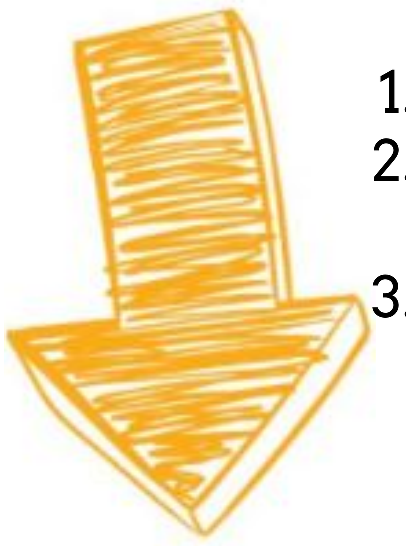
Eskolako laburpen-testuak biltzeko baliabideak eta euskarazko laburpenen corpusa

Unai Atutxa-Barrenetxea: Mikel Iruskieta, Olatz Ansa
unai.atutxa@ehu.eus; mikel.iruskieta@ehu.eus; olatz.ansa@ehu.eus
Ixa taldea, HiTZ Zentrua, Euskal Herriko Unibertsitatea (UPV/EHU)

Zertarako eskolako laburpen-testuen corpusa?



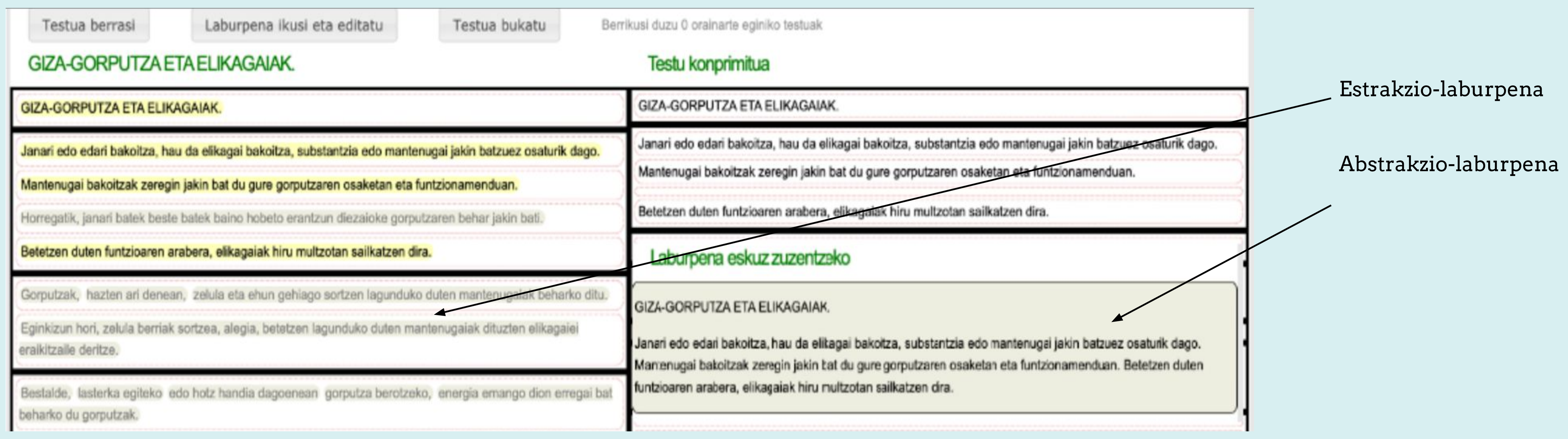
1. Ikasketa-maila ezberdinetako ikasleen laburpenak aztertzeko
2. Ikasketa-maila bakoitzean izan beharreko laburpen-gaitasuna zehazteko
3. Curriculumeko oinarrizko jakintzen eta helburuen arteko loturak esplizitu izan ditzan irakasleak

CLARIAH-EUS ikerketa-azpiegituraren ekarpena

- Tesietan egindako datuak erraz bilatzeko tresnak sortzea
- Beste ikertzaileentzat ikusgarri eta berrerabilgarri jartzea
  - CLARINen VLOn eskuragarri jarri
  - CC BY-NC 4.0 lizentziapean

Corpusa biltzeko erabilitako baliabideak

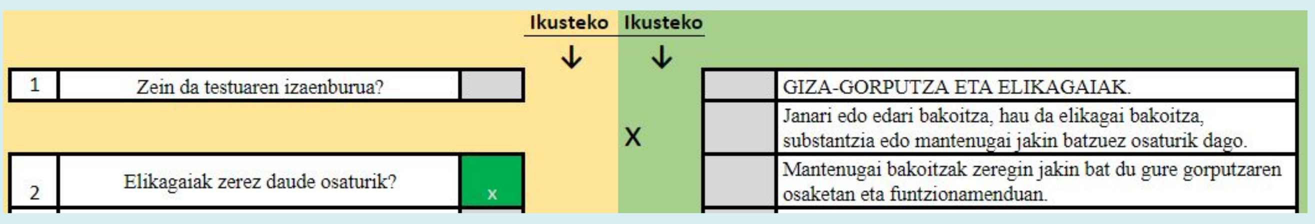
Compress-eus
laburpenak jasotzeko tresna (Atutxa et al., 2017)

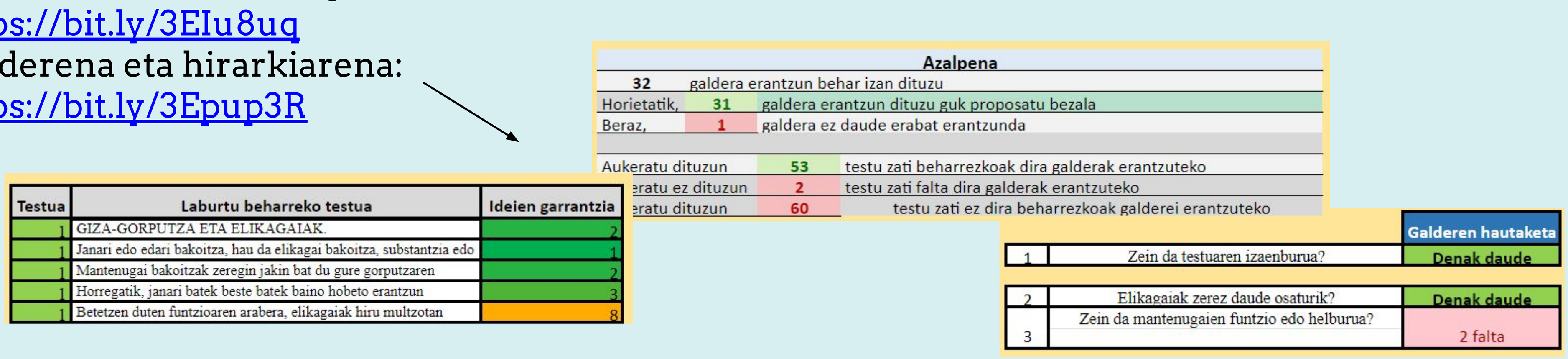


http://ixa2.si.ehu.eus/compress-eus

Hierarkiaren atzeraelikadura automatikoa
(Atutxa, 2022)

- HIMAM metodoa: hierarkia mailan oinarritua: https://bit.ly/3TLT1tj
- GOM metodoa: galderetan oinarritua
  - Galderak erantzuten laguntzekoa: https://bit.ly/3EIu8uq
  - Galderena eta hirarkiarena: https://bit.ly/3Epup3R





Laburpena lantzeko tailerra (Atutxa, 2022)

- Compress-eus
- Atzeraelikadura automatikoa
- Laburpenak egin eta ebaluatzeko irizpideak
- Ebaluaziorako errubrika

https://atutxaunai.wixsite.com/lab-tailerra2022

Corpusaren ezaugarriak

Lehen Hezkuntzako testuak
Laburtu beharreko testuak:

- LH 5 eta LH 6. mailako testuak, adin honetan egokia delako laburpena lantzen hasteko (Sanz-Moreno, 2005)
- Euskarazko 13 testuren laburpenak

LH eta unibertsitateko ikasleak laburtzaile
Zergatik?

- LHko ikasleek eskolan landu beharreko testuak direlako
- Unibertsitateko ikasleak etorkizuneko irakasleak direlako, testu horien edukiak landu beharko dituztenak

Compress-eus erabilita
Laburpen guztiak tresna hau erabiliz egin dira

- Ikasle bakoitzaren estrakzio eta abstrakzioak biltzen ditu
- Automatikoki gordetzen ditu TXT formatuan
- Laburpenen azterketa errazten du
- Atzeraelkadura automatikoa eskaintzeko, informazioa ematen digu automatikoki

Bildutako corpusa
1654 estrakzio eta abstrakzio guztira

Atzeraelikadurarik gabe
Lehen Hezkuntza: eragile geologikoak
7 testu -> 31+74+28+49+48+30+44
Lehen Hezkuntza: elikagaiak
4 testu -> 102+97 +71+82
Unibertsitatea: elikagaiak
4 testu-> 193+184+176+169
Atzeraelikadurarekin
Unibertsitatea: multidokumentuak
2 testu -> 217+79

ixa
Universidad del País Vasco Euskal Herriko Unibertsitatea
HiTZ Hizkuntza Teknologiako Zentroa Basque Center for Language Technology
CLARIAH-EUS