

# Bota bertsoa, eta guk aztertuko dugu

## Azken urteetako Bertsolari Txapelketa Nagusien analisia

---

**Agirrezabal Zabaleta, Manex**

EHUko IXA taldeko ikertzailea

**Arrieta Kortajarena, Bertol**

EHUko IXA taldeko ikertzailea

**Astigarraga Pagoaga, Aitzol**

EHUko IXA taldeko ikertzailea

**Hulden, Mans**

EHUko IXA taldeko ikertzailea

**Zein dira azken txapelketetan gehien erabili diren errimak eta oinak? Zein neurri aukeratzeko joera dute bertsolariek kartzelako ariketan, eta zein bilakaera izan du horrek txapelketaz txapelketa? Doinuen erabileran ba al da aldaketa nabarmenik? Zenbat euskal hitz ulertu behar dira bertso bat ulertzeko? Gora egin al du euskara batuaren erabilerak azken Bertsolari Txapelketa Nagusietan? Horiek eta antzeko galderak erantzuten saiatu gara artikulu honetan, Bertsozale Elkartearen eta, bereziki, Xenpelar Dokumentazio Zentroaren eskutik lortutako azken zazpi txapelketetako bertsoen corpora oinarri izanik eta hastear den Euskal Herriko Bertsolari Txapelketa Nagusia aitzakiatzat harturik.**

EHUko Informatika Fakultateko IXA taldean hizkuntza eta informatika uztartuta egiten diren lanen barnean, bertsolaritza aztertzen ere badihardugu azken urteotan. Hala, bertsotan egiten laguntzeko arbel digitala (errima- eta sinonimo-bilatzaileak eta neurri-egiaztatzailea biltzen dituen, besteak beste) aurkeztu dugu berriki Bertsozale Elkartearekin elkarlanean (aurki, mugikorretarako ere prest egongo da).

Era berean, hizkuntzaren sorkuntzaren alorrean, bertsoen sorkuntza automatikoa ere lantzen ari gara. Lehen pausoak egin baditugu ere, urrats sendoagoak egin aurretik bertsoak xehe-xehe aztertzeke saiakera egin dugu, horien azterketa sakonak gerora sorkuntza hobe ekar dezakeelakoan.

Azterketa horiek egiteko, Xenpelar Dokumentazio Zentroak bildu eta sailkatutako



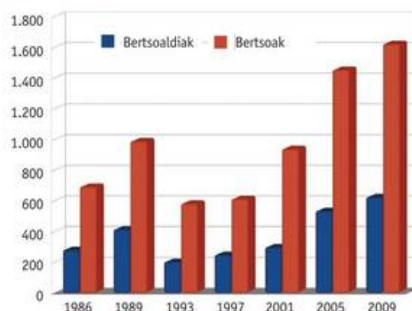
Arg. CC/www\_wkberri\_net

corpusa izan dugu oinarri. Guk erabili dugun corpusak 1986tik 2009ra arte egindako txapelketa nagusietako bertsoak hartzen ditu; corpus hori 2.600 bertsoalditan sailkatutako 6.887 bertsoak osatzen dute. 1. irudian ikus daitekeen moduan, datu-basean gordeta dauden bertsoaldiak --eta, ondorioz, bertsoak-- geroz eta gehiago dira.

Hainbat mailatan egin dugu azterketa, betiere bertsoaren ezaugarri nagusiak kontuan izanda: errimak, neurriak, doinuak, hitzak, kategoria morfosintaktikoak eta euskara batuaren erabilera.

## Errimak

Gehien erabiltzen diren errimak eta oinak zein diren aztertzeko, lerro bikoitietan soilik errima egiten duten neurriak hartu ditugu aintzat, gisa horretako bertsoekin corpusaren % 94 eskuratzeko genuelako, batetik, eta neurri irregularragoetako errimak lortu beharrak azterketa hau egiteko merezi ez zuen konplexutasuna eransten zuelako, bestetik.



1. irudia. Azterketarako erabili den datu-basean gordeta dauden bertsoak eta bertsoaldiak.

2. irudiko taulan ikus daitekeenez, txapelketaz txapelketa ez dira beti berdinak errima erabilienak, nahiz eta batzuk baliatzeko joera besteak baino handiagoa izan (esate baterako, eBGDRa errima lehen postuan agertzen da gehienetan).

Corpusa bere osoan hartuta (zazpi txapelketetako bertso guztiak kontuan hartuz) ere aztertu ditugu zein diren gehien erabili diren errimak eta oinak (3. irudian ikus daitezke datuok; oinen ezkerrean ageri den zenbakiak oin hori errima horretan zein proportziotan erabili den adierazten du; adibidez, "ela" errima erabili den kasuen % 13,27tan, "bezela" izan da hautatutako oina). Kontuan izan behar da corpusean bertso gehienak azken bi txapelketetakoak direla, eta, ondorioz, neurri hauetan pisu handiagoa izango dutela bi txapelketa horietako datuek.

	1986	1989	1993	1997	2001	2005	2009
1. eBGDRa	1000	1000	1000	1000	1000	1000	1000
2. eBa	1000	1000	1000	1000	1000	1000	1000
3. eNMA	1000	1000	1000	1000	1000	1000	1000
4. eBGDRa	1000	1000	1000	1000	1000	1000	1000
5. ePTKa	1000	1000	1000	1000	1000	1000	1000
6. eBa	1000	1000	1000	1000	1000	1000	1000
7. eNMA	1000	1000	1000	1000	1000	1000	1000
8. ePTKa	1000	1000	1000	1000	1000	1000	1000
9. eNMA	1000	1000	1000	1000	1000	1000	1000
10. eBGDRa	1000	1000	1000	1000	1000	1000	1000

2. irudia. Txapelketa bakoitzean gehien erabilitako errimak. Oharra: eBGDRa errimak eba/eda/ega/era hartzen ditu barnean, bodegero legearen arabera ontzat jotakoak; horren arabera, "arriba", "hobe da", "errega" eta "tankera" hitzek, adibidez, errimatu egiten dute. Berdintsu gertatzen da ePTKa, eNMA eta antzeko errimekin.

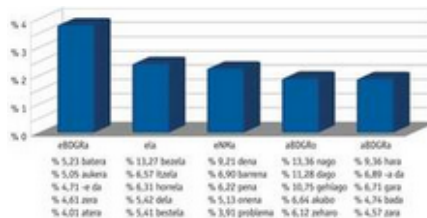
Bestalde, corpus osoan gehien errepikatzen diren hiru oinak, eta, beraz, erabilienak, "bezela", "gabe" eta "begira" hitzak dira.

## Neurriak

Neurrien azterketari dagokionez, kartzelako



ariketan gehien erabilitakoak zein diren aztertu dugu, hori baita neurri librean kantatzen den ariketa puntuagarri bakarria.

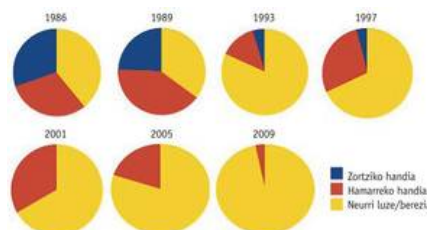


3. irudia. Txapelketa guztiak kontuan hartuta gehien erabili diren erriak.

4. irudiko grafikoan ikus daitekeen moduan, neurri luze eta berezietarako joera, espero bezala, geroz eta handiagoa da. Azpimarratzekoa deritzogu, era berean, 2001. urteko txapelketatik aurrera (corpuseko datuen arabera) ez dela zortziko handian kantatu, eta 2009. urteko txapelketan hamarreko handia ere oso gutxi erabili zela (% 3 soilik). Datu horien kariatara, badirudi etorkizuneko kartzeletan zortziko handiak eta hamarreko handiak ez dutela lekurik izango.

## Doinuak

Doinu librean kantatzen diren bertsoaldiak soilik hartu ditugu kontuan azterketa honetan; kanpoan geratu dira, horrenbestez, puntu-erantzunetan erabilitako doinuak.



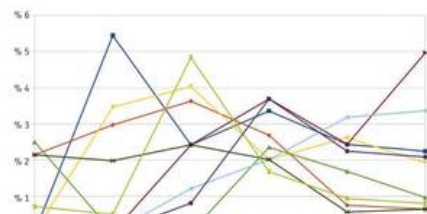
4. irudia. Kartzelako ariketan erabilitako neurriak.

5. irudian, maiz entzuten diren hamar doinuren erabileraren bilakaera ikus daiteke (ehunekotan). Aipatzekoak dira "Triste bizi naiz eta" doinu ezagunaren erabilera urria, eta "Haizea dator ifarraldetik" eta "Baserrian jaio nintzen" doinuen gorakada nabarmena. (Oharra: 1989. urteko txapelketa ez dugu kontuan hartu, corpusean agertzen diren bertsoaldien ia laurdenak ez duelako doinua dokumentatuta.)

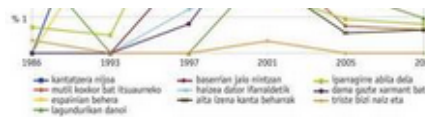
## Hitz erabilienak

Bertsoa egiteko erabiltzen diren hitzei dagokienez 6. irudiko grafikoan, lema-kopuru zehatz bat erabilia, bertsoa zer proportziotan osa daitekeen erakusten da. Bertan ikus daiteke bertsoen corpuseko 500 lema erabilienak nahiko direla bertso baten % 70 osatzeko eta 1.000 lema erabilienak bertsoaren % 80 osatzeko. Modu argiago batean esanda, euskara-ikasle batek bertso-corpus honetan gehien erabiltzen diren 500 lemak ezagutzearekin bertso baten % 70 ulertuko luke (ahozkotasunak eragindako trabak eta sintaxiaren ulergarritasun-mugak kontuan hartu gabe, noski).

Bestalde, esan beharra dago txapelketetako corpus honek Zipf-ren legea betetzen duela. Hizkuntzaren prozesamenduaren ikuspegitik, Zipf-ren legeak honako hau dio: hizkuntza naturaleko edozein corpus hartuta, maizen



agertzen den hitza  $X$  aldiz agertzen bada, maizen agertzen den hurrengo hitza  $X/2$  aldiz agertuko da, eta hurrengoa  $X/4$  aldiz, eta hurrengoa  $X/8$  aldiz...

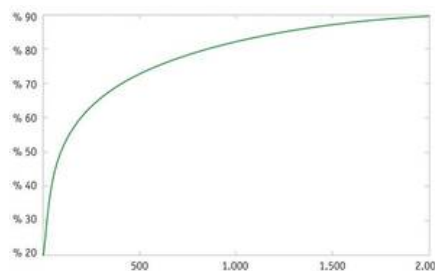


5. irudia. 10 doinuren erabileraren bilakaera.

## Kategoria morfosintaktikoak

Hitzen kategoria morfosintaktikoak ere aztertu ditugu, gehien erabiltzen direnak zein diren jakin eta urtez urte aldaketa nabarmenik izan den ala ez ikusteko.

7. irudian ikus daitekeenez, izenak eta aditzak (aditz nagusiak, laguntzaileak eta trinkoak batera zenbatuta) dira gehien erabiltzen direnak alde handiarekin. Aipagarria iruditzen zaigu, bestalde, adjektiboen erabileraren bilakaera; izan ere, behera egin du txapelketaz txapelketa, nahiz eta aldea ez den oso esanguratsua.

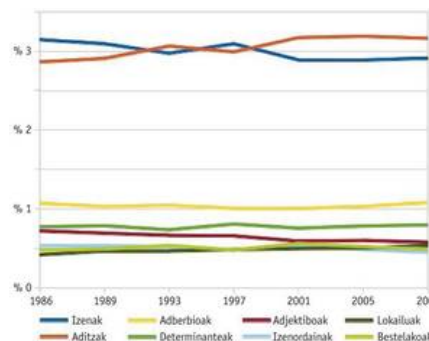


6. irudia. Lema erabilien proportzioa bertsoetan.

## Euskara batuaren erabilera

Azkenik, bertsoen corpusean euskara batuaren erabilera zenbatekoa den jakiteko, IXA taldearen lematizatzailearekin aztertu dugu corpusa, eta lematizatzaileak ezagutzen dituen hitzen bilakaerari erreparatu diogu.

8. grafikoan ikus daitekeen moduan, hitz ezagunen kopuruak gora egin du txapelketaz txapelketa. 2005eko txapelketan jo zuen goia (% 89), eta, 2009koan proportzio horrek apur bat behera egin bazuen ere, antzeko mantentzen dela ikus daiteke. IXA taldeko lematizatzaileak hitzak ez ezagutzearen arrazoiak askotarikoak izan daitezke; gure zenbatespenen arabera, ordea, euskara batua ez erabiltzea da gehien gertatzen dena (% 80ko kasuetan). Izen berezi ezezagunak (% 13), erdarakadak (% 6) edo transkribatze-akatsak (% 1) dira gainerakoak. Datuon arabera, ezin ziurta genezake hitz ezagunen gorakada euskara batua gehiago erabiltzearen ondorioz gertatzen denik (eta ez, esaterako, erdarakada gutxiago erabiltzearen poderioz), baina gure intuizioak eta eskuz aztertu dugun lagin batek joera horixe dagoelako ustea berretsi digu.



7. irudia. Bertsoetan erabilitako hitzen kategoria morfosintaktikoak.

Azken txapelketetako neurriek, gure iritziz, bi motatako aurreikuspenak iradokitzen

dituzte, nahiz eta dauzkagun datuak behar bezain zehatzak ez diren eta ondorioak ateratzeko goizegi dela iruditzen zaigun: joera hori alderantzikatu egingo da aurrerantzean, eta euskalkietako hizkera gehiago erabiltzera joko dute berriz bertsolariak; edo euskara batuaren erabileran goi-muga (% 90) horren bueltan ibiliko dira hemendik aurrera ere. Nolanahi den, gure ustez, euskara batuaren erabilerak are gehiago gorantz egitea da gaitzena, bertsolaritzaren gisako ahozko jardun batean.

## Ondorioak

Azken zazpi txapelketa nagusietako bertsoak modu estatistikoan aztertzeak zenbait joera azalertzeko aukera eman digu. Datuon azterketa pausatuagoa eta xeheagoa egiteak mereziko duen arren, estreinako honek ere utzi dizkigu kontu esanguratsu batzuk. Neurriaren aukeraketan eta euskara batuaren erabileran, esaterako, aurrez geneuzkan intuizioak egiazkoak zirela berresteko

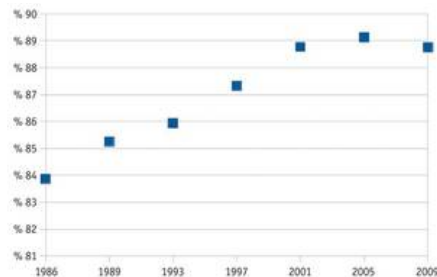
balio izan digu: neurri berezi eta luzeetarako geroz eta joera handiagoa dago, eta euskara batuaren erabileran ere badirudi gorakada ia etengabea izan dela orain arte. Doinuei dagokienez, badirudi badagoela geroz eta doinu gutxiago erabiltzeko joera, baina datuotan lausotu antzera agertu zaigu ezaugarri hori, eta ez gara ausartu horri buruzko ondoriorik plazaratzen, oraingoz.

Mantendu egingo al dira joera horiek aurtengo txapelketan, ala alderantzikatu? Eta hurrengoetan? Zer beste interpretazio interesgarri egin daitezke bertsoen corpusa oinarri gisa hartuta? Txapelketakoak ez diren bertsoaldiak aztertzeak zer ondorio azalertzeko lituzke? Eta txapelketakoak direnak txapelketakoak ez direnekin konparatzeak?

Lan handia dago egiteke alor honetan oraindik, baina bertsogintzaren azterketa xehea egiteko bertsoak egoki dokumentatzen jarraitzeak daukan garrantzia ukaezina da gure ustez, baldin eta ikusi nahi bada behintzat zer bilakaera hartzen duten hurrengo urteetan artikulu honetan aipatutako joerek eta modu pausatuago batean aztertzea merezi duten bestelakoek.

## Idatzi zuk zeuk Gai librean atalean

Gai librean aritzeko, bidali zure artikulua [aldizkaria@elhuyar.com](mailto:aldizkaria@elhuyar.com) helbidera  
Hauek dira Gai librean atalean Idazteko arauak



8. irudia. IXA taldearen lematizatzaileak ezagututako hitzen proportzioa.