

# HIZKUNTZA ALDAERAK EUSKARAZ: AURREKARIAK, SAILKAPENA ETA EBALUAZIO PROPOSAMENA

Jaione Bengoetxea, Itziar Gonzalez-Dios, Rodrigo Agerri  
 HiTZ Zentroa - Ixa, Euskal Herriko Unibertsitatea UPV/EHU  
 {jaione.bengoetxea@ehu.eus}

## MOTIBAZIOA

Egungo hizkuntza teknologiek zailtasunak dituzte hizkuntza estandarretik urruntzen diren hizkerak ulertzeko

Aldaerak esplizituki lantzeko beharra HPan

## HELBURUAK

- Aldaerak kontuan hartzen dituen euskarazko lehenengo **datu multzoak** biltzea
- Egungo hizkuntza teknologiek aldaerak **ulertzeko** duten **gaitasuna ebaluatzea** eta hobetzea
  - Eskuragarriagoak eta inpartzialagoak diren tresnak garatuz

## SAILKAPENA HIZKUNTZALARITZAN

- Hizkuntzalaritzan, aldaeren sailkapen zabal eta sakonak daude:
  - Hainbat ezaugarrien arteko elkar ekintza oinarri hartuta
  - Erregistroak oinarri hartuta

Dialektoa

Batua

Tokian-tokikoa

Erregistroa

Formala

Informala

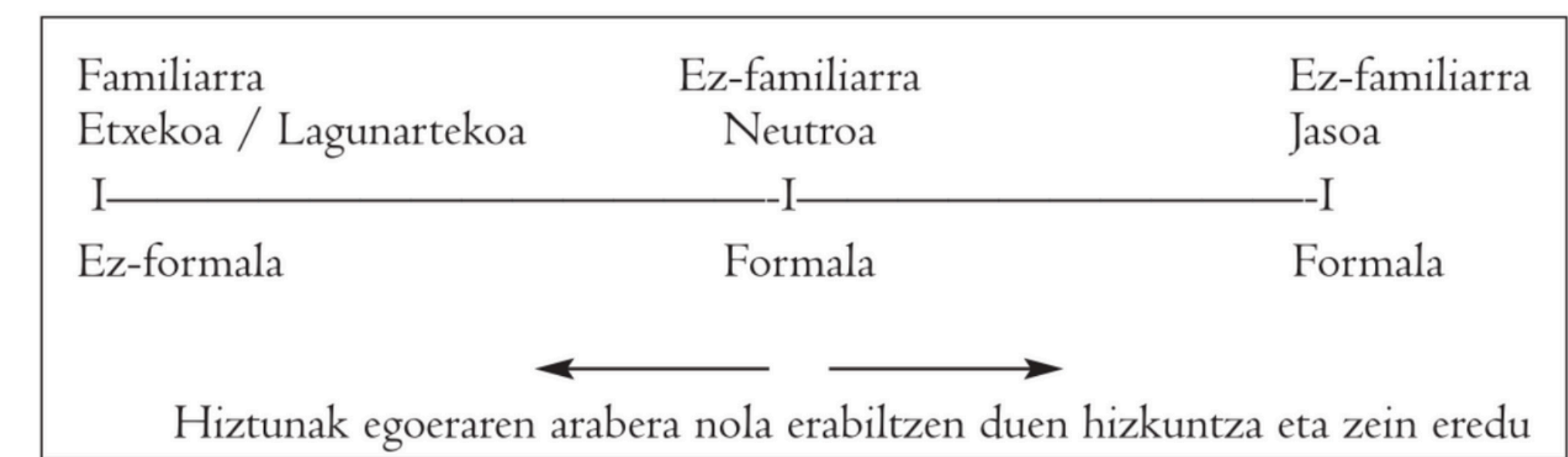
Hitzeko komunikazioa

Idatzizkoa

Ahozkoa

Garzia, J. (2018)

- Aldaera guzti hauek HPan lantzea konplexua da euskaraz:
  - Estandarretik urruntzen den datuen kopurua mugatua da

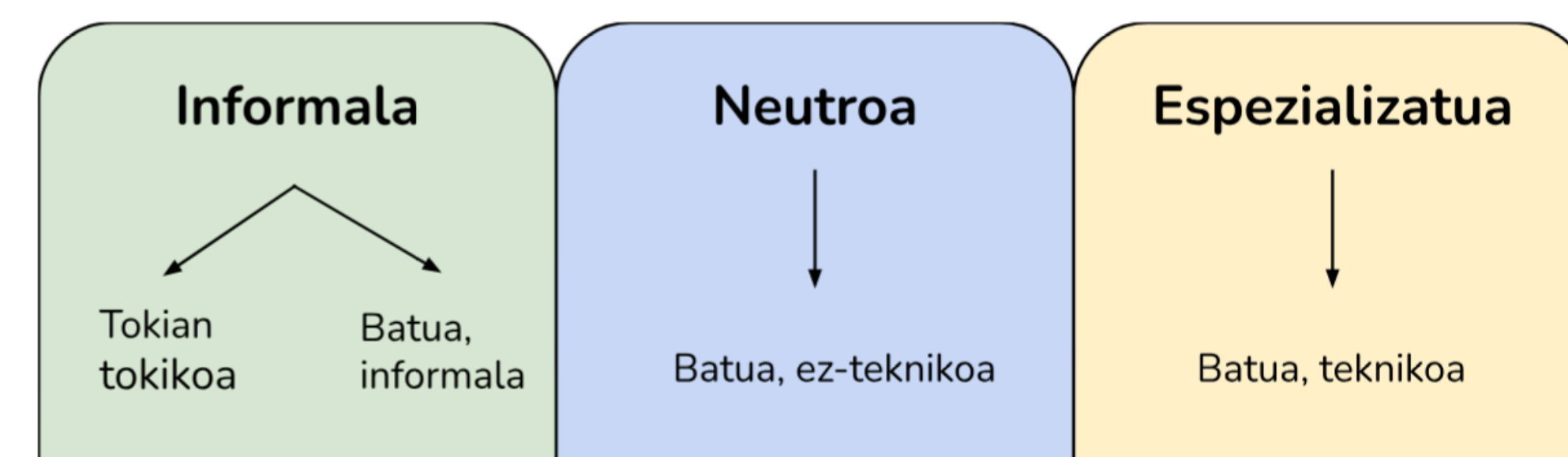


Hiztunak egoeraren arabera nola erabiltzen duen hizkuntza eta zein eredu

Barrios et al. (2008)

## SAILKAPENA HIZKUNTZAREN PROZESAMENDUAN

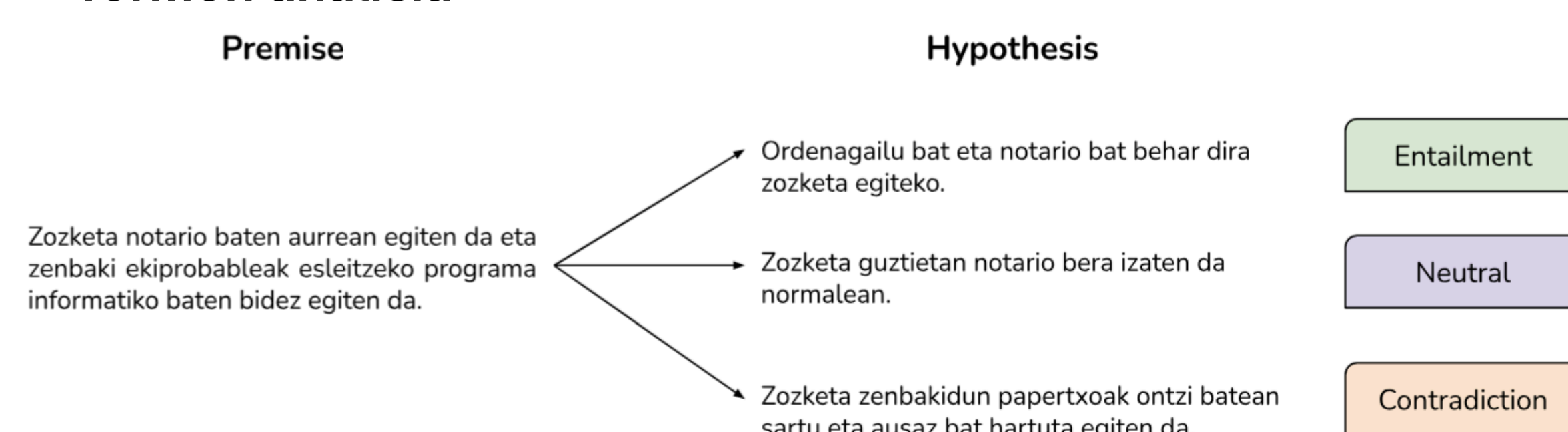
- Sailkapen sinplifikatuago baten beharra:
  - Erabilgarri dauden datuetan oinarritutako sailkapen proposamena
  - Hizkuntzalaritzako sailkapenetan oinarrituta, gure proposamena egungo euskal hizkuntza teknologietan erabili daitekeen sailkapena da



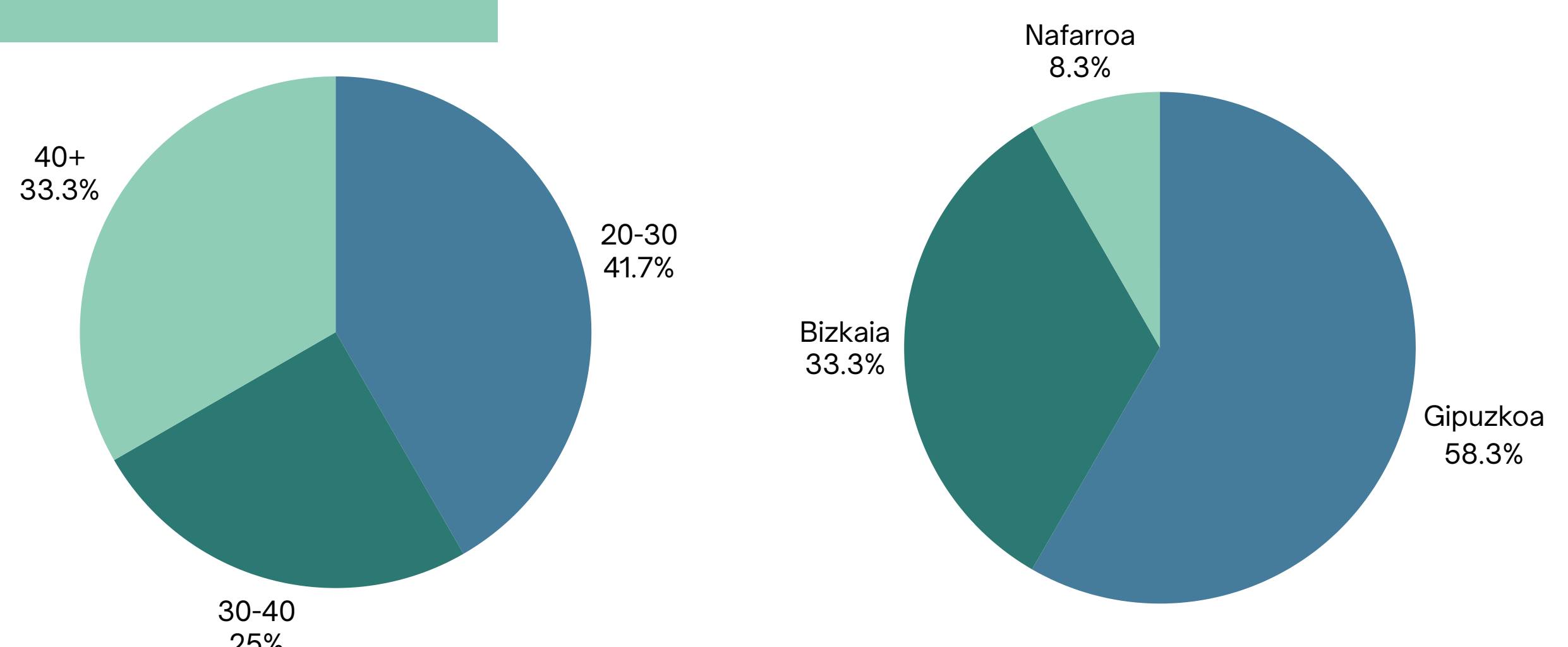
## EBALUAZIOA

- Hizkera estandarrean ez dauden premisek duten eragina ebaluatu nahi da NLI atazan
- XNLI-eu datu multzoa (Heredia et al., 2024)
  - Euskal hiztunei eskatu premisen parafraasiak eskuz sortzeko
  - Guida lerro gutxi: esaldi hau modu "informalago" batean nola esango zenuke?
- Helburuak:
  - Aldakortasuna: hizkuntza teknologien **errendimendua ebaluatu**, NLI atazan aldaerak gehitzean
  - Parafraasiak: hiztunek parafraasi informalak sortzeko erabili dituzten

### formen analisia



## BERRIDAZKETAK



### Jatorrizkoa esaldia

Gaztetape hondartzara harriz betea dago; baina olagarru polit batekin egin genuen topo.

Egia da, bidenabar, liburua irakurrita edukita hobeto ulertzen direla hainbat pasarte, seguruenik.

### Berridatzitako esaldia

Gaztetape hondartza harriz beteta dau; baie amorrotz polit bat topa gendun.

Egie dek, de paso, liburue irakurrita eukita hobeto ulertzeittukela hainbat pasarte, seguruna.

## ETORKIZUNEN LANA

- Ebaluazioa: konparatu NLI atazaren emaitzak hizkuntza estandarrekin vs hizkuntza aldaerekin
- Euskarazko aldaerak barnebiltzen dituzten datu-multzo gehiago bildu
- Hizkuntza teknologien errendimendua hobetu euskararen aldaera ezberdinetan

## ESKERRAK

- DISARGUE - TED2021-130810B-C21f inanciado por MCIU/AEI /10.13039/501100011033 y por la Unión Europea Next GenerationEU/ PRTR
- TED2021-130810B-C21 financiado por MCIU/AEI /10.13039/501100011033 y por la Unión Europea Next GenerationEU/ PRTR