	0,00000	0,00000	0,03571	0,03571	0,00000	0,03571
Ba04	0,00000	0,02273	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000	0,04545

2. Irudia: Berben probabilitateekin egindako matrizearen zatia.

Prozesu hau eginda, hiru eratako azterketak egin dira; bata, berben azterketa orokorretik abiatuta eta beste biak informatzaileen erabileretatik abiatuta. Lehen azterketan agertzeko probabilitaterik handienak daukaten berbak agerian jarri ditugu eta hauek erkatu ditugu Larrabatzun lortutako emaitzekin; esan behar dugu bide batez, azterketa hau bat datorrela Gamindek (2010b) egindakoarekin.

Informatzaileak sailkatzeko eta balizko egitura soziala azaleratzeko hiru metodo erabili ditugu gorago esan bezala. Matrizea eginda informatzaileak sailkatzeko prozedura hautatu behar da. Multzokatze azterketetan mota bi bereizten dira gehienetan (Pardo eta Ruiz, 2002, Pérez López, 2005), multzokatze hierarkikoak eta ez hierarkikoak. Aldakortasunaren azterketetan metodo hierarkikoak erabili izan dira nagusiki (Goebel, 2010). Era berean, metagarriak edo urrigarriak izan daitezke (Aurrekoetxea, 1995; Terrádez, d.g.). Aldakortasunaren gaineko ikerketetan metagarriak erabili dira gehienetan, hau da, multzokatze urrats bakoitzean hurbilen daudenak multzo berean sartzen dira beste multzo bat osatzeko eta horrela harik eta multzo bakarra lortu arte. Multzokatze ez hierarkikoetan edo K-batezbestekoetan (KM) aldeaz aurretik ikertzaileak multzo kopurua finkatu behar du. Hauek gain, badago logika lausoan oinarritzen den multzokatzea ere (fuzzy C-means, FCM).

Teknika hauen helburu nagusia da elementu batzuk multzokatzea talde homogeneotan, euren artean daukaten antzekotasunetan edo desberdintasunetan oinarrituta. Eurak aplikatzerakoan urrats batzuk egin behar dira, Ayugari (d.g.) jarraikirik honako hauek proposa daitezke:

1. Aldagaien aukeraketa
2. Multzokatze metodoa
3. Antzekotasun eta distantzia neurriak
4. Sailkatze algoritmoa
5. Multzoak lortu
6. Emaizen interpretazioa

Multzokatze metodoa aukeratu ondoren, kasuen artean erkaketak egin ahal izateko, antzekotasunen eta desberdintasunen distantzia neurria aukeratu behar da. Honetan asko agertzen dira bibliografian (Pardo eta Ruiz, 2002, Pérez López, 2005), baita aldakortasuna aztertzeko erabili direnetan ere (Goebel, 1992, 2010; Aurrekoetxea, 1995, 2005; Clua, 2010). Gure lantxo honetan kasu guztietan distantzia euklidearra aukeratu dugu.

Multzokatze hierarkikoa egiteko sailkatze metodoa aukeratu behar da hurrengo, hauetan ere asko daude (Pardo eta Ruiz, 2002, Pérez López, 2005), aldakortasunaren azterketetan gailendu dena Ward-en metodoa da (Goebel, 1992, 2010; Aurrekoetxea, 1995, 2005; Gaminde, 2007; Gaminde et al., 2016b).

Multzokatze hierarkikoaren emaitza dendograma bat izango da non multzo nagusiak agertuko diren; elementu bakoitza multzo bakar baten barruan agertuko da multzo horretako gainerako elementuekin daukan distantziaren arabera.

KM multzokatze mota erabiltzen badugu, gauza bera gertatuko zaigu baina oraingo honetan elementuen banaketa egingo da guk alde aurretik aukeratu ditugun multzo kopuruaren arabera. Teknika honek elementuen arteko berdintasunak eta multzoen arteko aldeak nabarmenarazten ditu.

Azkenik, lexiko aberastasuna kalkulatzeko hainbat indize proposatu dira (Roberto et al., 2012). Gure hemengo helburuetarako TTR (type token ratio) deritzona aukeratu dugu. Indize hau kalkulatzeko informatzaileek testuan erabiltzen dituzten lema desberdinak (L) eta lema guztiak (N) kontuan hartzen dira:

$$TTR = \frac{L}{N}$$

Laginaren tamaina txikiaren efektua apaltzeko (Roberto et al., 2012), $\log TTR$ izeneko indizea proposatu da. Indize hau kalkulatzeko da logaritmoak erabilita:

$$\log TTR = \frac{\log L}{\log N}$$

Indize hau da gure kasu honetan erabili duguna eta berau gure aurreko lan batean erabili genuenez gero (Gaminde et al., 2017a), hemengo emaitzak Larrabetzuko gazteekin lortutakoekin erkatu ahal izango ditugu.

3. Datuen azterketa

Datuen azterketa errazago azaltzeko atal hau hiru azpiataletan banatuta aurkezten dugu; lehen azpiatalean berben erabilera ikusiko dugu, bigarrenean informatzaileen sailkapena eta hirugarrenean aberastasun-indizeak.

3.1. Berben erabilera

Azpi-atal honetan gure informatzaileek komikien edukia narratzeko erabili dituzten berbak aztertuko ditugu. Denetara 2007 berba erabili dira eta lema desberdinak 221 izan dira. Lemen maiztasunak aztertzen baditugu ikusten dugu gehien erabiltzen dena “argazki” dela ($p= 0,0937$). Era berean, hamar gauzapenetik gora dauzkaten lemak 36 dira (lema guztien % 16,29), gainerako 185 lemek agertzeko oso probabilitate txikia daukate eta horietariko 84 behin baino ez dira erabili informatzaile guztien testuetan ($p = 0,0005$ bakoitzeko).

Lema	Kopurua	p
argazki	188	0,0937
atara	154	0,0767
aita	88	0,0438
ume	60	0,0299
ur	60	0,0299
jausi	57	0,0284
seme	56	0,0279
egon	55	0,0274
ahate	52	0,0259
joan	52	0,0259
harri	47	0,0234
jarri	46	0,0229
beno	42	0,0209
parke	42	0,0209
orduan	41	0,0204
ikusi	40	0,0199
egin	38	0,0189
gizon	35	0,0174
korrika	33	0,0164
kamera	32	0,0159
justu	29	0,0144
ez	25	0,0125
nahi izan	25	0,0125
agertu	23	0,0115
izan	22	0,0110
esan	21	0,0105
erori	16	0,0080
automatiko	15	0,0075
estropezu egin	15	0,0075
ipini	15	0,0075
laku	15	0,0075
bat-batean	14	0,0070
eman	14	0,0070
momentu	13	0,0065
atze	11	0,0055
ondo	11	0,0055

2. Taula: Testuetan gehien erabili diren lemen agerpen kopuruak eta probabilitateak.

2. taulan 36 lema horien agerpen kopuruak (kop.) eta agertzeko probabilitateak (p) ematen ditugu; euron artean agerpen kopuru osoaren % 74,84 biltzen dute².

Hemengo emaitzak Larrabetzuko gazteekin lortutakoekin (Gaminde et al., 2017) erkatzeko, informatzaile kopuruak desberdinak izan zirenez gero lehen hoge berben rankingei erreparatuko diegu. Berbekin batera herri bakoitzean lortu diren agerpen probabilitateak ere adieraziko dira (3. taula).

Bilbo		Larrabetzu	
Berba	p	Berba	p
argazki	0,0937	argazki	0,1158
atera	0,0767	atera	0,0640
aita	0,0438	joan	0,0419
ume	0,0299	egon	0,0369
ur	0,0299	seme	0,0369
jausi	0,0284	jausi	0,0345
seme	0,0279	ur	0,0320
egon	0,0274	gizon	0,0296
ahate	0,0259	orduan	0,0296
joan	0,0259	aita	0,0271
harri	0,0234	ume	0,0271
jarri	0,0229	harri	0,0246
beno	0,0209	ipini	0,0222
parke	0,0209	kamera	0,0222
orduan	0,0204	beste	0,0197
ikusi	0,0199	estropezu egin	0,0197
egin	0,0189	korrika	0,0197
gizon	0,0174	pato	0,0197
korrika	0,0164	paratu	0,0172
kamera	0,0159	parke	0,0172

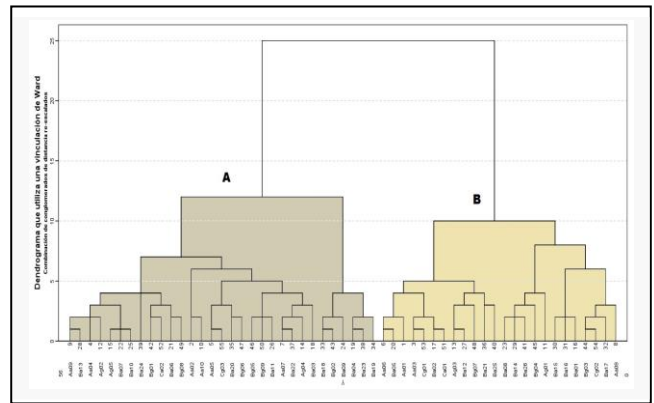
3. Taula: Bilboko eta Larrabetzuko lehen hoge berben probabilitateak.

Taula aztertzen badugu ikusten dugu rankingaren lehen hoge lemetatik 15 (%75) berberak direla herri bietan: “argazki”, “atera”, “aita”, “ume”, “ur”, “jausi”, “seme”, “egon”, “joan”, “harri”, “parke”, “orduan”, “gizon”, “korrika” eta “kamera”. Bilbon rankingaren gorengo leketan agertzen dira “ahate”, “jarri”, “beno”, “ikusi” eta “egin”; Larrabetzun, ostera, “ipini”, “beste”, “estropezu egin”, “pato” eta “paratu”.

3.2. Informatzaileen sailkapena

Agertzeko probabilitaterik altuena daukaten lemen arabera (10 gauzapan baino gehiago eduki dituzten 36 lema) informatzaileak sailkatzen baditugu, 3. irudiko dendograman erakusten diren multzoak lortzen dira. Bertan ikus daitekeen bezala, informatzaile gehienak A multzoan kokatuta agertzen dira (31, % 56,36); B multzoan 24 informatzaile agertzen dira (% 43,64).

² Emaitzak ez datoz guztiz bat Gaminderen 2010bko lanean aurkeztutakoekin, hemen aldaketa txiki batzuk erabili ditugulako lema aukeratzeko eta informatzaile kopuruak ez delako bat bera.

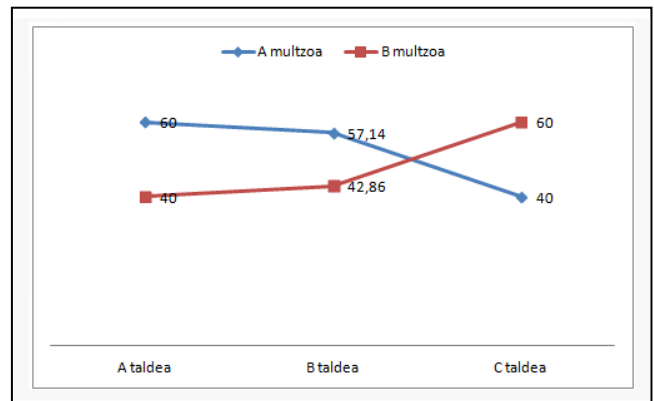


3. Irudia: Informatzaileen sailkapena gehien agertzen diren lemen erabileraren arabera.

Informatzaileak multzoetan zelan banatzen diren aztertzen badugu, euron ama hizkuntzaren arabera 4. taula osa dezakegu, bertan informatzaile kopuruak eta ehunekoak adierazten dira (4. irudiko grafikoa).

	A multzoa		B multzoa	
	Kopurua	%	Kopurua	%
A taldea	9	60	6	40
B taldea	20	57,14	15	42,86
C taldea	2	40	3	60

4. Taula: Informatzaileen banaketa multzoetan euron ama hizkuntzaren arabera.



4. Irudia: Informatzaileen banaketa multzoetan euron ama hizkuntzaren arabera.

Gauza bera generoaren arabera eginez gero, 5. taulako datuak lortzen dira. Datuok argi adierazten dute ez dagoela generoaren arabera banaketa argirik.

	A multzoa		B multzoa	
	Kopurua	%	Kopurua	%
Neskak	21	55,26	17	44,74
Mutilak	10	58,82	7	41,18

5. Taula: Informatzaileen banaketa multzoetan euron generoaren arabera.

K-batezbestekoan multzokatzearen arabera informatzaileak sailkatzen baditugu multzo nagusi bi eginik, A multzoan 29 informatzaile agertzen zaizkigu (% 52,73) eta B multzoan 26 (% 47,27).

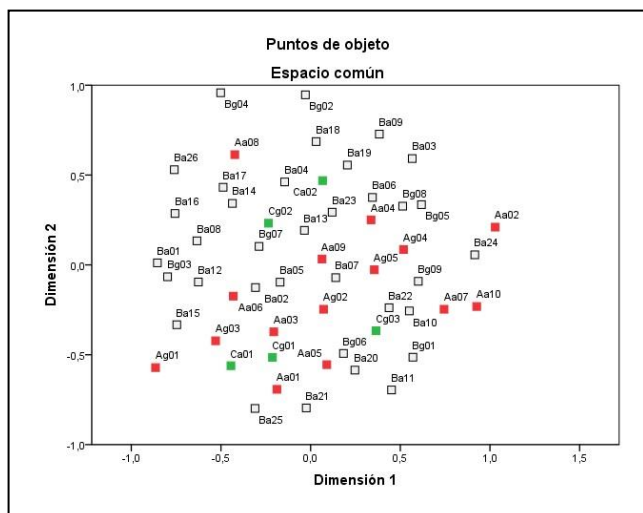
Multzokatze hierarkikoarekin egin dugun bezala, informatzaileak multzoetan zelan banatzen diren aztertuta, euron ama hizkuntzaren arabera 6. taulako emaitzak lortzen dira; aurrekoaren bidez lortutako emaitzekin erkatuta, aldeak B multzoan baino ez zaizkigu agertzen eta eurak oso txikiak dira.

Multzokatze hierarkikoarekin egin dugun bezala, informatzaileak multzoetan zelan banatzen diren aztertuta, euron ama hizkuntzaren arabera 6. taulako emaitzak lortzen dira; aurrekoaren bidez lortutako emaitzekin erkatuta, aldeak B multzoan baino ez zaizkigu agertzen eta eurak oso txikiak dira.

	A multzoa		B multzoa	
	Kopurua	%	Kopurua	%
Neskak	20	52,63	18	47,37
Mutilak	9	52,94	8	47,06

6. Taula: Informatzaileen banaketa multzoetan euron generoaren arabera.

Azkenik, dimentsio aniztun eskala izeneko estatistikoaren bidez informatzaileak euron tasun lexikoen arabera dimentsio biko lau baten proiektatu ditugu (5. irudia), bertan ikusten den bezala, informatzaile guztiak nahasirik agertzen dira eta ez dago talde homogeenorik bereizterik.



5. Irudia: Informatzaileen hedapena dimentsio aniztun eskala probaren arabera.

Gauza bera generoaren arabera egin bezala, 7. taulako datuak lortzen dira. Teknika honen bidez lortutako emaitzek erakusten dute banaketa orekatuagoa multzokatze hierarkikoaren bidez lortutakoekin baino. Beraz, ezin talde homogeenorik egin daiteke ez informatzaileen ama hizkuntzaren arabera ez eta generoaren arabera ere.

	A multzoa		B multzoa	
	Kopurua	%	Kopurua	%
A taldea	9	60,00	6	40,00
B taldea	18	51,43	17	48,57
C taldea	2	40,00	3	60,00

7. Taula: Informatzaileen banaketa multzoetan euron ama hizkuntzaren arabera.

3.3. Aberastasun indizeak

Azpiatal honetan informatzaileen aberastasun indizeen emaitzak aurkeztuko ditugu; erabili dugun aberastasun indizea metodologia atalean azaldu dugun logTTR izan da. 8. taulan informatzaile guztien lexiko aberastasun indizearen batezbestekoa, desbideratzea, minimoa maximoa eta heina erakusten dira.

N	\bar{x}	sd	min.	max.	heina
55	0,883	0,033	0,807	0,946	0,138

8. taula: Lexiko aberastasunaren indizearen batezbestekoa, desbideratzea, minimoa, maximoa eta heina.

Datuok Larrabetzuko gazteekin lortutako datuekin erkatzen baditugu (Gaminde et al., 2017a), ikusten dugu aldea oso txikia dela; Larrabetzuko gazteen lexiko aberastasunaren indizeen batezbestekoa 0,878 da, desbideratzea 0,036 eta N = 32.

Datuok informatzaileen ama hizkuntzaren arabera aztertzen baditugu, 9. taulan agertzen diren batezbestekoak eta desbideratzeak lortzen dira. Bertan ikusten den moduan, C taldeak dauka indizearik altuena gero B taldeak eta baxuena, berriz, A taldeak dauka; aldeak ez dira estatistikoki esanguratsuak. 6. irudiko grafikoa Bilboko gazteen indizeen batezbestekoak eta Larrabetzuko gazteenak batera ematen ditugu.

AH.	N	\bar{x}	sd
A	4	0,878	0,031
B	3	0,882	0,033
C	4	0,898	0,039

9. taula: Lexiko aberastasunaren indizearen batezbestekoak eta desbideratzeak informatzaileen ama hizkuntzaren arabera.

6. Irudia: Lexiko aberastasunaren indizearen batezbestekoak eta desbideratzeak informatzaileen ama hizkuntzaren arabera.

Datuak informatzaileen generoaren arabera aztertuta 10. taulako emaitzak lortzen dira. Bertan erakusten den bezala, mutilen indizea altuxeagoa da neskena baino; aldea ez da estatistikoki esanguratsua. Bilboko gazteen aberastasun indizeen batezbestekoak generoaren arabera erakusten dira 7. irudiko grafikoa Larrabetzuko datuekin batera erkatu ahal izateko.

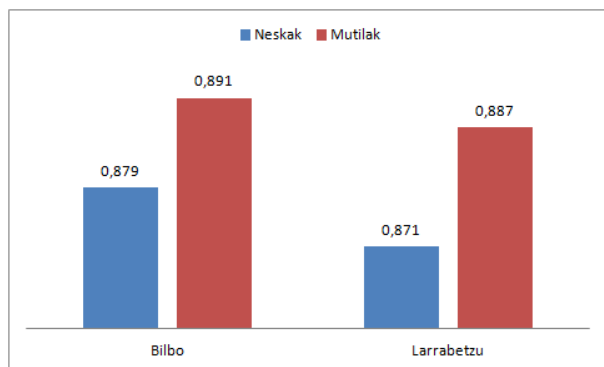
Generoa	N	\bar{x}	sd
Neskak	6	0,879	0,034
Mutilak	5	0,891	0,031

10. taula: Lexiko aberastasunaren indizearen batezbestekoak eta desbideratzeak informatzaileen generoaren arabera..

7. Irudia: Lexiko aberastasunaren indizearen batezbestekoak eta desbideratzeak informatzaileen generoaren arabera.

Beraz, esan dezakegu, aztertu dugun bat-bateko testu honen lexiko aberastasunari dagokionez, ez dagoela alde handirik ez informatzaileen ama hizkuntzaren arabera ez eta generoaren arabera ere.

4. Ondorioak



Azken azpiatal honetan datuen azterketatik atera daitezkeen ondorio nagusiak laburbilduko ditugu eta aldi berean aurrera egiteko iradokizun batzuk ere plazaratuko ditugu.

Hemen egin dugun berben erabileraren gaineko azterketa eta Gaminderen (2010b) lanean argitaratutakoa bat datozela ikusi dugu; bestetik, Larrabetzuko gazteen datuekin alderatuz gero, maiztasunik handienetako berbak bat datoz % 75ean, rankingak antzekoak izanik. Beraz, testu honen lexikoari dagokionez, herriaren tamainak eta hiztunen jatorriak ez daukate garrantzi handirik.

Informatzaileen sailkapena egiten dugunean, ez da agiri inongo azpi multzo nabarmenik, ez informatzaileen ama hizkuntzaren arabera ez eta euron generoaren arabera ere. Haatik, informatzaileak oso nahasirik agertzen dira bai multzokatze hierarkikoa, bai K-batezbestekoak eta bai dimentsio anizkun eskala erabiltzean. Larrabetzuko gazteak multzokatzean gauza bera agertu zitzaigun.

Lexiko aberastasunaren indizeen arabera egindako azterketek ez dute alde nabarmenik adierazi; agertu izan zaizkigun alde txikiak ez dira estatistikoki esanguratsuak izan ez informatzaileen ama hizkuntzaren arabera ez eta euron generoaren arabera ere.

Gaminderen (2010b) lanean gramatika arloko arauak eta ezaugarriak erkatzean bai agertu ziren alde nabarmenak informatzaileen ama hizkuntzaren arabera. Bertan adierazten zen bezala, A taldeko informatzaileek azpi multzo bat osatzen zuten eta B eta C taldeetakoek beste bat. Ondorioz esan dezakegu lexikoa eta gramatika bide bitatik ari direla bilakatzen eta desberdin jokatu behar dela euron azterbideetan edo gurago bada, bataren eta bestearen emaitzak ez direla guztiz parekarriak.

Aurrerantzean esan berri dugun hau baiestekotzat edo ezestekotzat azterketa metodologia berberarekin azterketa heda daiteke Euskal Herri osora, eta osora diogunean esan nahi dugu Lanestosatik Tuterara eta handik Baionara behintzat. Hori egiteko materialak baditugu, izan ere, 500 bat komikiren grabazioak lortuak dauzkagu hizkuntzaren eremu osoan, arazo bakarra aztertzeko astia da. Geroak esango ahal digu azterketa horrek zer emango digun!

5. Aipamenak

- Aurrekoetxea, G. (1992) Nafarroako euskara: azterketa dialektometrikoa, *Uztaro* 5, 59-109.
- Aurrekoetxea, G. (1995) *Bizkaieraren egituraketa geolinguistikoa*, UPV/EHU-ren Argitalpen Zerbitzua, Leioa.
- Aurrekoetxea, G. (2003) "Euskalkiak estandarren uholdepean, Arratiako kasua" in *Mendebalde Kultura Alkartea: Ahozketasuna aztergai*.
- Aurrekoetxea, G. (2004) "Estandar eta dialektoen arteko bateratze-joerak, ikuspuntu teorikotik begirada bat", *Uztaro* 50.
- Aurrekoetxea, G. (2005) "Nafarroako Euskararen Sailkapenaz" in Etxebarria, P. eta Knör, H. (2005) *Nerekin yaio nun. Txillardegiri omenaldia*. Euskaltzaindia eta EHU, Iker 17, Bilbo. (109-124)
- Aurrekoetxea, G. (2008) "Bariazio soziolinguistikoa Dimako euskaran", *Euskalingua* 12.
- Aurrekoetxea, G. (2009). Bariazio sinkronikoa aztertzeko metodologia(k). *Lapurdum XIII*, 43-59.
- Aurrekoetxea, G. (2016). Distantzia geografikoaren eta hizkuntza distantziaren arteko korrelazioa. in Aurrekoetxea, G., Makazaga, J. M. eta Salaberri, P. (arg.). *Hire Bordatxoan: Txipi Ormaetxea Omenduz*. UPV/EHU. Bilbo

- Aurrekoetxea, G. eta Ormaetxea, J. L. (2006) "Euskararen Atlas Sozio-Geolinguistikoa" ikerketa proiektua", *Euskalingua* 9.
- Aurrekoetxea, G. eta Ormaetxea J. L. (eds.) (2010) *Tools for Linguistic Variation*, Bilbao: UPV-EHU, *Supplements of the Anuario de Filología Vasca "Julio de Urquijo"*, LIII.
- Aurrekoetxea, G.; Gandarias, Leire, Gaminde, Iñaki eta Iglesias, A. (2014a) Variación prosódica en vasco: áreas acentuales. in Yolanda Congosto Martín, María Luisa Montero Curiel y Antonio Salvador Plans (eds.) *Fonética Experimental, Educación Superior e Investigación*, Arco Libros, Madrid.
- Aurrekoetxea, G.; Gaminde, I.; Gandarias, L. eta Iglesias, A. (2014b) Prosodic variation in the Basque language: intonational areas. in Díaz-Negrillo, A. & Díaz Pérez Francisco Javier (eds) *Specialisation and Variation in Language Corpora*, Peter Lang AG, International Academic Publishers, Berna.
- Ayuga, E. (d.g.) "Análisis de Conglomerados". http://ocw.upm.es/estadistica-e-investigacion-operativa/matematicas-y-estadistica-aplicada/contenidos/OCW/Anal_Multivar/Mat_Clase/anal_mult_2.pdf
- Boersma, P. eta Weenink, D. (2016) Praat: doing phonetics by computer [Computer program]. <<http://www.praat.org>>
- Clua, E. (2010) Relevancia del análisis lingüístico en el tratamiento cuantitativo de la variación dialectal. in Aurrekoetxea, G. eta Ormaetxea, J. L. (2010) *Tools for Linguistic Variation*. UPV/EHU, Bilbo.
- Clua, E.; Iglesias, A.; Usobiaga, I. eta Salicrú, M. (2017) Nola identifikatu ezaugarri esanguratsuenak bariazio dialektalean: Dia Tech nabarmentzeko aukera in Iglesias, A. eta Ensunza, A. (2017) *Gotzon Aurrekoetxea lagunarterik hara*. UPV/EHU. Bilbo.
- Eguskiza, N.; De Pablo, I. eta Gaminde, I (2017) "Hizkuntz hizkeren sailkapena metodo kuantitatiboen bidez" agertzeko Psikodidaktika kongresuan
- Ensunza, A. (2015) Busturialdeko euskararen hizkuntza-aldakortasuna denboran eta espazioan. Doktorego tesia; UPV/EHU
- Etxebarria, A.; Gaminde, I., Olalde, A. eta Gaminde, U. (2016) "Hizkuntza aldakortasuna Larrabetzuko aditz morfologian" in Iglesias, A.; Romero, A. eta Ensunza, A. *Linguistic variation in the Basque language and education – II / Euskararen bariazioa eta bariazioaren irakaskuntza – II*. UPV/EHU, Bilbo.
- Gaminde, I. (2002) Bizkaiko Euskararen Ezaugarri Fonologiko Batzuen Inguruan. *Euskalingua* 1
- Gaminde, I. (2007). Bizkaian Zehar: Euskararen Ikuspegi Orokorra. Mendebalde Kultura Alkartea eta Bizkaiko Foru Aldundia: Bilbo
- Gaminde, I. (2010a) Bizkaiko Gazteen Prosodiaz: Euskaraz eta Gaztelaniaz. Mendebalde Kultura Alkartea eta Bizkaiko Foru Aldundia. Bilbo.
- Gaminde, I. (2010b): "Bilboko gazte euskaldunen ezaugarri linguistikoez" in *Mendebalde Kultura Alkartea: Bilbon Mundua Ikusi*. Bilbo. (35-59)
- Gaminde, I.; Romero, A.; Legarra, H. (2012) Gramatika eta hizkuntza bariazioa Bermeon, Bermeoko Udala: Bermeo.
- Gaminde, I.; Romero, A.; Etxebarria, A. eta Eguskiza, N. (2016a) "Bizkaiko aditz laguntzaileen bilakaeraren azterketaz" in Iglesias, A.; Romero, A. eta Ensunza, A. *Linguistic variation in the Basque language and education – II / Euskararen bariazioa eta bariazioaren irakaskuntza – II*. UPV/EHU, Bilbo.
- Gaminde, I., Etxebarria, A., Eguskiza, N., Romero, A. eta Unamuno, L. (2016b) "Lexikoaren bariazioa eta multzokatze-azterketa" *Euskalingua*-28, 19-3
- Gaminde, I.; Olalde, A.; Etxebarria, A.; Eguskiza, N. eta Gaminde, U. (2017a): Hizkuntza Aldakortasuna Larrabetzun. Larrabetzuko Udala. Bilbo.
- Gaminde, I.; Eguskiza, N.; Romero, A. eta Etxebarria, A. (2017b) Informatzaileen sailkapenerako arazo metodologiko batzuen gainean (Agertzeko in III. jardunaldiak Bariazioaz eta hezkuntzaz.)
- Goebel, H. (1992) "Problèmes et méthodes de la dialectométrie actuelle (avec application à l' AIS)", in Euskaltzaindia/Académie de la Langue Basque éd., *Nazioarteko dialektologia biltzarra. Agiriak/Actes du Congrès International de Dialectologie (Bilbo/Bilbao 1991)*, Bilbo/Bilbao, 429-475.
- Goebel, H. (2010) Introducción a los problemas y métodos según los principios de la Escuela Dialectométrica de Salzburgo. in Aurrekoetxea, G. eta Ormaetxea, J. L. (2010) *Tools for Linguistic Variation*. UPV/EHU, Bilbo.
- Iglesias, A. (2014) Igorreko hizkeraren azterketa dialektologikoa, Doktorego tesia, UPV/EHU
- Léonard, J. L., Heinsalu, E., Patriarca, M. eta Darlu, P. (2015) "Modeling regional variation from eas: complexity and communal aggregates" in Aurrekoetxea, G.; Romero, A. eta Etxebarria, A. *Linguistic Variation in the Basque an Education-I/Euskararen Bariazioa eta Bariazioaren Irakaskuntza-I* (50-58or.) Bilbo. UPV/EHU
- Lujanbio, O. (2016). Hizkuntza-aldakortasuna euskararen Nafarroa ipar-mendebaldeko bi udalerritan egindako azterketa. [Doktorego tesia]. UPV/EHU.
- Pardo, A. eta Ruiz, M. A., (2002) SPSS 11 Guía para el análisis de datos. McGraw-Hill, Madril.
- Pérez López, C. (2005) Métodos estadísticos avanzados con SPSS. Thomson, Madril.
- Roberto, J. A.; Martí, A. eta Salamó, M. (2012) "Análisis de la riqueza léxica en el contexto de la clasificación de atributos demográficos latentes" *Procesamiento de Lenguaje Natural*, 48, 97-104
- Terrádez, M. (d.g.) Análisis de Conglomerados. Universitat Oberta de Catalunya, <http://www.uoc.edu/in3/emath/docs/Cluster.pdf>