

BasqueParl: Eusko Legebiltzarreko hitzaldien hizkuntza-azterketa eta zenbaketa



Sustatu

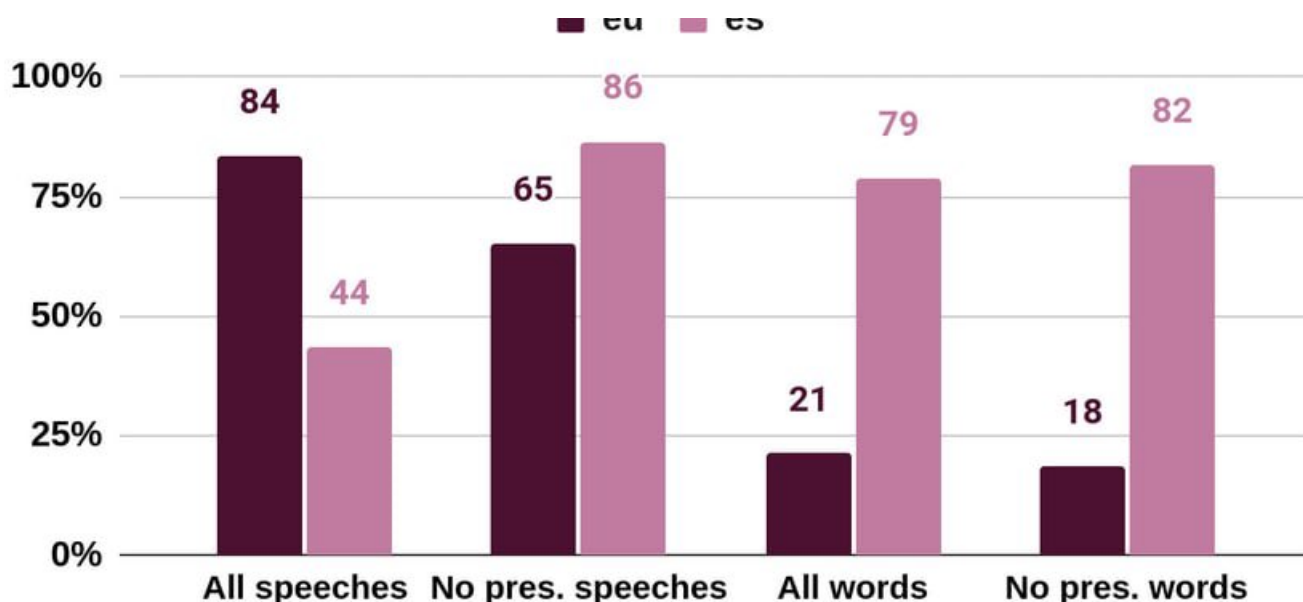
2022-06-21 : 09:37

Nayla Escribano (https://twitter.com/nayla_escribano) Hitz (<https://hitz.eus>) hizkuntzalaritza ikertegiko ikertzaileak korpus interesgarri bat prestatu eta landu du: BasqueParl (<https://github.com/ixa-ehu/basqueparl>), 14 milioi hitzezko korpus elebidun bat, Eusko Legebiltzarrean (<https://www.legebiltzarra.eus/portal/eu>) 2012tik 2020ra egindako hitzaldien transkripzio multzoa. Laburbilduz, esandako 14 milioi hitz horietan, 3 milioi izan dira euskaraz, eta 11 milioi erdaraz.

Field	Field type	Speeches	Paragraphs	Words	Lemmas	Entities
Overall	Overall	41,417	342,666	13,872,105	5,090,573	349,890
Language	Basque (eu)	34,571	133,599	2,938,061	1,375,686	122,439
	Spanish (es)	18,016	209,067	10,934,044	3,714,887	227,451
Gender	Female	30,857	185,119	6,452,503	2,424,765	179,030
	Male	10,559	157,481	7,416,852	2,664,596	170,782
Party	EAJ-PNV	28,530	143,685	4,330,467	1,701,415	145,400
	EH Bildu	3,638	58,391	2,423,245	1,009,715	56,440
	PP	3,562	51,683	2,672,062	858,798	56,376
	PSE-EE	2,834	44,738	2,327,350	769,767	43,994
	EP	1,467	24,235	1,171,948	410,966	24,738
	UPyD	1,368	19,110	909,007	322,220	22,270
Speaker	President	22,772	60,326	753,231	315,138	62,317

(<https://media.sustatu.eus/media/argazkiak/za1/hitza1.jpeg>)

Escribanok berak laburbildu zituen aurkikuntza nagusiak Twitterreko hari batean (https://twitter.com/nayla_escribano/status/1536276960896327683). Horra datu batzuk. Hitz kopuruan proportzioak %79 / %21 diren arren euskararen kaltetan (biribilduz 11 milioi hitz gaztelera, 3 milioi euskaraz), "hitzaldiak"! kontzeptua ikusita, euskara ageri da hitzaldien %84an, eta gaztelania bakarrik %44an. konparaketak, portzentaia eta zenbaki absolutoetan:

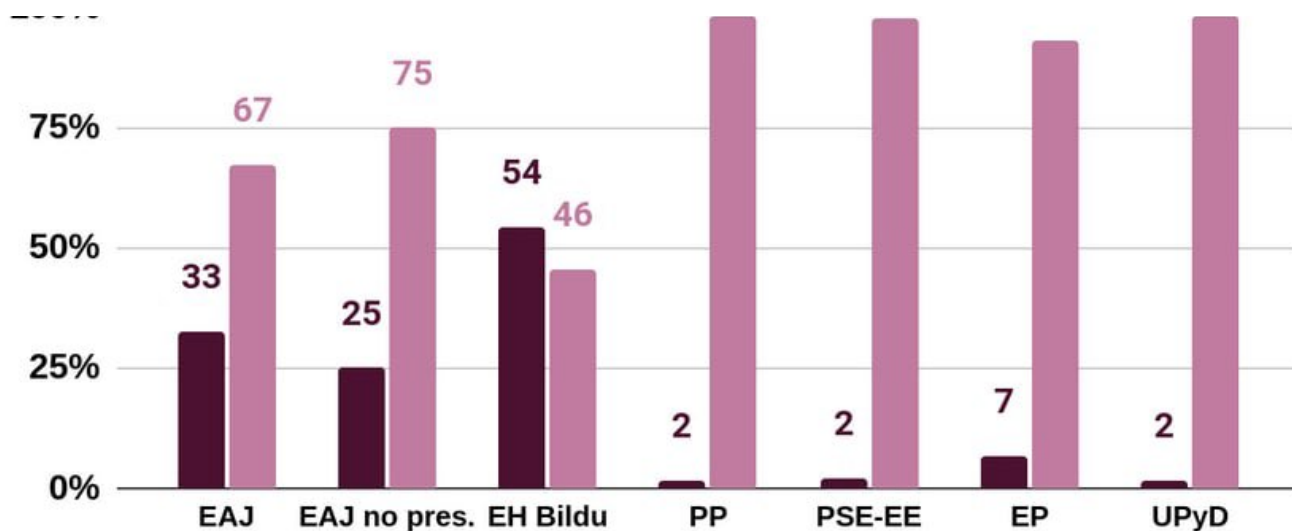


Field	Field type	Speeches	Paragraphs	Words	Lemmas	Entities
Overall	Overall	41,417	342,666	13,872,105	5,090,573	349,890
Language	Basque (eu)	34,571	133,599	2,938,061	1,375,686	122,439
	Spanish (es)	18,016	209,067	10,934,044	3,714,887	227,451
Gender	Female	30,857	185,119	6,452,503	2,424,765	179,030
	Male	10,559	157,481	7,416,852	2,664,596	170,782
Party	EAJ-PNV	28,530	143,685	4,330,467	1,701,415	145,400
	EH Bildu	3,638	58,391	2,423,245	1,009,715	56,440
	PP	3,562	51,683	2,672,062	858,798	56,376
	PSE-EE	2,834	44,738	2,327,350	769,767	43,994
	EP	1,467	24,235	1,171,948	410,966	24,738
	UPyD	1,368	19,110	909,007	322,220	22,270
Speaker	President	22,772	60,326	753,231	315,138	62,317

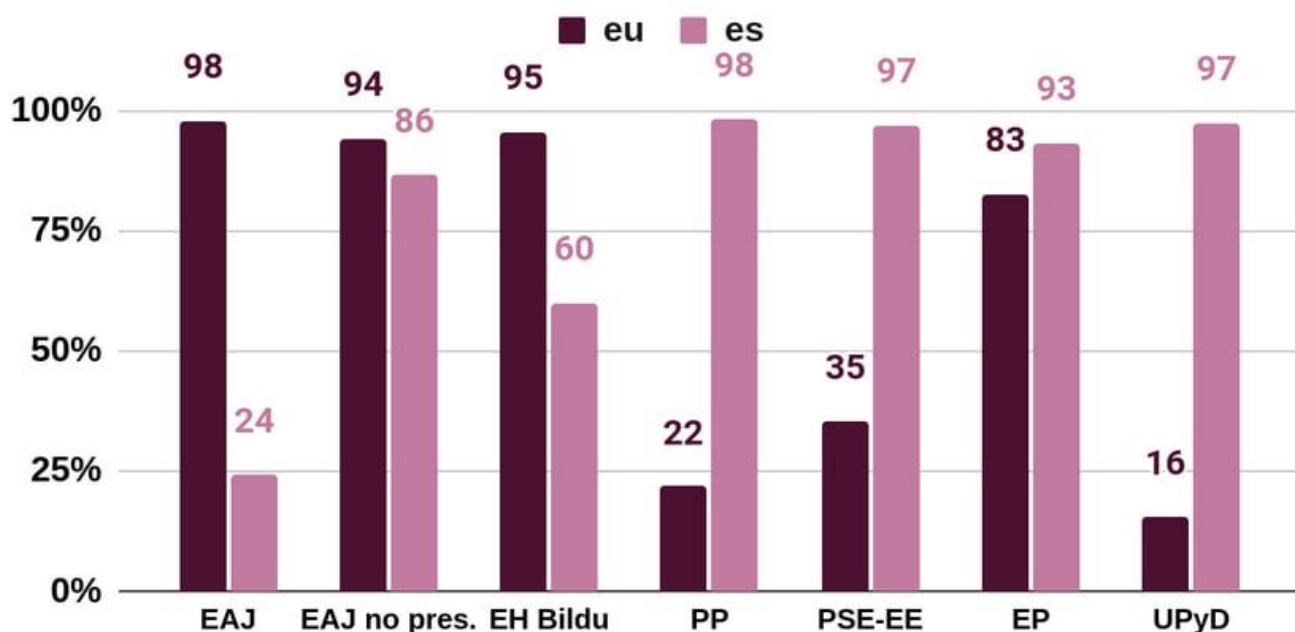
Nola da posible paradoxa? Euskarazko ohar tarazpen laburrak oso ohikoak direlako, eta bereziki Legebiltzarreko presidentziatik halako oharrek egiten direlako ohikoan (eta azterketako urteetan, 2012tik, EAJko politikari bat da presidentea: Bakartxo Tejeria). "Hitzaldi" gehienak berriz, presidentziako ohar horiek dira, eta gehienak euskara hutsezkoak dira. Adibidez; "Orain erresoluzioa bozkatuko dugu" edo "Eskerrik asko Iriarte anderea".

Alderdika esaten diren hitzetan, bereizketa egin du ikertzaileak presidentziaren hitzak alde batera utziz gero EAJren datuak alderatzeko. Ondorio politikoak ateratzeko, begiratu grafikoari eta zeuk atera:





Goiko taula hitz kopuruari dagokiona da. Beherago itsatsi duguna hitzaldiena. Ikus, adibide gisa, kontrastea Podemosen jardunean: euskara izan, badago beren hitzaldien %83an, baina euskarazko hitzen proportzioa %7a da. EAJk ere beherakada nabarmena du, hitzaldien %98tik hitzen %33ra.



Izan ere, hitzaldi asko formulazio honekin egiten dira (adibidea ikerketakoa da, ez da esaten noren hitzaldia den):

Basque Spanish

Gauzak egiten dira eta uste dut nik, nik ere eskubidea dudala Gobernuak eta beste erakundeek egiten dutena esateko. Zeren beti ver el vaso medio vacío o medio lleno, pues cambia un poco la perspectiva y vernos siempre en modo Gobierno, creo que no es nada objetivo. Se hacen cosas, se harán cosas y esta vez creo

que me deberían reconocer que de la iniciativa primera a lo que hemos acordado, no nos hemos dejado nada o creo que casi nada. Entonces, bueno, sólo quería aclarar eso eta eskerrak berriro.

Hizkuntza analiziaz gain, hizlarien generoa ere aztertu du Escribanok eta hara datua: emakumezkoen ahotsa gutxiengoa izatetik alderantzizko egoerara igaro da hamarkada batean:

Irakur hari osoa (https://twitter.com/nayla_escribano/status/1536276943942963200), interesgarria da oso.

Ikerketako datuak libre egin dituzte Escribanok eta Hitz zentroak, Legebiltzarreko hitzaldiak eta informazioa libreak baitira:

- Korpusa: <https://github.com/ixa-ehu/basqueparl> (<https://github.com/ixa-ehu/basqueparl>)
- arXiv Interneteko artxiboan: <https://arxiv.org/abs/2205.01506> (<https://arxiv.org/abs/2205.01506>)
- Demo bat bilaketa eta kalkuluetarako: <http://legebiltzarra.ix.a.eus> (<http://legebiltzarra.ix.a.eus>)



CodeSyntaxek kudeatua, Eusko Jaurlaritzaren Hezkuntza, Hizkuntza Politika eta Kultura Sailak (Hizkuntza Politikarako Sailburuordetzak) (<https://www.kultura.ejgv.euskadi.net/r46-704/eu/>) diruz lagundua.