



Lapurdum

Euskal ikerketen aldizkaria | Revue d'études basques |
Revista de estudios vascos | Basque studies review

19 | 2016
Numéro XIX

Fonema kontsonantikoen banaketa estatistikoa euskal hitzetan zehar

Aritz Irurtzun



Édition électronique

URL : <https://journals.openedition.org/lapurdum/3438>

DOI : 10.4000/lapurdum.3438

ISSN : 1965-0655

Éditeur

IKER

Édition imprimée

Date de publication : 1 janvier 2016

Pagination : 341-355

ISBN : 978-2-95534-134-6

ISSN : 1273-3830

Référence électronique

Aritz Irurtzun, «Fonema kontsonantikoen banaketa estatistikoa euskal hitzetan zehar», *Lapurdum* [Linean], 19 | 2016, Sarean emana-----an 01 janvier 2021, kontsultatu 03 septembre 2021. URL: <http://journals.openedition.org/lapurdum/3438> ; DOI: <https://doi.org/10.4000/lapurdum.3438>

Fonema kontsonantikoen banaketa estatistikoa euskal hitzetan zehar

Aritz IRURTZUN
CNRS-IKER

1. Sarrera¹

Lan honetan Xarlesek horrenbeste maite dituen euskal hitzen azterketa bat eskainiko da, baina bulegoan beti modu irriгарrian kontaktzen dizkidan historia eta istorioetan ez bezala, nik hitzen mami eta erabilerari beharrean haien formari egingo diet so; beraz, idorragoa izango da nire ekarpena. Idorra eta umila, ondoren datorrena euskal hitzetan aurki daitezkeen murriztapen fonotaktikoetarako eta fonemen distribuziorako lehen hurbilpen enpiriko bat besterik ez baita, ondoko lanetan halabeharrez findu eta garatu beharko dena.

2. Fonemen banaketa hitzean zehar

Hitzak aztertzeke lehen arazo nagusia corpusaren osaera da. Euskaraz, badira azken urteotan garatu diren hainbat corpus erabilgarri; horien artean UPV/EHUren Ereduzko Prosa Gaur², Egungo Testuen Corpora³, Euskal Klasikoen Corpora⁴, edota Elhuyarren Web-Corpusak⁵ ditugu. Horiek izugarri corpus handiak dira milioika hitzez osatuak.

-
1. Lan hau garatzeko D. Taylorren Corpus of Historical American English-en ikerketan oinarritu naiz, hark egindako grafemen analisia (ikus https://github.com/Proofreader/word_list_tools) euskarazko fonemen analisirako moldatuz. Lan hau ondoko diru-laguntzei esker burutu da: FFI2013-43823-P eta FFI2012-38064-C02-01 (MINECO), AthEME 613465 (Europar Komisia), IT769-13 (Eusko Jaurlaritzza). Eskerrik asko A. Egurtzegi, M. Martínez-Areta eta J. Salaberriari haien iruzkinengatik.
 2. URL: <http://webcorpusak.elhuyar.org/>
 3. URL: <http://www.ehu.eus/etc/>
 4. URL: <http://www.ehu.eus/ehg/kc/>
 5. URL: <http://webcorpusak.elhuyar.org/>

Nik hala ere, azterketarako Astigarraga-eta (2009) laneko Xuxen hiztegiko corpusa erabili dut, izan ere, alde batetik nire helburuak betetzeko hitz aski badituelako (80000 hitzetik gorakoa da), baina beste aldetik ere testu-masa handitan oinarritutako corpusek baino euskal hitzen errepresentazio zehatzagoa ematen duela esan liteke, izan ere corpus honek ez ditu web-corpusetan oinarritutako corpus berriek dituzten flexio-aldagai guztiak hitz desberdintzat hartzen (“aita”, “aitak”, “aitaren”, “aitari”...), eta mailegu berriak barne hartzen baditu ere, bestelako corpusak baino “kontserbadoreagoa” da. Esan gabe doa, idealena agian Euskaltzaindiaren Orotariko Euskal Hiztegiko sarrerak aztertzea izango zen, hori baita euskal hitzen erreferentzia nagusia, baina tamalez, eskuragarri dagoen dokumentuan hitzak berak ez daude analisirako egokia izan daitekeen moduren batean.

Bigarren arazo nagusia, corpuseko testuen (hitz idatzien) transformazio fonologikoa egitea da, hartarako euskararen arau fonotaktikoak kontuan izan ditut, jarraian agertuko diren zehaztapenak eginik.

Azkenik, fonema desberdinek hitzetan zehar duten banaketa aztertzerakoan arazo nagusia hitzen luze-labur desberdintasuna da; hitzen hasierako edota bukaerako fonemen batzaz bestekoak ateratzea eta hauek konparatzea sinplea da, baina gainontzekoan nola konparatu hiru fonemaz osatutako hitz bat (demagun “uda” /uda/) eta 5 fonemaz osatutako beste bat (demagun, “udare” /udare/)? /u/ bietan da lehen elementua, baina “uda”-n askoz garrantzitsuagoa da “udare”-n baino, lehenean 1/3 baita baina bigarreanean 1/5 besterik ez. Bestalde “uda”-ko /a/ hirugarren posizioan izateaz gain azken posizioan ere bada, aitzitik “udare”-ko /a/ hirugarren posizioan dago, baina ez da azken ebakina, erdigunekoa baizik.

Honela bada, datu erdi-kuantitatiboen analisirako hitzen deziletako (hamarrekotako) banaketa egin dut, fonema bakoitzak hitzaren zein ehuneko hartzen duen zehaztuz. Badira, jakina, analisi sinplea duten hitzak, hala nola “honenbeste”, non 10 fonema ditugun (<h> grafema /h/-tzat aintzat hartuta, Iparraldeko barnealdean bezala) eta haietako bakoitzak hitzaren % 10 hartzen duen:

(1)

Fonemak	h	o	n	e	n	b	e	ʃ	t	e
Posizioa	1	2	3	4	5	6	7	8	9	10
Ehunekoa	% 10	% 10	% 10	% 10	% 10	% 10	% 10	% 10	% 10	% 10

Badira, halaber, fonema gutxiago edota gehiagoz osatutako hitzak, hala nola “bide” (2) edo “dialektikotasun” (3):

(2)

Fonemak	b	i	d	e
Posizioa	1	2	3	4
Ehunekoa	% 25	% 25	% 25	% 25

(3)

Fonemak	d	i	a	l	e	k	t	i	k	o	t	a	ʃ	u	n
Posizioa	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Ehunekoak	% 6,66	% 6,66	% 6,66	% 6,66	% 6,66	% 6,66	% 6,66	% 6,66	% 6,66	% 6,66	% 6,66	% 6,66	% 6,66	% 6,66	% 6,66

Honela bada, hitzen konparaziorako banaketa xehe hori hamar deziletara ekarri dut, fonema bakoitzak zein posizio eta zein heinetan okupatzen duen zehaztuz.

(4)

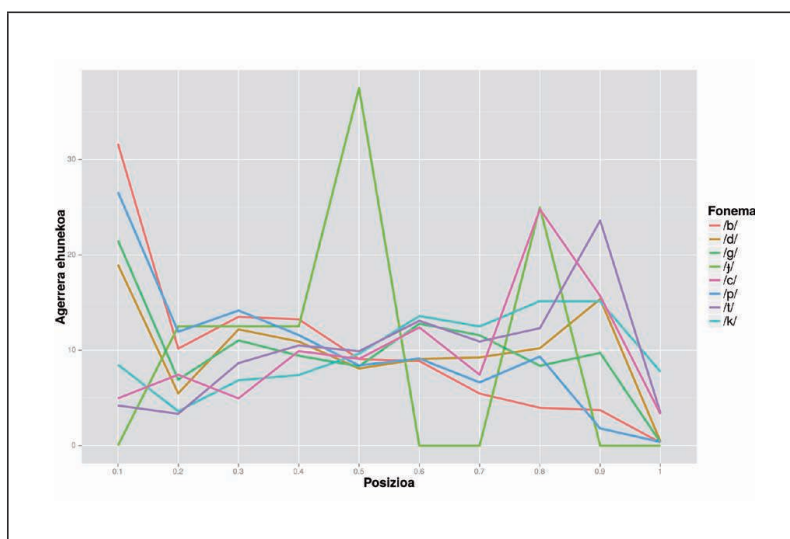
Fonemak	d	i	a	l	e	k	t	i	k	o	t	a	ʃ	u	n
Posizioa	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Ehunekoak	% 6,66	% 6,66	% 6,66	% 6,66	% 6,66	% 6,66	% 6,66	% 6,66	% 6,66	% 6,66	% 6,66	% 6,66	% 6,66	% 6,66	% 6,66
Dezilak	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15

Ondoko orrietan ematen dut haien distribuzioaren berri. Oraingo honetan analisia kontsonanteetara mugatuko dut, bokal eta diptongoen azterketa etorkizunerako utziz⁶.

2.1. Leherkariak:

Leherkarien sailean bi multzo garden desberdin daitezke sabaikari eta ez-sabaikarien artean; azken hauek oso balio baxuak erakusten dituztelako hitz osoan zehar; /j/-ren 8 agerpen besterik ez ditugu (hainbat posiziotan agerpen bakar bat ere gabe eta hitzaren erdialdera pilatuta); aldiz, /c/ oparago agertzen zaigu (orotara 121 agerpen) eta hitzaren posizio guztietan banaketa parekoxeagoa du, 8. eta 9. deziletan balio altuenak erakutsita.

6. Analisia PERL, LibreOffice Calc (SansDec'-en script bat), R eta ggplot2-rekin egin da.



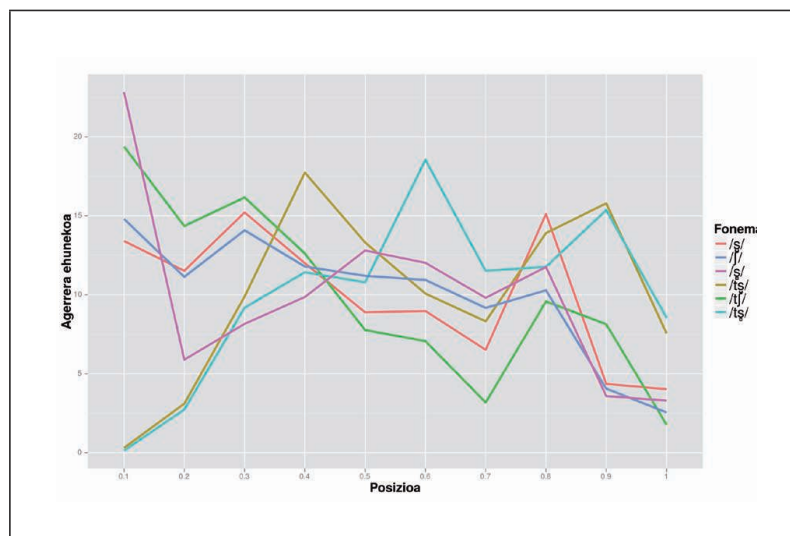
Leherkari ez sabaikarrituen artean agerpen gehien dituen /k/ da, guztira >> 48.682 agerpenekin, jarraian /t/ >> 38.615 agerpenekin, /g/ >> 20.880, /d/ >> 20.047 eta /b/ >> 14.962, agerpen gutxien dituen /p/ izanik >> 10.104. Fonemok hitzean zehar duten distribuzioa ondoko taulan zehazten da agerrera kopuruak dezilka emanaz. Ondoko irudian fonema bakoitzaren agerrera-banaketen ehunekoak grafikoki adierazten dira, hots, fonema bakoitzaren agerrera guztiak kontutan izanda, haien ehuneko zenbat diren lehen dezilean (0,1), bigarrean (0,2), eta abar:

Posizioa	/j/	/c/	/b/	/d/	/g/	/k/	/p/	/t/
0,1	0	6	4737	3806	4493	4135	2687	1624
0,2	1	9	1522	1098	1444	1754	1206	1288
0,3	1	6	2022	2444	2305	3331	1432	3342
0,4	1	12	1982	2188	1968	3600	1171	4057
0,5	3	11	1358	1622	1737	4687	850	3823
0,6	0	15	1327	1819	2668	6609	923	5056
0,7	0	9	817	1856	2417	6072	670	4217
0,8	2	30	593	2047	1749	7368	944	4756
0,9	0	19	557	3081	2030	7359	182	9115
1	0	4	47	86	69	3767	39	1337
Guztira	8	121	14962	20047	20880	48682	10104	38615

Esan bezala, grafikoan fonema bakoitzaren agerreren distribuzioa islatzen da, ez agerrera-kopurua bera. Honela bada, /j/-k bosgarren dezilean duen gailurra 3 agerpeni besterik ez zaio zor (agerrera guztien (8) % 37,5). Grafikoetan agerreren zenbatekoak

beharrean agerrera-kopuruak ilustratu izan bagenitu agerpen handiko fonemon distribuzioak agerpen urrikoen balioak zapalegi egingo zituzketen, grafika lauak uzteraino. Horregatik, datu gordinak tauletan eman dira, eta erlatiboak grafikoetan. Areago, teknika hau erabilia, euskal leherkariaren distribuzioan interesgarria den behaketa bat egin dezakegu. Ezagun da euskal leherkariak banaketa fonotaktiko osagarria izan dutela hitzaren hasiera eta amaierari dagokionez: historikoki leherkari ahostunak besterik ez dira izan hitz hasieran, eta leherkari ahoskabeak hitz amaieran (ik. besteak beste, Martinet (1955), Mitxelena (1977), Lakarra (1995, 2011), Martínez-Areta (2008)). Hemen egindako azterketak behaketa hori indartzen du, leherkari ahostunek askoz agerpen gehiago dituzte lehen dezilean hamargarrenean baino (adibidez /b/-ren agerpenen % 31,66 lehen dezilean dira, hamargarren dezilean % 0,31 besterik ez izanik, eta antzekoa gertatzen da /g/-rekin (% 21,51 vs. % 0,33) zein /d/-rekin (% 18,98 vs. % 0,42)). Leherkari ahoskabeek, aldiz, bestelako portaera dute; /k/-k antzeko balioak ditu hitzaren hasiera zein amaieran (% 8,49 vs. % 7,73), eta halaber egiten du /t/-k (% 4,20 vs. % 3,46), aldentzen den bakarra /p/ izanik (% 26,59 vs. % 0,38) (gogoan izan, hala ere, deziletako agerpenenez ari gara, ez lehen/azken posizio absolutuez). Areago, grafikoak ere argi uzten du ahostun-ahoskabe bikoteetarako esperoko genituzkeen distribuzioak agertzen zaizkigula: /b/-k /p/-k baino balio altuagoak ditu lehen dezilean (bai agerpen kopuru absolutuan (/b/ = 4737, /p/ = 2687), bai eta fonema bakoitzaren agerpenen ehunekoari dagokionez (/b/ = % 31,66, /p/ = % 26,59)) eta aurkakoa gertatzen da hitzaren azken dezilean (/b/ = 47, % 0,31 vs. /p/ = 39, % 0,38). Eta berdina gertatzen da beste leherkari bikoteekin: lehen dezilean /d/-k (3806, % 18,98) /t/-k (1624, % 4,2) baino balio altuagoak ditu, baina hitz amaierako dezilean alderantziz /d/-k (86, % 0,42) /t/-k (1337, % 3,46) baino balio baxuagoak ditu. /g/-k/ /k/-k bikoteak ere distribuzio berbera erakusten du: hitz hasierako dezilean /g/-k (4493, % 21,51) /k/-k (4135, % 8,49) balio altuagoak ditu bai kopuruan eta bai agerpenen ehunekoan, baina hitzaren amaieran alderantzizko egoera dugu /g/-k (69, % 0,33) /k/-k (3767, % 7,73) baino askoz balio apalagoak izanik.

2.2. Txistukariak



Txistukarien sailean leherkarien sailean ikusi dugun antzeko egoera agertzen zaigu: oro har, euskaraz txistukari frikariak agertu ohi dira hitz hasieran eta afrikatuak hitz amaieran. Salbuespentzat >> /tʃ/ jarri ohi da, hitzaren hasieran ohikoagoa izanik hitz amaieran baino (ik. besteak beste, Mitxelena (1977))⁷. Gure datuek ere behaketa hau babesten dute:

Posizioa	/s̺/	/ʃ/	/s̺/	/tʃ/	/tʃ/	/tʃ/
0,1	2493	226	6777	10	524	12
0,2	2142	170	1750	99	388	238
0,3	2829	215	2423	315	437	799
0,4	2232	180	2928	565	341	994
0,5	1653	171	3803	424	210	939
0,6	1669	167	3571	321	191	1613
0,7	1212	140	2913	265	86	1002
0,8	2811	157	3492	443	259	1024
0,9	811	62	1066	503	220	1336
1	748	39	980	241	48	744
Guztira	18600	1527	29703	3186	2704	8701

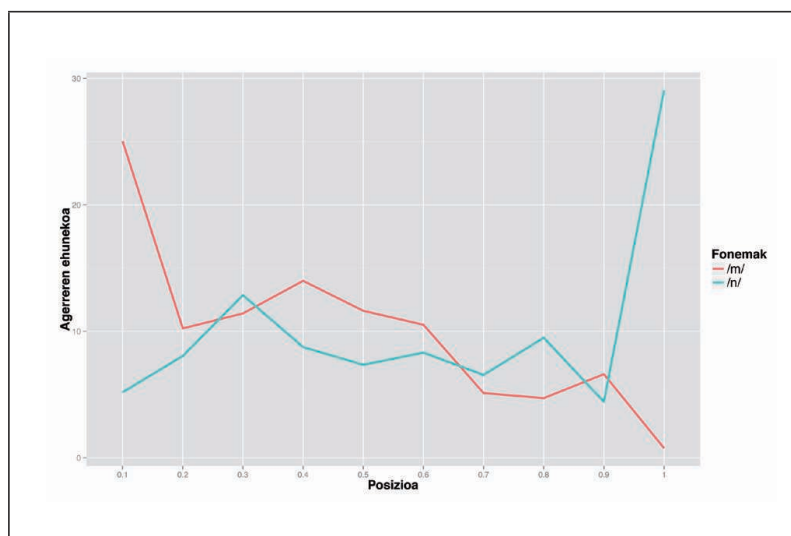
Ikus daitekeen bezala, bi multzo garden ageri dira:

- (i) txistukari frikariak eta /tʃ/-k, linealki ez bada ere, gainbeherako joera dute grafikan (eta nabarmenki, balio altuagoak dituzte lehen dezilean azken dezilean baino: /s̺/ = 2493, % 13,40 vs. 748, % 4,02, /ʃ/ = 226, % 14,8 vs. 39, % 2,55, /s̺/ = 6777, %22,81 vs. 980, % 3,29 eta /tʃ/ = 524, % 19,37 vs. 48, % 1,77)⁸.
- (ii) /tʃ/-k eta /tʃ/-k, aldiz, kontrako joera erakusten dute, hitzaren hasierako balioak oso baxuak izanik (/tʃ/ = 10, % 0,31, /tʃ/ = 12, % 0,13) eta, balio altuenak hitzaren erditsutan izan ostean, hitzaren azken dezilean hasierakoan baino askoz balio altuagoak erdietsita (/tʃ/ = 241, % 7,56, /tʃ/ = 744, % 8,55).

7. Aipatu beharrekoa da ere Nafar-Lapurteraz /ʃ/ oso ohikoa dela hitz hastapenean (eskerrik asko J. Salaberriari ohar hau egiteagatik).

8. Aipatu beharrekoa da txistukarien analisisian <x> grafemaren analisi berezitu bat egin dela benetakoko /ʃ/ txistukariari dagozkionak (“ada/ʃ/ka”) grafema berarekin idatzi arren /kʃ/ sekuentziari dagozkionetatik (“fa/kʃ/”) bereiziz. Azken hauek sekuentzia bifonemiko gisara tratatu dira analisisian.

2.3. Sudurkariak



Sudurkarien sailean /m/ eta /n/ besterik ez dira aztertu, aztergai izan den corpusak ez baitu >>/p/-ren (<ñ>) agerpenik (<in> sekuentziak badira, jakina, baina esan bezala jarrera zuhurrarekin egin ditut kalkuluak, eta <in> grafema sekuentziak /in/ fonema sekuentzien gisara aztertu ditut inongo bustidurarik gabe (funtsean, Iparraldeko euskalkietan egiten den gisara).

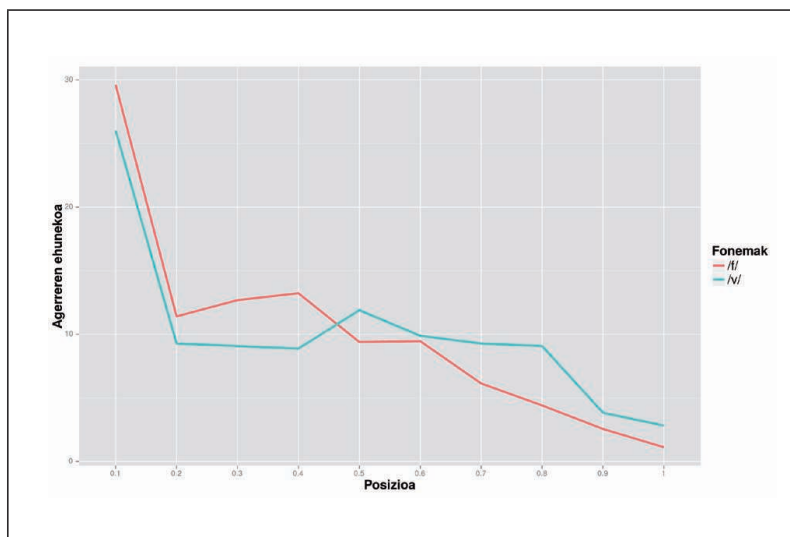
Honela bada, hauxe da sudurkari ezpainbikari eta apikarien distribuzioa hitzetan zehar:

Posizioa	/m/	/n/
0,1	2961	2450
0,2	1210	3805
0,3	1350	6079
0,4	1655	4137
0,5	1357	3476
0,6	1244	3930
0,7	605	3097
0,8	558	4488
0,9	782	2091
1	90	13733
Guztira	11830	47286

Ikus daitekeenez, hitzaren erdiko tarte nagusian antzeko bilakabideak erakusten badituzte ere, lehen eta azken deziletan alde nabarmenak daude: /m/-ren agerrera oso oparoa da hitz hasieran (2961 agerpen, % 25,02), bigarren deziletik aurrera balio apalago eta

uniformeagoak izanik hitzaren azken dezileraino, non 90 agerpen (% 0,7) besterik ez duen. /n/ aldiz balio apalekin hasten da (2450, % 5,18) eta hirugarren dezileko gailur baten ostean aski homogeneo mantentzen da % 7,47ko bataz bestekoan azken dezila iritsi arte, non izugarri handitzen den agerrera kopurua (13733), /n/-en agerrera guztien % 29,04.

2.6. Ezpainkariak



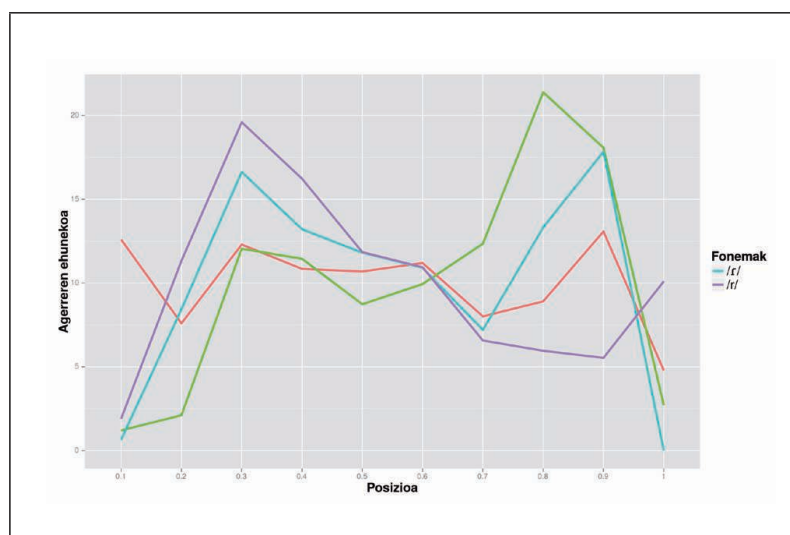
Ezpainkarien sailean /f/ eta /v/ aztertu dira. Jakina da bi fonema hauek ez direla aitzineuskararen agertzen (ik. besteren artean, Martinet (1955), Mitxelena (1977), Lakarra (1995, 2011), Hualde (1997), ikus ere Egurtzegi (2013) gainbegiratu baterako); /f/ Erdi Arotik hona gero eta gehiago agertzen den fonema da eta /v/ frantsesaren eraginez berrikitan sartutakoa dateke Iparraldeko hizkeretan, maileguetan beti ere (ik. Epelde-eta (2012)). Hemen oro har erabili den jarrera kontserbadorea alde batera utzi eta /v/-en berariazko azterketa egin da aintzat hartuta corpuseko <v>-ak /v/-ekin bat datozela. Ikusiko denez, oso balio baxuak ematen ditu hala ere, eta kontrako irizpidea hobetsiz gero aski litzateke hemen /v/-tzat jotako agerrera /b/-ren multzoarekin batzea (Hegoaldeko hizkeretan egiten den gisara). Nolanahi ere, informazio ahalik eta xeheena ematearren /v/-ren balioak ematea hobetsi dut zuzenean /b/-renekin batu ordez (esan bezala, kontrako bidea berreskuratzea sinplea da).

Hona, bada, bi fonemon distribuzioa:

Posizioa	/f/	/v/
0,1	927	129
0,2	357	46
0,3	397	45
0,4	414	44
0,5	294	59
0,6	296	49
0,7	192	46
0,8	138	45
0,9	80	19
1	35	14
Guztira	3130	496

Deigarria da bi fonemok oso pareko distribuzioa dutela, oso balio altuak hitzaren hastapenean (/f/ = 927, % 29,61 eta /v/ 129, % 26), eta jaitsiera bizkor baten ostean beheranzko mailakatuak erakusten dute bietan azken dezileko balio baxuenetara iristeraino (/f/ = 35, % 1,11 eta /v/ = 14, % 2,82).

2.7. Urkariak



Urkarien sailean bi bikote aztertu dira: /r/ eta /r/ dardarkari soil eta anizkunak, eta /l/ eta /k/ albokari apikari eta sabaikariak hurrenez hurren. Dardarkarien saila aztertzeko irizpide

fonetikoagoa erabili da fonologikoa baino coda posizioan agertzen diren dardarkari guztiak /r/-en sailean aztertuz, nahiz eta etimologikoki eta sakonekoko egituran /r/ bati egokitu. Hau egiteko arrazoiak honakoak izan dira: (i) alde batetik oso gutxi dira sakonean hitz amaieran dardarkari soila baina azalean anizkuna daukaten hitzak⁹, (ii) hitz multzo txiki batentzat erabakia arbitrarioa litzateke, bi moldeetara erabiltzen baitira¹⁰, (iii) etimologikoki /r/ bati egokituta ere hitz-barneko coda posizioan izanik [r] ahoskatzen diren hitzentzat sakoneko-egituraren analisia ez da hutsala (labur esateko, hitzun gehienontzat oharkabekoa den etimologiaz gain zein da “nortasun” hitzeko dardarkariaren sakoneko forma /r/ dela esateko arrazoa?). Ohar metodologiko hau emanik, eta 7. eta 8. oin-oharretako zehaztapenetako xehetasunekin informazio handirik galtzen ez delako usteaz, hauexek dira urkarien sailean agertzen zaizkigun kontsonanteen banaketak:

Posizioa	/r/	/r/	/l/	/k/
0,1	142	539	3267	4
0,2	1841	3202	1970	7
0,3	3630	5542	3190	40
0,4	2884	4589	2813	38
0,5	2579	3350	2773	29
0,6	2381	3094	2904	33
0,7	1573	1858	2078	41
0,8	2908	1685	2310	71
0,9	3896	1565	3392	60
1	1	2856	1241	9
Guztira	21835	28280	25938	332

Ikus daitekeenez, /r/-k, eta /k/-k antzeko banaketa erakusten dute: oso balio baxuak hitzaren lehen dezilean eta “M gisako” distribuzioa bi gailur nabarmenekin hitzaren hirugarren eta 8-9 deziletan (bigarren gailurra aurrekoa baino altuagoa izanik). Bikote naturalka begiratzen baditugu bi dardarkariak antzeko patroia erakusten dute hitzaren hasierako partean, oso balio baxuak hasieran (/r/ = 142, % 0,65 eta /r/ = 539, % 1,9), zalantzarik gabe euskarak hitzen hasierako dardarkarien aurka izan duen murriztapen fonotaktiko zurrunaren isla dena. Hala

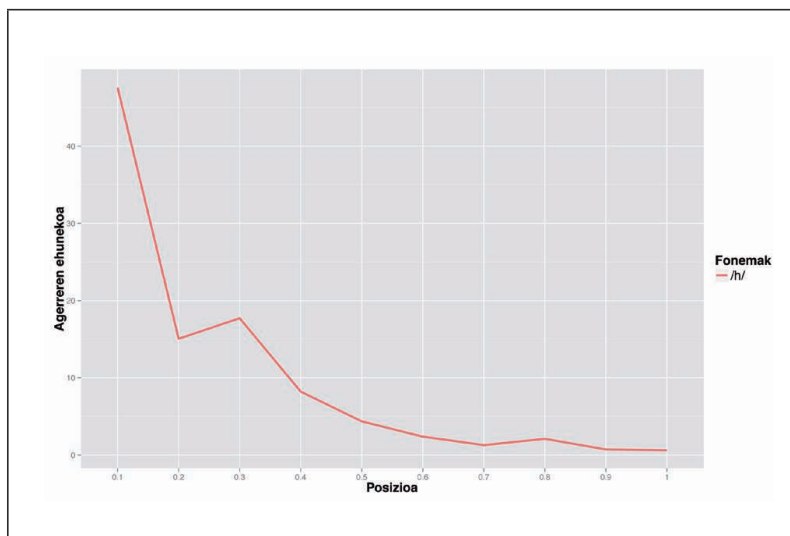
-
9. Corpus osoan dauden amaierako dardarkari guztietatik (guztira 2858) 31 besterik ez dira oinarrian /r/ dutenak, haietako asko “haur” erakuslea, “nor”, “zer” izenordainak edota beste bi lexema bereziren (“paper”, “ur”) gainean eraikiak: aloger, berrpaper, bertsopaper, edozer, gaztangazur, gazur, grefier, grossier, guhaur, hihaur, idazpaper, inor, kauter, kultur, ler, lixapaper, lizpaper, lunbar, nahinor, natur, nehor, nor, pader, paper, plater, segur, ur, xukapaper, zer, zuhaur, zur.
10. Hauek liriateke hemen sartzekoak: edonor, kaier, konfesor, ofizier, ostaler, partialier, plazer, putanir, Vladimir, Zesar.

ere, aztergai izan den corpusean, eta euskararen oro har, gero eta ohikoagoak dira dardarkari (anizkunez) hasten diren formak (“ramadan”, “rally”, “relax”...), murriztapen fonotaktiko honen zurruntasunaren ahultzea erakusten dutenak. Dardarkariekin segituz, badira beste bi fenomeno aipagarri: alde batetik, lehen aipatu bezala, /r/-k oso balio altuak lortzen ditu 8-9 deziletan. Hau agian euskarazko eratorpen-atzizki usatu batzuen osaerari (“-ari/-lari”, “-era”, “-keria”...) zor zaio. Gero /r/-k bere baliorik baxuena lortzen du azken dezilean, agerpen bakarra eta, agerpenen % 0,004arekin¹¹. Aitzitik, dardarkari anizkunak amaierako posizioan balioen igoera erakusten du, azken dezilean 2856 agerrera izanik (/r/-ren agerrera guztien % 10,09).

Albokariei dagokionez, alde nabarmenak ikusten dira distribuzioan: hasteko, /l/ oso fonema erabilia da, guztira 25938 agerrerarekin. >>/k/ ordea askoz urriagoa da, 332 aldiz besterik ez da agertzen corpusean (analisirako corpuseko <ll> grafemen azterketa egin da albokari sabaikariari dagozkionak (“artillari”) eta maileguren batetik datorren /l/ geminatua desberdinduz (“Allah”), azken hauek azterketatik kanpo utzi dira, ez baita emana soinu hauek euskaraz ahoskatzen direnik. Aurrekoetan bezala ere, ez da goiko bokalek eragindako bustidura prozesurik aintzat hartu eta <il> sekuentziak /il/ gisara interpretatu dira. /l/ eta /k/-ren konparaketarekin jarraituz, argi ikusten da /l/-ren agerrerak askoz handiagoak izateaz gain hauek hitzean zehar askoz modu homogeenagoan barreiatuak daudela (hamar deziletan zehar %10eko batuz besteko desbiazio estandarra 2,62koa besterik ez da). Aurrerago esan bezala, /k/-k aldiz askoz distribuzio malkartsuagoa du, “M-gisako” banaketa egiten duten bi gailur gardenekin. Honela bada, fonema honen banaketaren desbiazio estandarra 6,67koa da. /l/-k balio uniforme-xamarrak erakusten ditu hitzaren hasieratik (azken dezilean balio apalagoak baditu ere). /k/ aldiz oso balio baxuekin hasten da lehen dezilean (4 agerpen, agerpen guztien % 1,2), eta 3-4 eta 8-9 deziletako gailurren ostean oso balio baxuekin amaitzen du (9 agerpen, % 2,71).

11. Esan bezala, neutralizazioa aintzat hartu da analisirako. Ikus ere 7. eta 8. oin-oharrak.

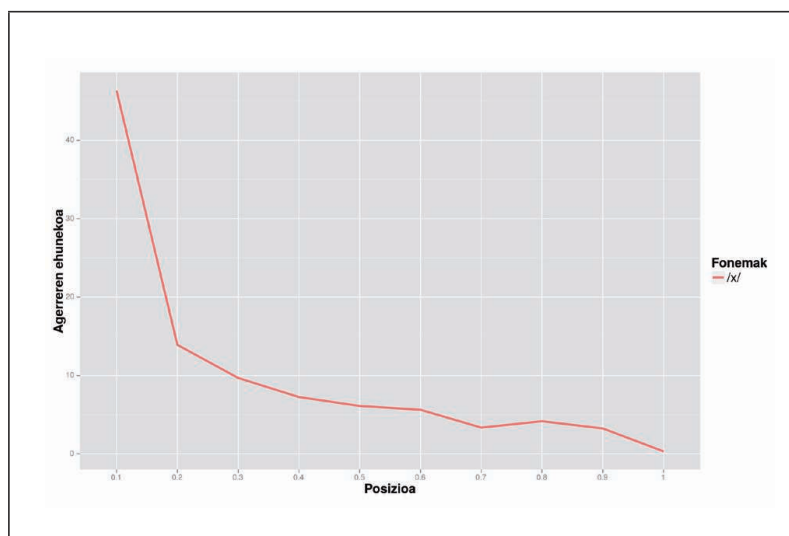
2.8 Hasperena



Ezaguna da hasperena euskarak bere-berea duen fonema izanagatik haren agerpena ez dela homoginoa hitzean zehar, izan ere Mitxelenaren lanetatik hona, ez dira gutxi izan hasperen (etimologikoa) hitzaren lehen bi silabetara mugatua dagoela esan duten lanak (*cf.*, besteak beste, Mitxelena (1957), Lakarra (2009, 2015), edota Egurtzegi (2014), eztabaidarako). Hemen egindako azterketak ere halakoxe irudi garden bat ematen du hitzean aurrera egin ahala hasperenak geroz eta urriagoak izanik (aipagarria da hasperenen % 47,58 lehen dezilean agertzen zaizkigula, eta % 80tik gora lehen hiru deziletan), honela, grafikan ikus daitekeenez, distribuzioa gainbeherakoa da erabat (eta dezilen 10eko batz bestekotikako desbiazio estandarra 14,54koa).

Posizioa	/h/
0,1	3385
0,2	1071
0,3	1260
0,4	584
0,5	310
0,6	169
0,7	91
0,8	149
0,9	51
1	44
Guztira	7114

2.9 Frikari belarra



Azkenik, frikari belarraren distribuzioa dugu (corpuseko <j>-k halakotzat hartuta). Oro har, gutxi erabilitako fonema dugu (1848 agerpen besterik ez), distribuzio normaletik oso urrun:

Posizioa	/x/
0,1	856
0,2	257
0,3	179
0,4	134
0,5	113
0,6	104
0,7	62
0,8	77
0,9	60
1	6
Guztira	1848

Alde handiak ikus daitezke lehen dezilaren eta gainontzekoen artean: oro har, ehuneko 10eko batz bestekotikako desbiazio estandarra 13,3koa da, eta hemen nabarmena da lehen dezileko balioaren altua, izan ere 2. deziletik 10. dezilerako balioen batz bestekoen desbiazio estandarra 3,99koa da, askoz aldakortasun urriagoa erakutsiz.

3. Ondorioak eta etorkizunerako langaiak

Lan honetan fonema desberdinek euskal hitzek duten distribuzioaren lehen analisi bat egin nahi izan da. Oro har fonemen agerreran alde nabarmenak ikusi dira: kontsonante erabilienak /k/, /n/ eta /t/ dira, 48682, 47286 eta 38615 agerpenekin hurrenez hurren. Beste muturrean >>/j/ genuke, 8 agerpenekin besterik ez. Banaketari dagokionez ere alde nabarmenak aurkitzen dira fonemen artean; oro har banaketa ez-homogeneo eta ez normalak ikusi ditugu, fonema desberdinen arteko alde handiekin.

Deskripzio hau eginik, etorkizuneko langai bat izango da lan honetan deskribatu diren patroiak hautemangarriak ote diren euskaldunentzat; hau da, aurkitu diren distribuzio-patroi estatistikoekin bat datozen logatomoak “naturalagotzat” jotzen diren murriztapen fonotaktikoak errespetatuta ere fonemen distribuzio-patroi orokorrekin bat ez datozen logatomoak baino. Hori, etorkizun hurbileko lanean.

Erreferentziak

- Astigarraga, Aitzol, Koldo Gojenola, Kepa Sarasola & Aitor Soroa. 2009. TAPE:Testu-analisirako PERL erremintak, Bilbo: Udako Euskal Unibertsitatea.
- Egurtzegi, Ander. 2013. Phonetics and Phonology. In M. Martínez-Areta (arg.), *Basque and Proto-Basque. Language-Internal and Typological Approaches to Linguistic Reconstruction*. 119–172. Frankfurt am Main: Peter Lang.
- Egurtzegi, Ander. 2014. *Towards a Phonetically Grounded Diachronic Phonology of Basque*. Doktorego tesia: UPV/EHU.
- Epelde, Irantzu, Bernard Oyharçabal & Jasone Salaberria. 2012. La base de données NORANTZ: vers un observatoire des parlers basques d'Aquitaine. *Lapurdum* 16. 27–67. DOI: 10.4000/lapurdum.2353
- Hualde, José Ignacio. 1997. Aitzineuskararen leherkariak. *ASJU* 31(2). 411–424.
- Hualde, José Ignacio & Jon Ortiz de Urbina. 2003. *A Grammar of Basque*. Berlin & New York: Mouton de Gruyter.
- Lakarra, Joseba A. 2015. Hiru hasperen haboro. In M.J. Ezeizabarrena & R. Gómez (arg.), *Eridenen du zerzaz kontenta: Sailkideen omenaldia Henrike Knörr irakasleari (1947-2008)*. 349–378. Bilbo: UPV/EHUren Argitalpen Zerbitzua.
- Lakarra, Joseba A. 2011. Aitzineuskararen berreraiketaz: egindakoaz eta eginkizunez. In A. Sagarna, J.A. Lakarra & P. Salaberri (arg.), *Euskaltzaindiaren XVI biltzarra. Piriniotako hizkuntzak, lehena eta oraina*. 617–694. Bilbo: Euskaltzaindia.
- Lakarra, Joseba A. 2009. *h3 > h1, *h2 > h1 eta horiei datxekien zenbait fenomenoz. *Lapurdum* 13. 211–231. DOI: 10.4000/lapurdum.2091
- Lakarra, Joseba A. 1995. Reconstructing the Pre-Proto-Basque Root. In J.I. Hualde, J.A. Lakarra & R.L. Trask (arg.), *Towards a History of the Basque Language*. 189–206. Amsterdam & Philadelphia: John Benjamins.
- Martinet, André. 1955. *Économie des changements phonétiques*, Berna: A. Francke.
- Martínez-Areta, Mikel. 2008. *El consonantismo proto-vasco*. Bilbo: UPV/EHUren Argitalpen Zerbitzua.

-
- Martínez-Areta, Mikel (arg.). 2013. *Basque and Proto-Basque. Language-Internal and Typological Approaches to Linguistic Reconstruction*. Frankfurt am Main: Peter Lang.
- Mitxelena, Koldo. 1957. À propos de l'accent basque. *Bulletin de la Société de Linguistique de Paris* 53. 204–233 [Berrargitarapena: J.L. Lakarra & I. Ruiz Arzalluz (arg.), 2011, *Obras completas VII: Fonética y fonología. Morfosintaxis. Dialectología*. 123–152. Donostia-San Sebastian & Vitoria-Gasteiz: Gipuzkoako Foru Aldundia, University of the Basque Country].
- Mitxelena, Koldo. 1977 [1961]. *Fonética histórica vasca*. Donostia-San Sebastian: Gipuzkoako Foru Aldundia. 2. argitalpena.