

BasqueParl: Eusko Legebiltzarreko corpus elebiduna

Nayla Escribano*, Jon Ander González*, Julen Orbegoza-Terradillos**, Ainara Larrondo-Ureta**,
 Simón Peña-Fernández**, Olatz Perez-de-Viñaspre* eta Rodrigo Agerri*
 * HITZ Center - Ixa | ** Gureiker (Euskal Herriko Unibertsitatea UPV/EHU)
 {nayla.escribano, olatz.perezdevinaspre, rodrigo.agerri}@ehu.eus

Motibazioa

- Corpus berria **diskurtso politikoa** aztertzeko
- **Hizkuntza erabilera** aztertzeko euskara-gaztelaniazko *code-switching* corpus batean
- **Hiztunen diskurtsoa** aztertzeko hainbat **kategoriatan**: hizkuntza, generoa, adina, alderdia...
- **Parlamentarien portaera** gizartearekin eta gizarteak haiengan espero dutenarekin alderatzeko

Ekarpenak

- **BasqueParl corpusaren argitalpena**:
 - *Code-switching*-a daukan corpus elebidun berria
 - 2 legealdiko Eusko Legebiltzarren transkripzioak (2012-2016 eta 2016-2020)
 - 14 M hitz
- **Metadatuak**: generoa, adina, alderdia eta hizkuntza
- **Lemak eta entitate izendatuak**

Paragrafo baten adibidea

Euskara Gaztelania

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, nonos hemos dejado nada o creo que casi nada. Entonces, bueno, sólo quería aclarar eso eta eskerrak berriro.

→ **Gaztelaniazko paragrafoa**

Corpusaren deskribapena

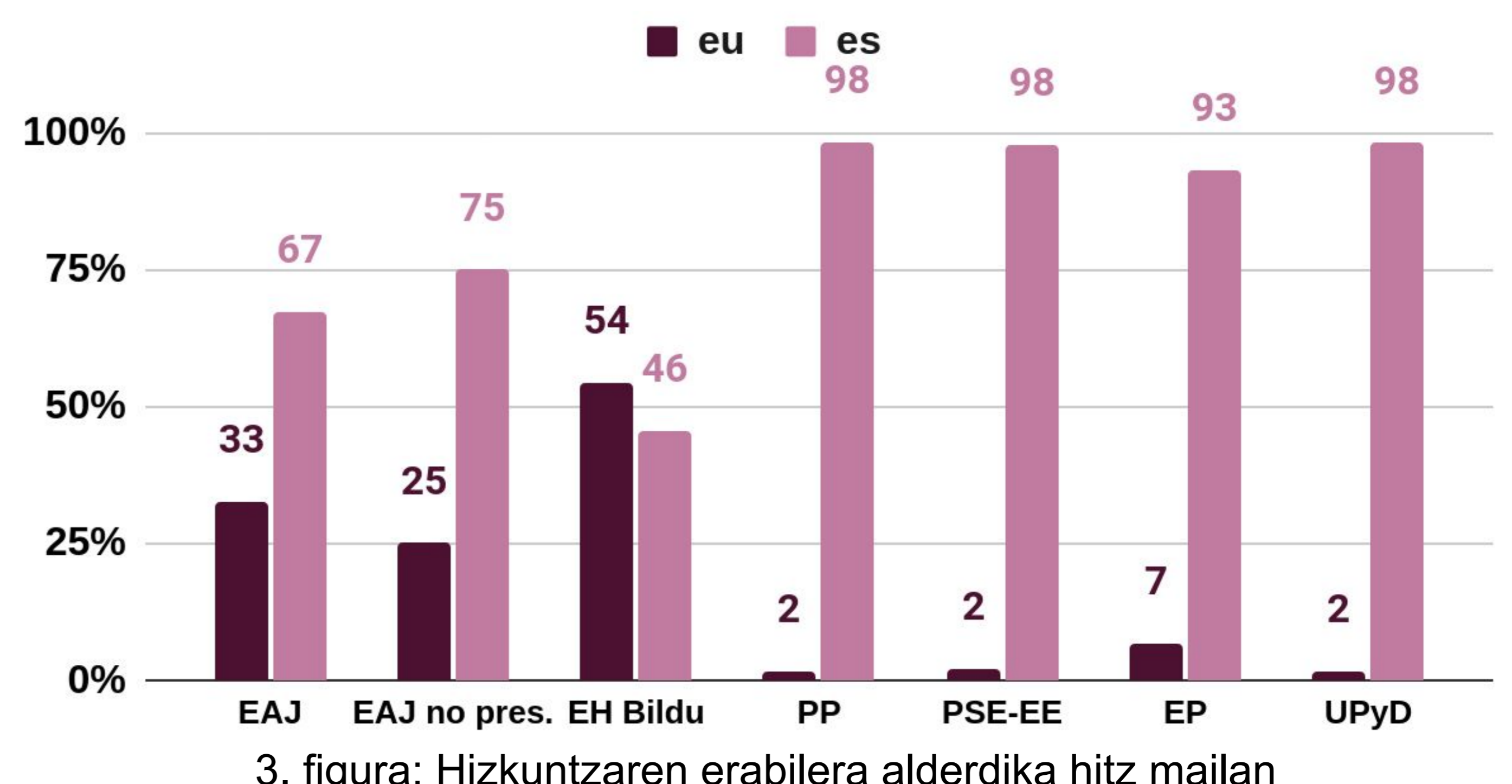
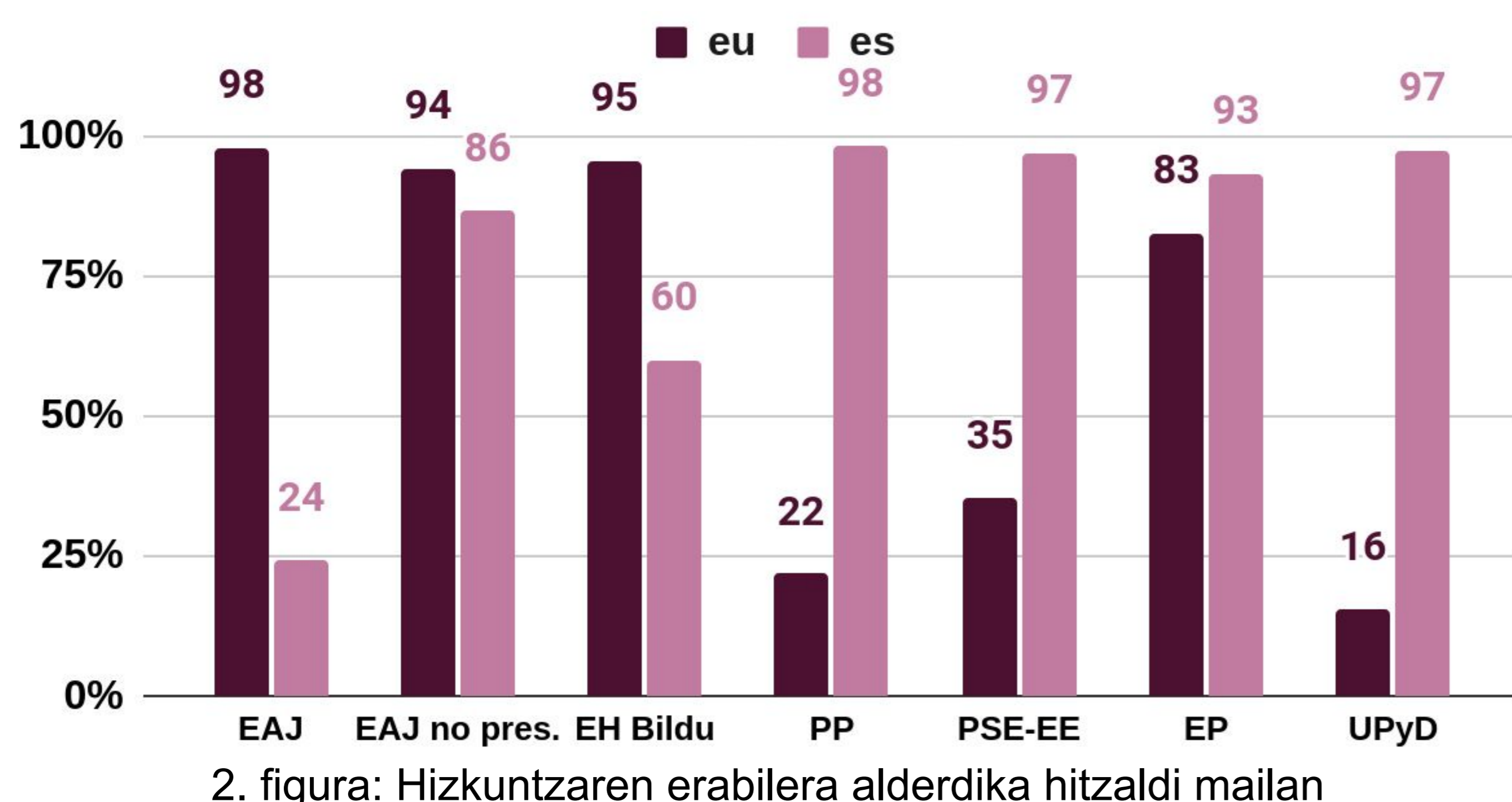
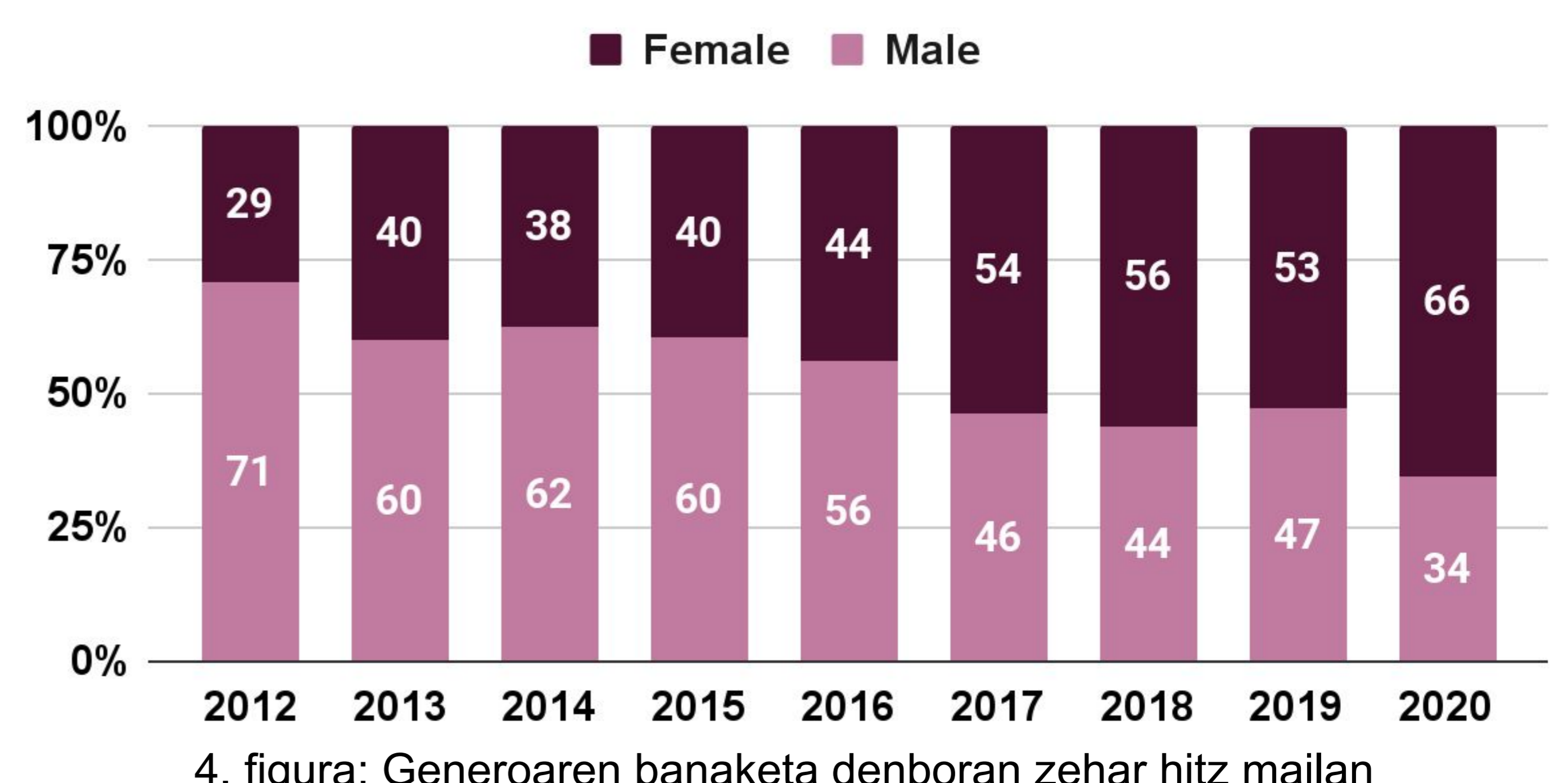
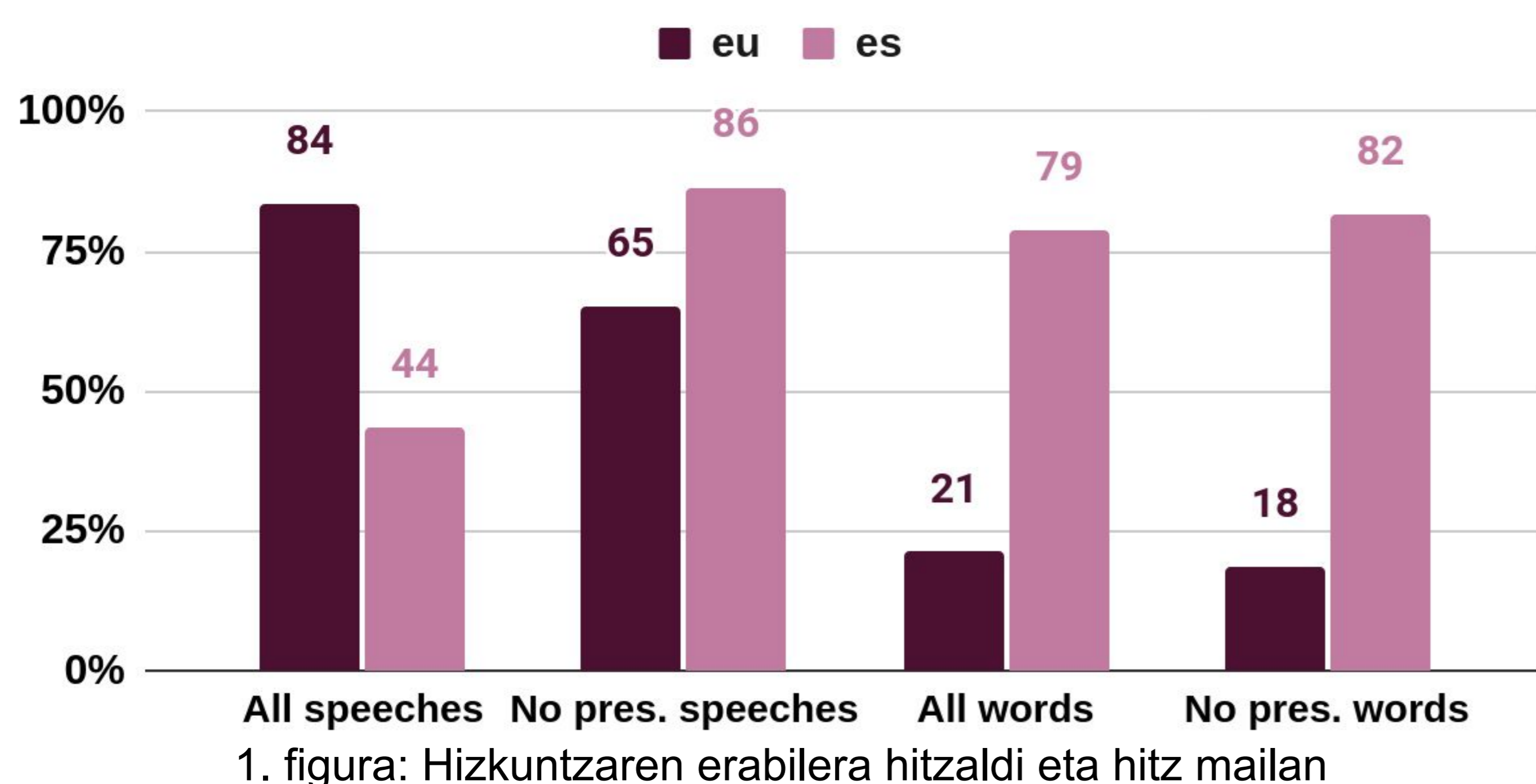
Arloa	Arlo mota	Hitzaldiak	Paragrafoak	Hitzak	Lemak	Entitateak
Guztira	Guztira	41,417	342,666	13,872,105	5,090,573	349,890
Hizkuntza	Euskara	34,571	133,599	2,938,061	1,375,686	122,439
	Gaztelania	18,016	209,067	10,934,044	3,714,887	227,451
Generoa	Emakumea	30,857	185,119	6,452,503	2,424,765	179,030
	Gizona	10,559	157,481	7,416,852	2,664,596	170,782
Alderdia	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
Hizlaria	Presidentea	22,772	60,326	753,231	315,138	62,317

Datuen analisia

Hitza: eu xor es

Paragrafoa: eu xor es

Hitzaldia: eu and/or es



Ondorioak

- **Euskara** askotan erabiltzen da hitzaldietan, baina hitz gehienak **gaztelaniaz** daude
- Antzeko emaitzak **alderdika**: EH Bilduk bakarrik erabiltzen ditu hitz gehiago euskaraz, eta alderdi espainiarren erabilera minimoa da
- **Emakumeen ordezkariak** hitz mailan alderantzikatu da urteetan zehar
- **Corpus librea**, eleaniztasunaren eta diskurtso politikoen analisiaren ikerketa publikorako