

BOTA BERTSOA, ETA GUK AZTERTUKO DUGU

Azken urteetako Bertsolari Txapelketa Nagusien analisia

MANEX AGIRREZABAL ZABALETA
BERTOL ARRIETA KORTAJARENA
AITZOL ASTIGARRAGA PAGOAGA
MANS HULDEN
EHUKo IXA taldeko ikertzaileak

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 corpusa 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, bertsoan egiten laguntzeko arbel digitala (errima- eta sinonimo-bilatzaileak eta neurri-egiaztatzailea biltzen dituen, besteak

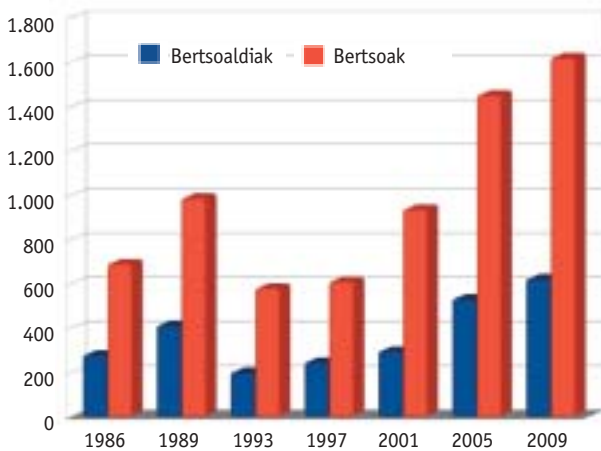
beste) aurkeztu dugu berriki Bertsozale Elkartearekin elkarlanean (aurki, mugikorretarako ere prest egongo da). Era berean, hizkuntzaren sorkuntzaren alorrean, bertsoen sorkuntza automatikoa ere lantzen ari gara. Lehen pausoak egin baditugu ere, urrats sendoagoak egin aurretik bertsoak

xehe-xehe aztertze saiakera egin dugu, horien azterketa sakonak gerora sorkuntza hobe ekar dezakeelakoan.

Azterketa horiek egiteko, Xenpelar Dokumentazio Zentroak bildu eta sailkatutako corpusa izan dugu oinarri. Guk erabili dugun corpusak 1986tik 2009ra arte egindako



ARG.: CC/WWW_WKBERRI_NET



1. irudia. Azterketarako erabili den datu-basean gordeta dauden bertsok eta bertsosaldiak.

txapelketa nagusietako bertsok hartzen ditu; corpus hori 2.600 bertsosalditan sailkatutako 6.887 bertsok osatzen dute. 1. irudian ikus daitekeen moduan, datu-basean gordeta dauden bertsosaldiak —eta, ondorioz, bertsok— geroz eta gehiago dira.

Hainbat mailatan egin dugu azterketa, betiere bertsokaren 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 bertsokekin corpusaren % 94 eskuratzan genuelako, batetik, eta neurri irregularagoetako errimak lortu beharrik azterketa hau egiteko merezi ez zuen konplexutasuna eransten zuelako, bestetik.

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

Corpusa bere osoan hartuta (zazpi txapelketetako bertsok 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 proportzioan erabili den adierazten du; adibidez,

“ela” errima erabili den kasuen % 13,27tan, “bezela” izan da hautatutako oina). Kontuan izan behar da corpusean bertsok gehienak azken bi txapelketetakoak direla, eta, ondorioz, neurri hauetan pisu handiagoa izango dutela bi txapelketa horietako datuek.

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 bakarra.

4. irudiko grafikoa ikus daitekeen moduan, neurri luze eta berezietarako joera, espero bezala, geroz eta handiagoa da. Azpimarratzeko deritzogu, era berean, 2001. urteko txapelketatik aurrera (corpuseko datuen arabera) ez dela zortziko handian kantatu, eta

2009. urteko txapelketan hamarrek handia ere oso gutxi erabili zela (% 3 soilik). Datu horien kariatara, badirudi etorkizuneko kartzelatan zortziko handiak eta hamarrek handiak ez dutela lekurik izango.

DOINUAK

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

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 bertsosaldien ia laurdenak ez duelako doinua dokumentatuta.)

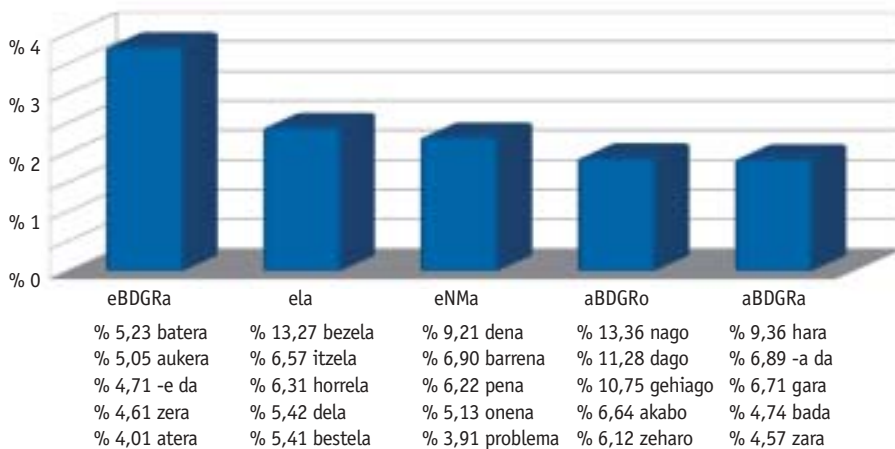
HITZ ERABILIENAK

Bertsok egiteko erabiltzen diren hitzei dagokionez 6. irudiko grafikoa, lema-kopuru zehatz bat erabilita, bertsok zer proportziotan osa daitekeen erakusten da. Bertan ikus daiteke bertsok corpuseko 500 lema erabilienak nahiko direla bertsok baten % 70 osatzeko eta 1.000 lema erabilienak bertsokaren % 80 osatzeko. Modu argiago batean esanda, euskara-ikasle batek bertsok-corpus honetan gehien erabiltzen diren 500 lema ezagutzearekin bertsok baten % 70 ulertuko luke (ahozkotatunak eragindako trabak eta sintaxiaren ulergarritasun-mugak kontuan hartu gabe, noski).

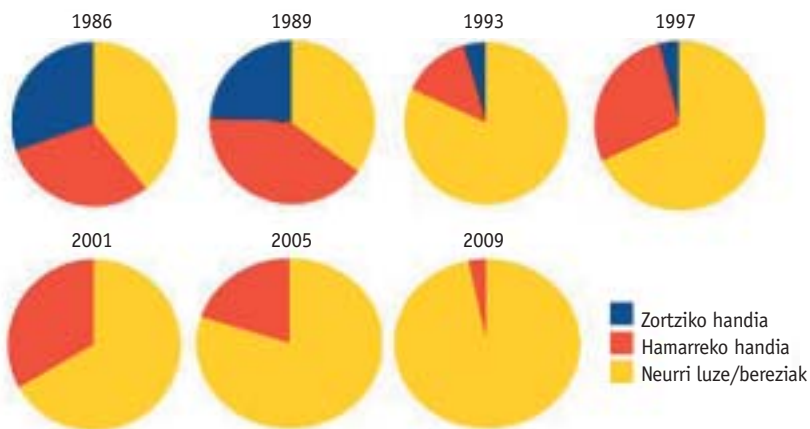
Bestalde, esan beharra dago txapelketetako corpus honek Zipf-ren legea betetzen duela. Hizkuntzaren prozesamenduan

	1986	1989	1993	1997	2001	2005	2009
1	eBDGRa	eBDGRa	ela	eBDGRa	eBDGRa	eBDGRa	ela
2	ela	iNMa	ala	eNMa	eNMa	ela	eBDGRa
3	uNMa	oSZa	eBDGRa	uNMa	aBDGRo	eNMa	aBDGRo
4	aBDGRBDGRa	iBDGRa	eNMa	ela	aBDGRa	ePTKa	eNMa
5	iPTKu	eNMa	ePTKo	aBDGRBDGRa	iBDGRa	aBDGRo	aBDGRa
6	ala	aBDGRBDGRa	aBDGRBDGRa	ePTKa	uNMa	uNMa	iBDGRa
7	iNMa	aNMa	aBDGRa	ala	ePTKo	ePTKo	ePTKa
8	uPTKa	ia	iNMa	iNMa	ala	aBDGRa	uPTKa
9	eNMa	aPTKu	aNMa	iBDGRa	aNMa	aBDGRBDGRa	iNMa
10	aBDGRBDGRia	ila	aBDGRo	aPTKu	ePTKiNM	iBDGRa	ePTKo

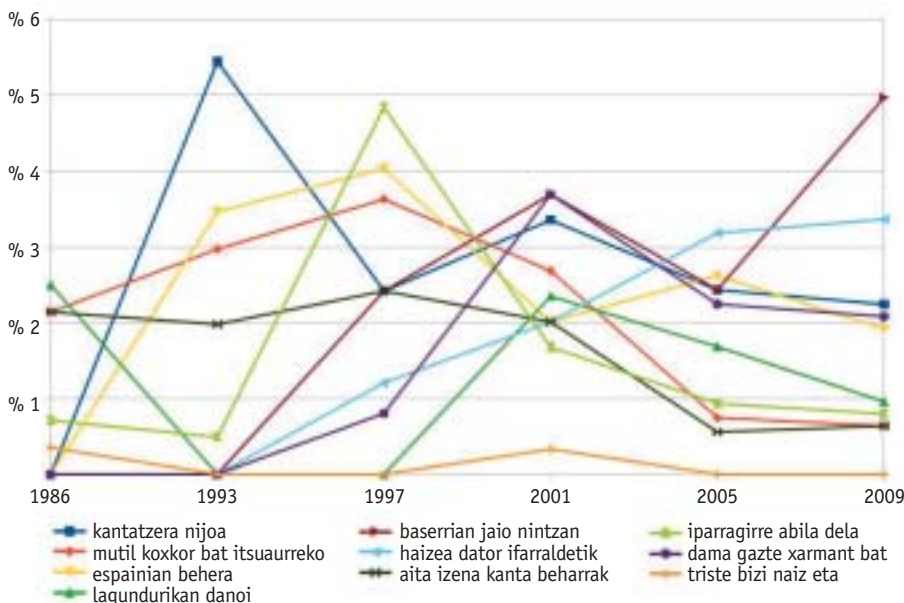
2. irudia. Txapelketa bakoitzean gehien erabilitako errimak. Oharra: eBDGRa errimak eba/eda/ega/era hartzen ditu barnean, bodegero legearen arabera ontzat jotakoak; horren arabera, “arriba”, “hobe da”, “ernega” eta “tankera” hitzek, adibidez, errimatu egiten dute. Berdintsu gertatzen da ePTKa, eNMa eta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



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



4. irudia. Kartzelako ariketan erabilitako neurriak.



5. irudia. 10 doinuren erabileraren bilakaera.

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 hurrengo hitza X/4 aldiz, eta hurrengo hitza X/8 aldiz...

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, adjektiboan erabileraren bilakaera; izan ere, behera egin du txapelketaz txapelketa, nahiz eta aldea ez den oso esanguratsua.

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. grafikoa 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 transkribatzeakatsak (% 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.

Azken txapelketetako neurriek, gure iritziz, bi motatako aurreikuspenak iradokitzen dituzte, nahiz eta dauzkagun datuak



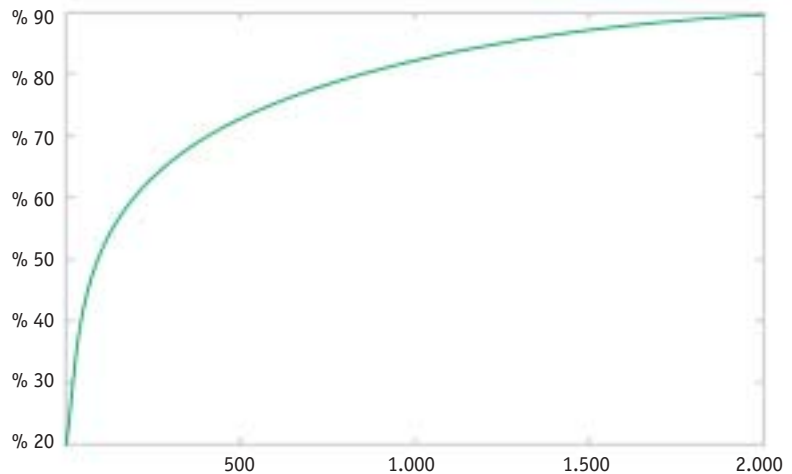
behar bezain zehatzak ez diren eta ondorioak ateratzeko goizegi dela iruditzen zaigun: joera hori alderantzikatu egingo da aurrerantzean, eta euskalkietako hizkera gehiago erabiltzera joko dute berriz bertso-lariek; edo euskara batuaren erabileran goimuga (% 90) horren bueltan ibiliko dira hemendik aurrera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

ONDORIOAK

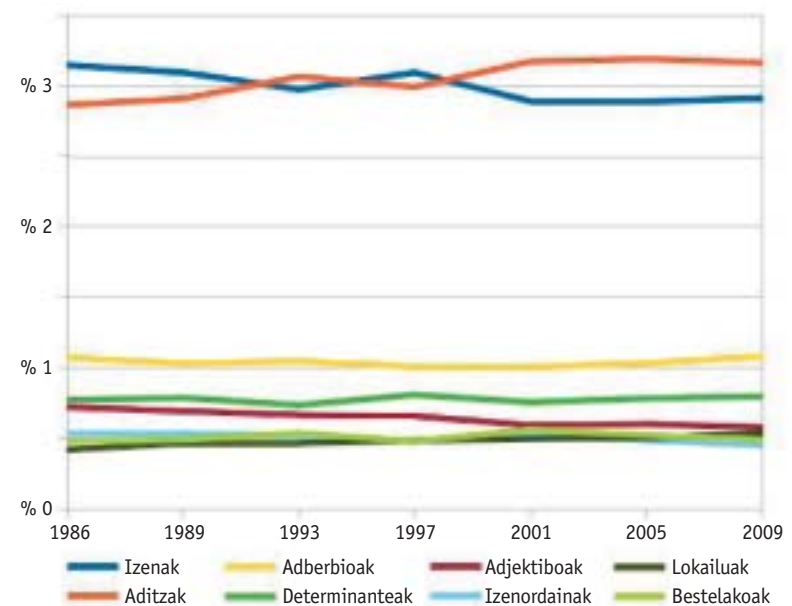
Azken zazpi txapelketa nagusietako bertsoak modu estatistikokoan aztertzeak zenbait joera azaleratzeko aukera eman digu. Datuon azterketa pausatuagoa eta xeheagoa egiteak mereziko duen arren, estreinako honenek ere utzi dizkigu kontu esanguratsu batzuk. Neurriaren aukeraketan eta euskara batuaren erabileran, 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 ondorioak 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 azaleratuko lituzke? Eta txapelketakoak direnak txapelketakoak ez direnekin konparatzeak?

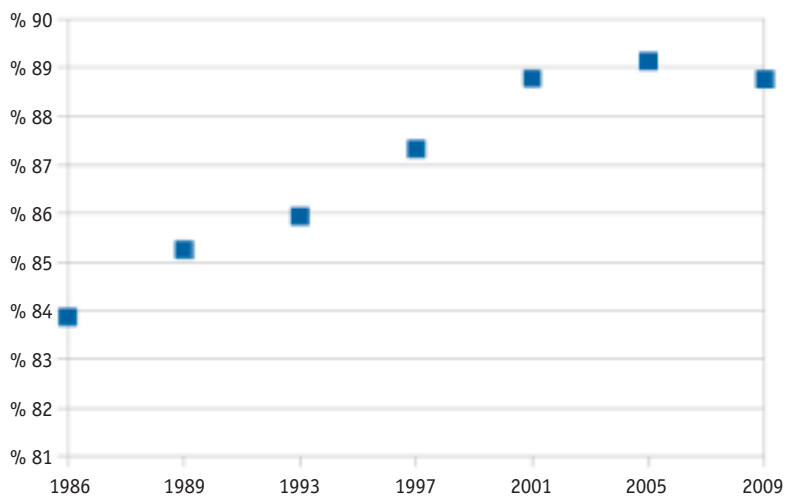
Lan handia dago egiteke alor honetan oraindik, baina bertso gintzaren azterketa xehea egiteko bertsoak egoki dokumentatzen jarraitzeak daukan garrantzia ukaezina da gure ustez, baldin eta ikusi nahi bada behintzat zer bilakaera hartzen duten hurrengo urteetan artikuluko honetan aipatutako joerek eta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



6. irudia. Lema erabiltzen diren bertsoen proportzioa.



7. irudia. Bertsoetan erabiltzen diren hitzen morfologia kategorien proportzioa.



8. irudia. IXA taldearen lematizatutako hitzen proportzioa.