

BIM eta SAHCOBA

Morfosintaktikoki etiketatutako euskarazko corpus historikoa eraikitzen

Ainara Estarrona¹, Izaskun Etxeberria¹, Ricardo Etxepare², Manuel Padilla-Moyano², Ander Soraluze¹, Kike Fernández¹

¹HiTZ Zentroa - Ixa, Euskal Herriko Unibertsitatea (UPV/EHU) / ²CNRS - IKER (UMR 5478)
 {ainara.estarrona}{izaskun.etxeberria}{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)

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

- Etiketatutako corpora zabaltzea
 - XVIII-XX mendeak
 - 12 milioi hitz
- Corpora arakatzeko bilaketa-interfazea garatzea
- Transkribatutako testuaren eta faksimilaren arteko lotura egitea

BIM armiarma

DB Anmota Lem Denak

Hitza 1 Mota Lema Kategoria ADT Hitz erorria edozien Kategoria Modua da Informazio morfologikoa

Hitza 2 Mota Lema Kategoria edozien Hitz erorria edozien Kategoria Modua da Ezauagari gehiago

Hitza 3 Mota Lema Kategoria edozien Hitz erorria edozien Kategoria Modua da Ezauagari gehiago

Bilatu Garbitu

Analisiak

Emaitzak (12)

1. neuen gogoa axea bideratzen iraraz. 420. Igutzeten edaten dena dantzaean estaztaran, esan dara alaianean. 239. Kibel ekabeari legia egin eta esan aita gure gara dura horretan. 101. Jokabidean dora. 472. Urak dantzaera ura dora. 473. Neure neskedo kee 264. Lurren lurrean dora. 265. Urdu egoik gogotzatzen badakinean. 266. Gogotzatzen gogotzatzen. 267. Gogotzatzen autengoez. 368. Otsaklean, urtelean dora arazak lezeraren 369. Ezauagari gehiago.

TEFRASIS Y SENTENCIAS (196)

362. Saria noiz, aspernetan gatz etzortu. 363. Samari osoa daga gestor, nok berrik ar bez tenet idioti. 364. Saria lekuetan oihala, dabelean sabelen obabu. 365. Biastan, zegik inoiz haurtan petan. 367. Ezozzen zozorek dau luperak.

362. Saria noz, aspernetan gatz etzortu. 363. Samari osoa daga gestor, nok berrik ar bez tenet idioti. 364. Saria lekuetan oihala, dabelean sabelen obabu. 365. Biastan, zegik inoiz haurtan petan. 367. Ezozzen zozorek dau luperak.

366. Dagaera belarren oihala, tabaketa tabaketa oihala. 367. Ezozzen zozorek dau luperak.

368. Otsalainen, urtelean dora arazak lezeraren 369. Ezauagari gehiago. 370. Ezauagari gehiago. 371. Neure neskedo kee 264. Lurren lurrean dora. 265. Urdu egoik gogotzatzen badakinean. 266. Gogotzatzen gogotzatzen. 267. Gogotzatzen autengoez. 368. Otsaklean, urtelean dora arazak lezeraren 369. Ezauagari gehiago.

Analisis komunitario de edoien-brain herriarrietan hitza (idoia u.s.) correspondiente al prefijo adnáin breka herriar hitza (MELV 1995), el 211 goranzko 362-363 = 196-197. 2. tanda goranzkoak, kuts lehengo 196-197.

