



Mikel Iruskietia¹, Jon Alkorta¹, Kike Fernandez¹, Ekain Arrieta¹, Rodrigo Agerri¹, Manex Aguirrezabal²
{mikel.iruskietia;jon.alkorta;ekain.arrieta;kike.fernandez;rodrigo.agerri;}@ehu.eus; manex.aguirrezabal@hum.ku.dk

¹HiTZ zentroa, Ixa taldea, Universidad del País Vasco/Euskal Herriko Unibertsitatea (UPV/EHU)

²Department of Nordic Studies and Linguistics (NorS), University of Copenhagen

ParlaMint 4.0

ParlaMint Europako zenbait parlamentuetako saioen transkripzioen corpus konparagarria da eta hasieratik parlamentuko saioak metadatu berak erabiliz osatu da. CLARIN-ERIC ikerketa azpiegituraren proiektu finantziatua da.

ParlaMint 4.0 datuetan:

- Corpusak 1,1 bilioi hitz ditu eta COVID garaia du aztergai
- 29 estatu eta autonomia erkidego bildu ditu; beraz, corpus eleanitza da
- Azpicorpusek 9 milioi eta 126 milioi hitz artean dituzte

Motibazioa

- Ikerketa azpiegiturak protokolo komunak eta estandarrak jarraituz europar mailako **ikerketa irekia** ahalbideratu nahi du:
 - Metadatu komunak: nork zer eta noiz esan duen adierazita
 - Eleanitza: esaldi bakoitza zein hizkuntzetan esan den etiketatuta
 - Aberastua: morfosintasia, NER, itzulpen automatikoa, sentimentuen analisia
 - Erabilterraza: ikergalderak erantzuteko tresna erabilterraz ugari
 - Curriculuma: eskolan erabiltzeko eskuliburuak, kasu arrakastatsuek...

Corpusaren antolamendua eta atalak

- Metadatuak: gobernuak, oposizioa, partidua, generoa, denbora...
- Mintzaldien transkribapenak: TEIrekin kodetuta, esaldiaren hizkuntza adierazita
- Automatikoki aberastua: informazio morfosintaktikoa, NER...

Mintzaldien transkribapenak

- Mintzaldiari dagokion legegintzaldia
- Eguna
- Mintzaldi-mota
- Egilea
- Esaldi bakoitzaren hizkuntza

Analisi linguistikoa

- Informazio morfosintaktikoa
 - <http://ufal.mff.cuni.cz/udpipe>
- Izen bereziak
- Rol semantikoak

Metadatuak

- Gobernuaren osaketa politikoa, 3 legegintzaldietan
- Alderdi politiko bakoitzari buruz:
 - Sortutako urtea
 - Alderdiaren izena
 - Laburdura
 - Webgunea
- Legebiltzarkideei dagokienez:
 - Jaioteguna
 - Sexua
 - Jaioterria
 - Alderdi politikoak
 - Webgunea



<https://www.clarin.eu/parlamint>

Euskal Legebiltzarreko eta gainontzeko corpusak eskuragarri

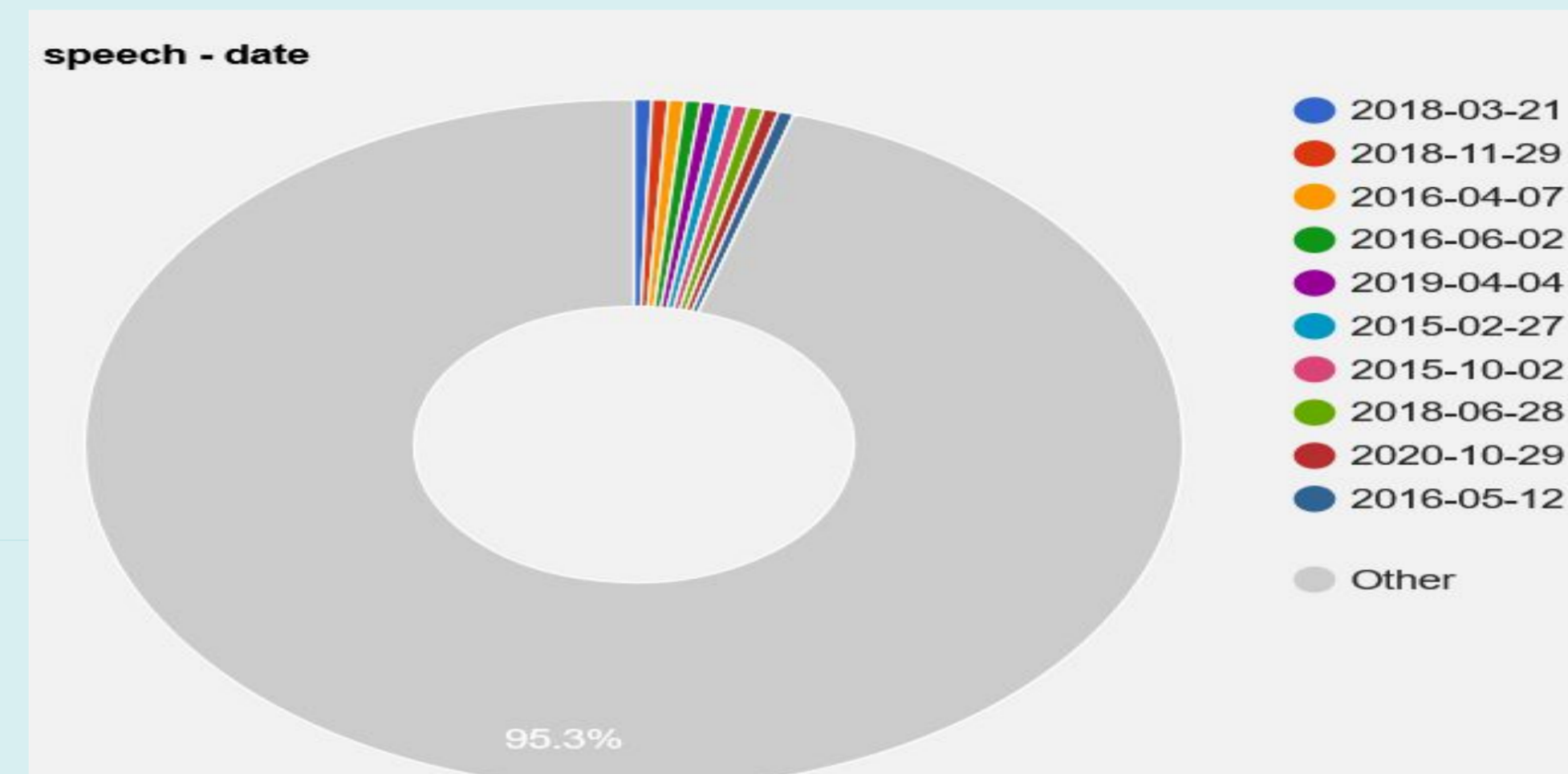
- ParlaMinten webgunean deskargatu
 - <http://hdl.handle.net/11356/1859> (corpusa)
 - <http://hdl.handle.net/11356/1860> (informazio ling.)
- Corpus analizatuak: https://www.clarin.si/ske/#dashboard?corpname=parlamint40_xx
- TEITOK: <https://lindat.mff.cuni.cz/services/teitok/parlamint-40/>
- Sketch Engine: <https://www.sketchengine.eu/parlamint-corpora-of-parliamentary-debates/>

Erabilera kasuak

Dobranić (2021)

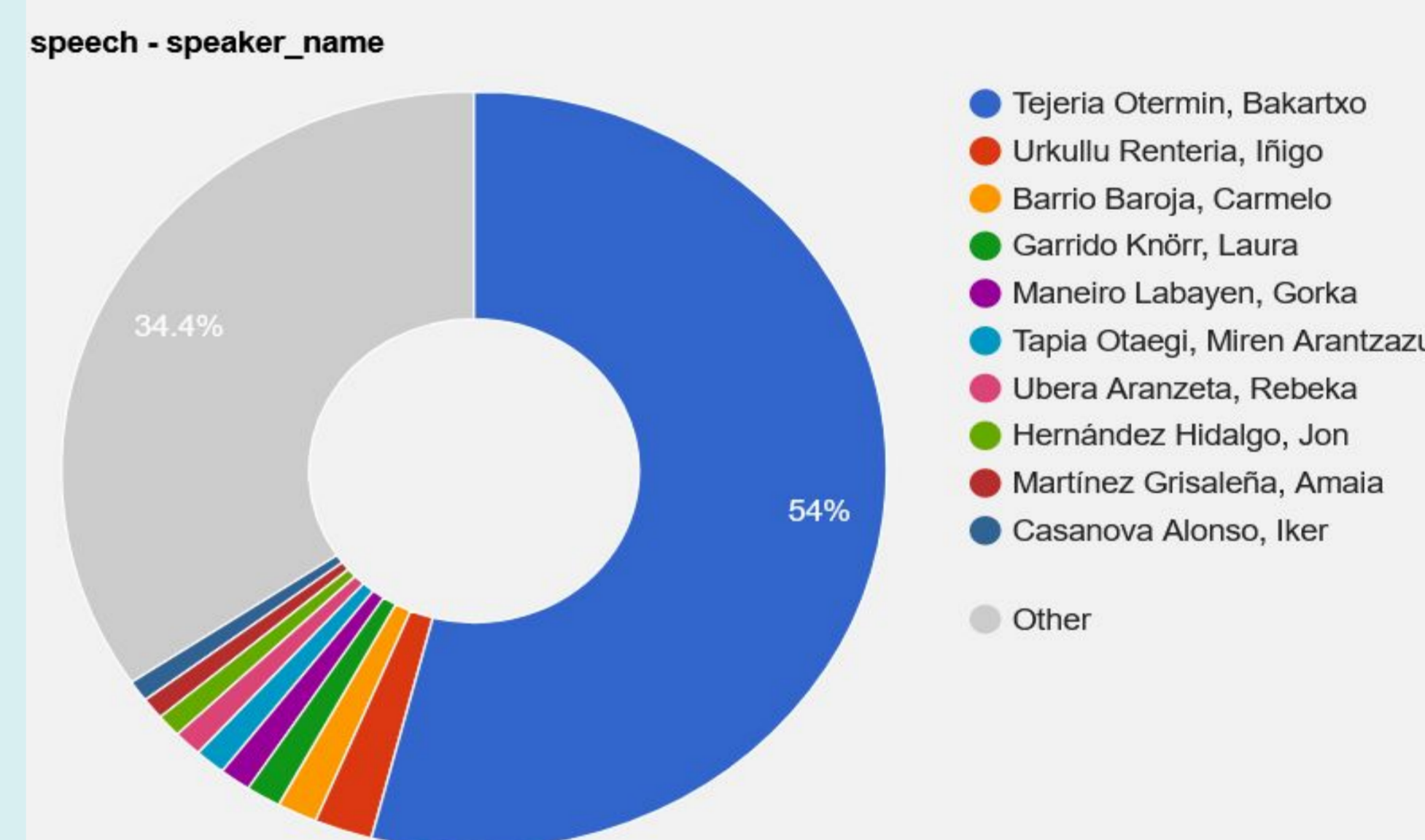
Bilkuren datuak:

- Izenburua
- Data
- Erakundea



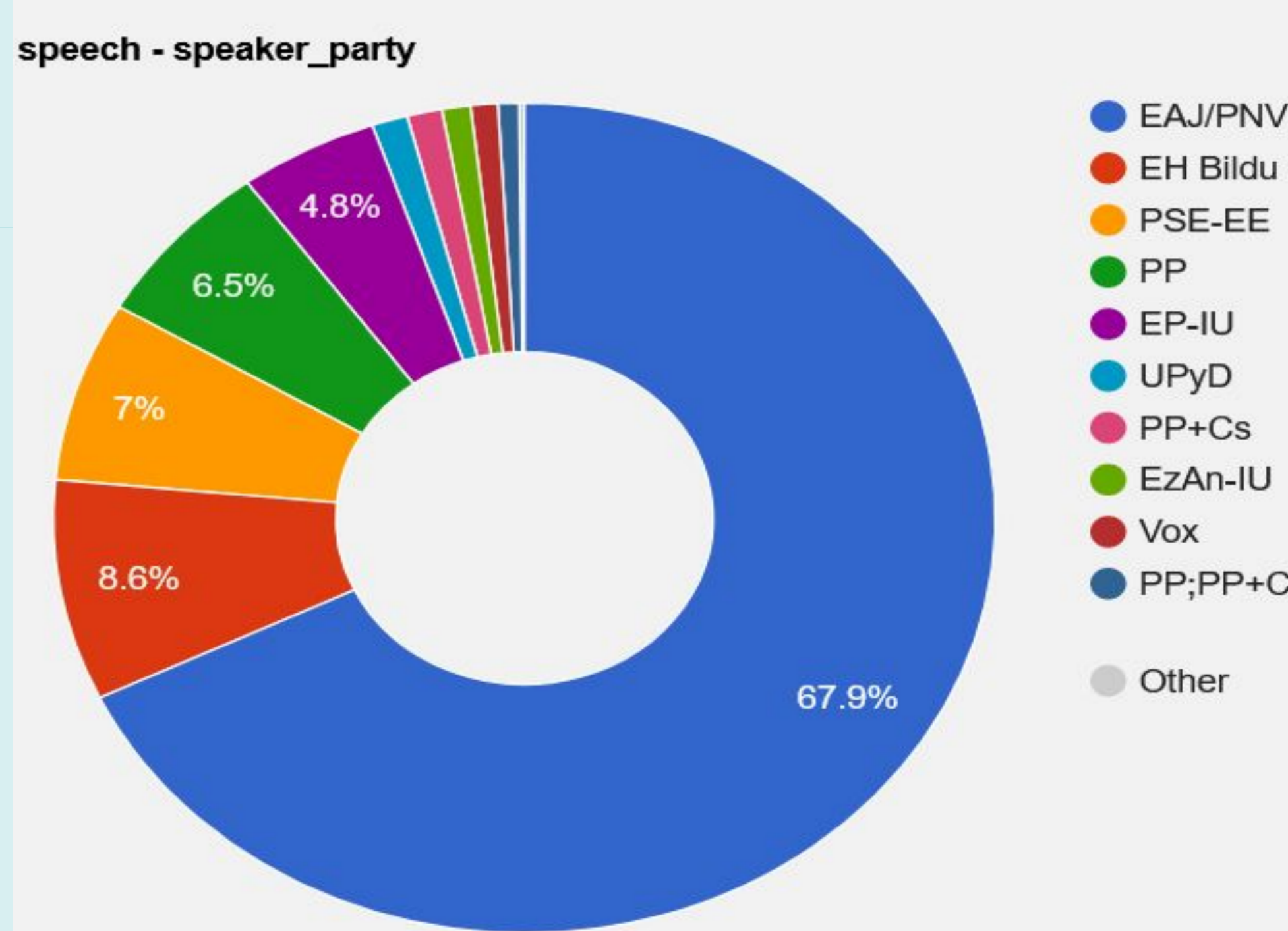
Legebiltzarkideenak

- Legebiltzarreko saioetara joandako kopurua eta ehunekoa
- Legebiltzarrean egindako galdera eta bultzatutako ekimenen kopurua
- Eztabaida bakoitzean parte hartu duten legebiltzarkideen kopurua
- Legebiltzarkide bakoitzak esandako hitz kopurua
- Legebiltzar taldekoak ez diren legebiltzarkideak



Legebiltzar taldeenak:

- Talde bakoitzaren eserleku kopurua
- Bozketetako bertaratzera
- Galdera eta ekimen kopurua
- Taldearekin bat ez datozen bozkek
- Hiztegi konplexutasuna



Legedia

Legeei buruzko honako datuak:

- Izenburua
- Kodea
- Data
- Egoera