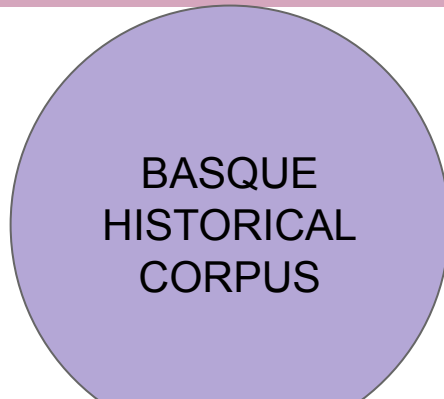


# BIM and SAHCOBA

## Building a morphosyntactically annotated Basque historical corpus

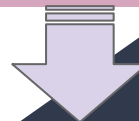
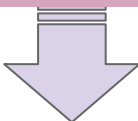
Ainara Estarrona<sup>1</sup>, Izaskun Etxeberria<sup>1</sup>, Ricardo Etxepare<sup>2</sup>, Manuel Padilla-Moyano<sup>2</sup>, Ander Soraluze<sup>1</sup>, Kike Fernández<sup>1</sup>

<sup>1</sup>HiTZ Center - Ixa, University of the Basque Country (UPV/EHU) / <sup>2</sup>CNRS - IKER (UMR 5478)  
{ainara.estarrona}{izaskun.etxeberrria}{ander.soraluze}{kike.fernandez}@ehu.eus  
{r.etxepare}{manuel.padilla}@iker.cnrs.fr



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

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



### GOALS

- Systematic diachronic study of a number of grammatical features of Basque
- Creation of a comprehensive morphosyntactically annotated historical corpus of Basque
  - 15th-18th centuries
  - 1 million words, all dialects

### GOALS

- Increasing the annotated corpus
  - 18th-20th centuries
  - 11 millions words
- Search interface to browse the corpus
- Link between the transcribed text and the facsimile

