

# BIM eta SAHCOBA

## Morfosintaktikoki etiketatutako euskarazko corpus historikoa eraikitzen

Ainara Estarrona<sup>1</sup>, Izaskun Etxeberria<sup>1</sup>, Ricardo Etxepare<sup>2</sup>, Manuel Padilla-Moyano<sup>2</sup>, Ander Soraluze<sup>1</sup>, Kike Fernández<sup>1</sup>

<sup>1</sup>HiTZ Zentroa - Ixa, Euskal Herriko Unibertsitatea (UPV/EHU) / <sup>2</sup>CNRS - IKER (UMR 5478)  
 {ainara.estarrona}{izaskun.etxeberria}{ander.soraluze}{kike.fernandez}@ehu.eus  
 {r.etxepare}{manuel.padilla}@iker.cnrs.fr

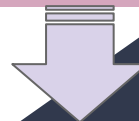
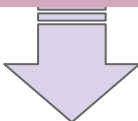


EUSKARAZKO  
CORPUS  
HISTORIKOA



**Basque in the Making (BIM): A Historical Look at a European Language Isolate**  
 (ANR: 17-CE27-0011 - BIM)  
 IN: Ricardo Etxepare (IKER-CNRS)

**Syntactically Annotated Historical Corpus in Basque (SAHCOBA)**  
 (MICINN: RTI2018-098082-J-I00)  
 IN: Ainara Estarrona (HiTZ-Ixa)



## HELBURUAK

- Euskararen gramatikaren ezaugarri zenbaiten azterketa diakroniko zehatza egitea
- Morfosintaktikoki etiketatutako corpus historikoa sortzea
  - XV-XVIII mendeak
  - Milioi bat hitz
  - Euskalki guztiak
  - Genero aniztasuna

## HELBURUAK

- Etiketaturako corpusa zabaltzea
  - XVIII-XX mendeak
  - 12 milioi hitz
- Corpusa arakatzeko bilaketa-interfazea garatzea
- Transkribaturako testuaren eta faksimilaren arteko lotura egitea

