

**Euskararen Prozesamendurako Erreferentziazko Corpusaren (EPEC)  
etiketatzea, segmentazio mailan.**

**Eskuliburua**

**Aldezabal I., Ceberio K., Esparza I., Estarrona A.,  
Etxeberria J., Izagirre E., Quintian M., Uria L.**

**2006ko ekainaren 2a**

## **Laburpena**

**Eskuliburu honetan EPEC corpusa segmentazio mailan etiketatzeko gidalerroak biltzen dira. Une honetan etiketatzeko prozesua amaitutzat ematen dugu, baina inoiz berriz ekinez gero, erabaki berriak hartu ahala, dokumentua eguneratzen joango da.**

## **Abstract**

**The guidelines for the tagging of the EPEC corpus at the segmentation level are set up in this manual. The tagging process is considered finished at the moment. In case of starting working on it again, the manual will be updated as new decisions are taken.**

# Aurkibidea

1. Sarrera.....	4
2. Zenbait oinarritzko kontzeptu.....	5
3. Etiketatzeko erabiltzen diren baliabideen azalpena.....	10
EULIA.....	10
1.- Sarrera.....	10
2. Corputa nola etiketatu.....	10
2.1 EULIAra sartu.....	10
2.2. Fitxategiak ireki.....	10
2.3. Fitxategia etiketatu.....	11
2.4. Euliatik irten.....	12
3.- Analisi berriak nola sortu.....	12
EDBL, diakritikoak, Anali eta kategoria-sistema.....	21
Haul-en zerrenda, Postposizioen zerrenda eta Entitateak.....	23
Oinarritzko bibliografia.....	24
4. Kasu arraroak nola etiketatuko ditugun, eta zer ez dugun etiketatuko.....	24
5. Etiketatzeko oinarriak eta irizpideak/erabakiak.....	26
Oinarriak.....	26
Irizpideak/Erabakiak.....	28
Irizpide orokorrak.....	28
0Hitz-forma bera, segmentazio desberdina (aurreko, ‘aurre + ko’).....	28
0Kategoria-bikoteak.....	29
0Izen berezi eta Leku-izenak.....	30
0DET DZH / IZE ZKI.....	31
0Puntua.....	32
0Laburtzapenak: siglak, laburdurak, sinboloak.....	32
0-ko DESK, -ko BNK, -ko GEL (ikus Euskaltzaindiaren 34. araua).....	32
0Elipsia.....	33
0Parentesi artekua.....	34
0Entitateak (maiuskulaz idatziriko hainbat izen).....	34
0Hitz arrotzak/zaharrak/arraroak/ezezagunak.....	35
0Anbigutasuna .....	35
0Metahizkuntza .....	36
0Liburuen, testuen... izenburuak, eta titularrak, atalak, azpiatalak... ..	37
0Aldaeretan bi balizko aukera baldin badaude.....	37
0Determinatzaile konplexuak.....	37
0Partizipio hutsarekin egiten diren erlatiboak.....	37
Kasu partikularrak.....	38

# 1. Sarrera

Lan honen helburua da 300.000 hitz inguru dituen EPEC deituriko corpora segmentazio mailan etiketatzeko argibideak eskaintzea. Zehazki, ondorengoak:

- Zenbait oinarritzko kontzeptu
- Etiketatze erabiltzen diren baliabideen azalpena:
  - EULIA, etiketatzeko tresna
  - EDBL (Euskararen Datu-Base Lexikala), DIAKRITIKOAK, ANALI (analizatzaile morfoloikoa), eta KATEGORIA-SISTEMA
  - HAULen zerrendak, POSTPOSIZIOen zerrendak eta ENTITATEAK (EDBLn eta, beraz, corpusean markatuta ez dauden bitartean)
  - OINARRIZKO BIBLIOGRAFIA. Batez ere Euskaltzaindiaren gramatikak eta arauak
- Kasu arraroak nola etiketatuko ditugun. Bestela esanda, zer ez dugun Cz markatuko.
- Etiketatze oinarriak eta irizpideak/erabakiak.

“Etiketatzeari” esaten dugunean, hauxe da berez esan nahi duguna: testu-hitz bakoitzarentzat ahal dela segmentazio-mailan *analisi egokia* (C: correct marka jarrita) edukitzea. Esan behar da Eulia tresnaren bidez etiketatzerakoan berez ez dugula “C” marka jartzen, baizik eta “√”. Ikur hau gero automatikoki “C” bihurtzen da testuan. 4. puntuaren ikusiko dugun bezala, batzuetan ezinezkoa da testu-hitz bati C marka jartzea.

Oinarritzko kontzeptuetan zehaztuko dugunaren arabera, analisi egokia esleitzea hauxe litzateke (testuan aurkitutako kasuistika kontuan hartuta):

- LEMATIZATZEA (LEMAZ edo LEMA KANONIKOZ HORNITZEA),
- ANALISI MORFOSINTAKTIKO EGOKIA EMATEA
- ZUZENTZEA

Estandarizatzea ez da helburu nagusia. Estandarizazioak lekua du, lehian dauden bi hitzak elkarren artean aldaera direnean eta EDBLk horri buruzko informazioa eskaintzen duenean.

Bestetik, kontuan izan behar da testu-hitza (*forma*) ez dugula inoiz aldatuko. Sarrera-, lema- eta morfema-eremuak bakarrik ukituko dira.

Etiketatu bitartean, hainbat ondorio eta emaitza linguistiko atera daitezke. Horiek, momentuz, ez ditugu lan honetan landuko. Bigarren pausu batean aztertzeke utzi ditugu.

## 2. Zenbait oinarritzko kontzeptu

### Zer da EPEC?

“Euskararen Prozesamendurako Ereduzko Corpusa”ren sigla da. 300.000 hitz inguruz osatuta dago. Heren bat *XX. mendeko euskararen corpusetik* hartuta dago, eta gainontzeko bi herenak *Euskaldunon Egunkaririk*

### Zer da segmentazio mailan etiketatzea?

Testuko hitza dagokion lema-, kategoria-, azpikategoria-, kasu- eta numero-informazioaz hornitzea. Hots, testuko hitz ororentzat ahal dela lematizazio eta analisi egokia izatea.

Beste hitz batzuetan esanda: testuko hitz bakoitzak C (correct) marka jarrita edukitzea. 4. puntuari ikusiko dugun bezala, batzuetan ezinezkoa da testu-hitz bati C marka jartzea. Eulia tresnaren bidez berez ez dugu “C” marka jartzen, “√” baizik. Ikur hau gero automatikoki “C” bihurtzen da testuan.

### Zer da lema?

Hitza atzizki flexiborik gabe, bere kategoria eta azpikategoriarekin; hiztegietan sarrera gisa dagoen hori. Adib.:

“eritziarekin” formaren lema: eritzi IZE ARR

“iritziarekin” formaren lema: iritzi IZE ARR

### Zer da analisisa?

“Segmentazio-mailako” etiketate-fase honetan, hitzaren analisizat hitz horrek duen flexioa joko dugu, kasua eta numeroa kontuan hartuta. Ikus daitekeenez, funtzio sintaktikoa ez da kontuan hartzen (har bazitekeen ere).

### Zer da lema kanonikoa? Zer da lematizatzea?

Askotan gertatzen da “jatorri etimologiko” beretik bi hitz (edo gehiago) etortzea (bi lema “lehiakide” edo aldaera egotea). Horrelakoetan, forma desberdin guztientzat lema komun bat hautatzen da. Lema komun hori lema kanonikoa deritzo. Adib.:

eritziarekin

eritziarekin -----> lema kanonikoa: “iritzi”

eritziarekin

Hitz batek beste hitz lehiakiderik ez duenean, lema eta lema kanonikoa gauza bera dira.

Testu bat lematizatzea, gehienetan lema kanonikoz lematizatzea esan nahi da. Hala ere, “lematizatze” kontzeptua zenbait hitzetan anbigua gertatzen da, ez bada zehazten lemaz edo lema kanonikoz ari garen.

### **Noiz hartzen dira hitzak jatorri berekotzat / aldaeratzat?**

Elkarren artean ALDAKETA FONETIKOAK eta TIPOGRAFIKOAK gertatzen direnean.  
Adib:

iritzi // eritzi // eritxi  
kajoi // kaxoi  
arabazozo // araba-zozo  
camping // kanpin  
facsimile // faksimile  
Ueu // UEU  
gizarajo // gizagaixo  
bost // bortz

### **Hortaz, zein aldaera mota desberdintzen dira?**

- Tipografikoak: marratxoa +/-, zuriunea +/-...  
arabazozo // araba-zozo  
Ueu // UEU  
mihi luze // mihi-luze
- Ez-tipografikoak: aldaketa morfofonologikoak, euskalkiak, mailegatze-egokitzapenak...  
iritzi // eritzi // eritxi  
kajoi // kaxoi  
camping // kanpin  
facsimile // faksimile  
-rekin // -rekila

Gehienetan bi aldaera mota hauetan lema kanonikoa estandarra izaten da, eta ez-kanonikoa ez-estandarra. BAINA EZ BETI:

bortz // bost. Biak dira estandarrak (eta ez dago argi euren artetik bat hartu behar den kanonikotzat).

kaxoi // kajoi. Biak dira ez-estandarrak (eta ez dago argi euren artetik bat hartu behar den kanonikotzat).

Etiketatzeari dagokionez, guk lema kanonikoa jarriko dugu baldin eta EDBLtik lotura egina badator. Gainontzean, hitzak berak duena jarriko diogu.

### **Zer dira erratak? Zer da zuzentzea?**

Erratak, idazleak idazterakoan egiten dituen sakatze-akatsak izan ohi dira. Eta EZ DIRA ALDAERAK. Adib.:

UEU // UEu  
esakintzean // eskaintzean

uztarrilean // urtarrilean

Horrelakoetan, orduan, ez dugu “estandarizatzeaz” hitz egingo, baizik eta “zuzentzeaz”. Alegia, hauetan “lema zuzendu” egiten dugu. Zuzentzearekin batera, jakina, lema egokia esleitzen diogu, eta lema hori, normalean, lema kanonikoa izango da. Zuzendu egiten dugunean, azken batean, beti sortu egiten dugu.

Zenbaitetan ez da erraza izango errata eta aldaera bereiztea. Adib. :

Erakaskuntzaren → hasiera batean *irakaskuntzaren* errata dirudi, baina corpusean begiratuta, badira *erakaskuntza* darabilten autoreak

Horrelakoetan, norberaren subjektibotasuna sartzen da jokoan.

### Zer da estandarizatzea?

Estandarizatzea eta Lematizatzea (lema kanonikoa esleitzea) ez dira beti gauza bera.

Ikusi dugun bezala, bi hitz elkarren artean aldaera izan arren, horietako bat ez da beti estandarra. Eta bestela ere gertatzen da. Hots, bi hitzen artean estandar/ez-estandar harremana egoteak ez du esan nahi hitz horiek elkarren artean aldaera direnik; askotan euren lemek ez dute zerikusirik elkarren artean. Adib.:

EZ-ESTANDARRA eta bere ESTANDARRA (lema kanonikoa)

-----  
bar

taberna

EZ-ESTANDARRA eta bere LEMA (kanonikoa)

-----  
bar

bar

ESTANDARRA eta bere LEMA (kanonikoa)

-----  
taberna

taberna

Honenbestez, estandarizatzea da, testu-hitz bat “okerra” bada estandartasunari begira (eta ez errata bati begira) dagokion hitz estandarrekin ordezkatzeta. Zuzentzea, berriz, esan dugun bezala, testu-hitza errata bada, horren ordezeko zuzena ematea da. Adib.:

*iritziarekin* testu-hitza estandarizatzen badugu, *iritziarekinen* analisisa markatuko dugu C gisa.

*barrean* testu hitza estandarizatzen badugu, *tabernanen* analisisa markatuko dugu C gisa.

Ordea:

*eritziarekin* testu-hitza lematizatzen (lema kanonikoz hornitzen) badugu, *iritziarekin* en analisisia markatuko dugu C gisa. Kasu honetan estandarizatzea bezalaxe litzateke.

*barrean* testu-hitza lematizatzen (lema / lema kanonikoz hornitzen) badugu, “bar + ean”en analisisia markatuko dugu C gisa.

*eritziarekin* testu-hitza lematizatzen (lemaz hornitzen) badugu, “eritzi + arekin”en analisisia markatuko dugu C gisa.

*eritziarekin* testu-hitza zuzentzen badugu eta “lema kanonizatzen” badugu, “iritzi + arekin”en analisisia markatuko dugu C gisa.

Lematizatzea, zuzentzea eta estandarizatzea zer diren ikusita, **EPEC-EN HELBURU NAGUSIA, LEMATIZATZEA (LEMAZ edo LEMA KANONIKOZ HORNITZEA), ANALISI EGOKIA EMATEA eta ZUZENTZEA DELA** esango dugu. Nahita jarri dugu “lemaz edo lema kanonikoz”. Izan ere, EPECen testu-hitz batek lema kanonikoaren aukera ematen badu, horrexekin lematizatuko dugu, baina ez badu ematen (nahiz eta guk jakin badagoela beste aldaera bat “kanonikotzat” har daitekeena), ez dugu aldaeraren analisisia sortuko. Alegia, bere horretan utziko dugu. Testu-hitza errata bat bada, aldiz, zuzendu (sortu) egingo dugu, bere lema jarrita. Eta azkenik, hitzak ez badu analisi egokirik (ez duelako sarrera EDBLn eta lexikorik gabe asmatzen ez duelako), sortu egingo dugu dagokion lema kanoniko eta analisiarekin.

Adibidez:

Testuan ‘*esklaberia*’ hitza badugu, eta automatikoki analisi egokirik etortzen ez bazaigu, hitza zuzena denez, dagokion kategoriarekin etiketatuko dugu.

Aldiz, ‘*esakintza*’ agertzen bada, ‘*eskaintza*’ hitzaren errata gisa, zuzendu egingo dugu bere forman ‘*esakintza*’ jarriz eta lema kanonikoan ‘*eskaintza*’.

Azkenik, *Alkiza* hitzaren analisisetan IZE LIB (leku-izen berezi) bezala bakarrik badatorkigu eta gure testuan abizena bada, IZE IZB (pertsoneko izen berezi) sortuko dugu.

Ohar bedi inongo momentutan ez dugula aipatu “forma”. Beti aritu gara “lema” eta “analisiaz”. Hain zuzen ere, **“FORMA” EREMUA EZ DUGU INOIZ ALDATUKO: BETI UTZIKO DUGU TESTUAN AGERI DEN BEZALA.**

## Testu-hitza bakoitzerako dugun informazioa

Eulia tresnak testu-hitza bakoitzeko analisirako ondoko informazioa ematen digu<sup>1</sup>:

A: Testu-hitza EDBLn dago sarrera estandar gisa. Beraz, ez dago proposamen desberdinik.

T0: Testu-hitza EDBLn dago sarrera ez-estandar gisa.

---

<sup>1</sup> ZT corpusa etiketatzerakoan marka berak agertzen dira, baina beren esanahia aldatu egiten da, lehenago Elhuyarrekoek tratatu egiten dutelako, aldaera eta erratena corpusean bertan aldatuz (beti ere originala gordetzen delarik).



T1/T2/Tn: Testu-hitza ez-estandarra da eta ez dago EDBLn landuta. Proposamena ematen da testu-hitzetik abiatu eta hainbat erregela aplikatuta. Proposamen hau EDBLn sarrera estandarra da. Zenbakiak zenbat erregela aplikatu diren adierazten dute (oro har)

G: Testu-hitza ez du ezagutu eta lexikorik gabe asmatzen saiatzen da.

E: Ematen diren analisietatik bat ere ez badagokio testu-hitzari, etiketazaileak sortu egiten du dagokiona. Horrelakoetan hitz honi E marka jartzen zaio automatikoki.

A eta T0 kasuak ez dira arazo segmentazioari begira, normalean asmatu egiten baitu bai leman bai analisisian. T zenbaki baxudunak normalean ez dira arazo, baina batzuetan huts egiten dute (ikus beheko adibidea: Biarnesetik). Hortik aurrera, proposamena susmagarria da testu-hitzarekiko (lema, eta analisisa oso desberdina izaten delako normalki), beraz, erreparatu egin behar zaio markatutako analisiari. G kasuak, berriz, beti dira erreparatzekoak.

Adibide batzuk:

TESTU-HITZA	PROPOSAMENA
pertsonaletan	(A) pertsonal ADJ + etan (lema: pertsonal // lema kanonikoa: pertsonal)
zapatila	(A) zapatila IZE ARR (lema: zapatila // lema kanonikoa: zapatila) Hobetsiak = [txapin]
Atlantiar	(T0) atlantiar ADJ ARR (lema: atlantiar // lema kanonikoa: atlantiar) Estandarrak = [atlantiko] Erro-kode = LD_MA
idekitzen	(T0) ireki (ideki) ADI + tzen (lema: ideki // lema kanonikoa: ireki)
artzea	(T1) har (ADI) + tze + a (lema: ar // lema kanonikoa: har)
biarnesetik	(T1) biarnes (ADJ ARR) + eta + tik (lema: biarnes // lema kanonikoa: biarnes)
erakaskuntzan	(G) erakaskuntza (IZE ARR) + an (lema: erakaskuntza // lema kanonikoa: erakaskuntza)
esakintzearen	(G) Hainbat analisi oker
igantenean	(G) Hainbat analisi oker
Zapateroren	(G) Zapatero (IZE IZB) + rekin (lema: Zapatero)

(biarnesetik, eskaintzearen eta igantenean testu-hitzen forma zuzenak hauek liriateke berez: biarnesetik (ABL sing), eskaintzearen (“eskaini” lemarekin), igantenean (“igantene” lemarekin??).

Beraz, gure kezka-iturri nagusiak horiek dira: ez testuko hitza ez proposatutako hitza EDBLn sarrera gisa landuta ez daudenak. Bestela esanda, proposamenak lexikorik gabe edo erregela gehiegi erabiliz sortzen direnean.

TESTU-HITZA ESTANDARRA DENEAN, PROPOSAMENA testuko hitz bera da, eta dagokion lematizazioarekin agertzen da (lema kanonikoa izan ala ez). Adib.: bostetan // bortzetan. Biek bere lematizazio egokia izango dute: bost + etan // bortz + etan), nahiz eta ez den bi lemen artean bat hautatu kanonikotzat.

### 3. Etiketatzeko erabiltzen diren baliabideen azalpena

## EULIA

### 1.- Sarrera

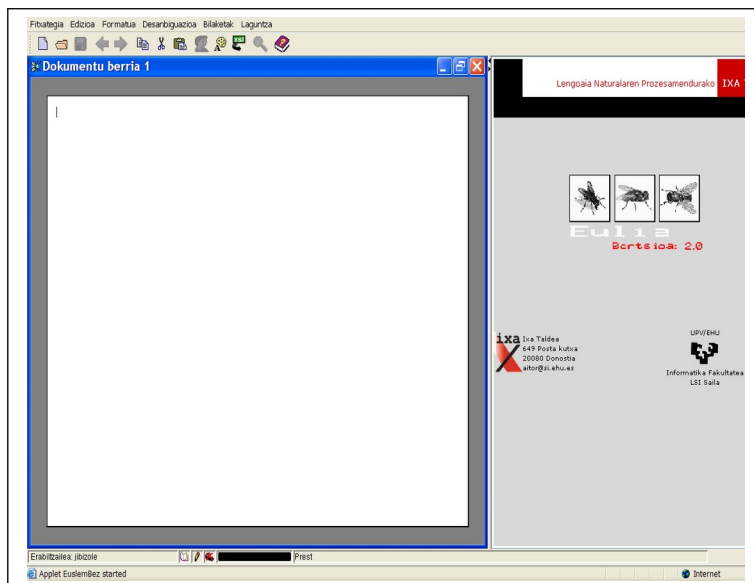
Hizkuntzaren prozesamendurako hainbat tresna baliatuz etiketatu den corpusa tratatzeko ingurunea da EULIA. Corpus hauek sortu, bistaratu, kontsultatu eta aldatzeko erabil daitezke ingurune hau. Eskuliburu honen helburuari heldu eta segmentazio mailan etiketatzeko, EULIAN eman beharreko pausoak azalduko ditugu atal honetan. Bi egiteko bereziko ditugu jarraian; batetik, hitz bakoitzaren etiketatzea bera, eta bestetik, EULIAk proposatzen dituen analisisetan egokirik eman ez, eta dagokiona sortzeko jarraibidea.

### 2. Corpusa nola etiketatu

#### 2.1 EULIAra sartu

EULIA interfazea honako helbide honetan aurkitzen da:  
<http://sisx04.si.ehu.es/eulia/>. Kontuan izan behar da tresna hau bistaratu ahal izateko, ordenagailuak JAVA instalaturik izan behar duela.

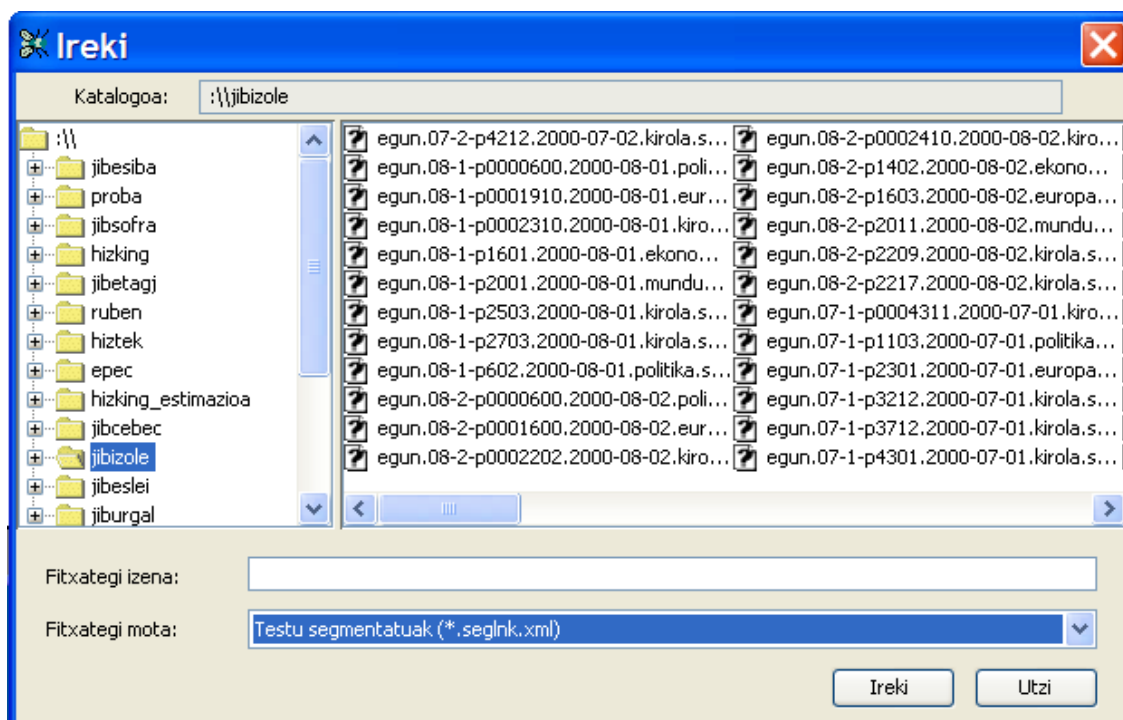
Ingurunera sartu aurretik, erabiltzailea eta pasahitza sartzeko eskatuko digu (etiketatzaile bakoitzak berea izango du). Datu hauek sartu ondoren, EULIAren hasiera-orrira helduko gara:



1. irudia

## 2.2. Fitxategiak ireki

Fitxategiak irekitzeko ikurraren gainean klikatu eta etiketazaile bakoitzeko karpeta bana agertuko da. Etiketazaileak bere karpetaren barruan dituen fitxategiak etiketatu behar ditu, banaka.



2. irudia

Kontuan izan “**Fitxategi mota**” jartzen duen lekuan **Testu segmentatuak (\*.seglnk.xml)** aukeratu behar dela.

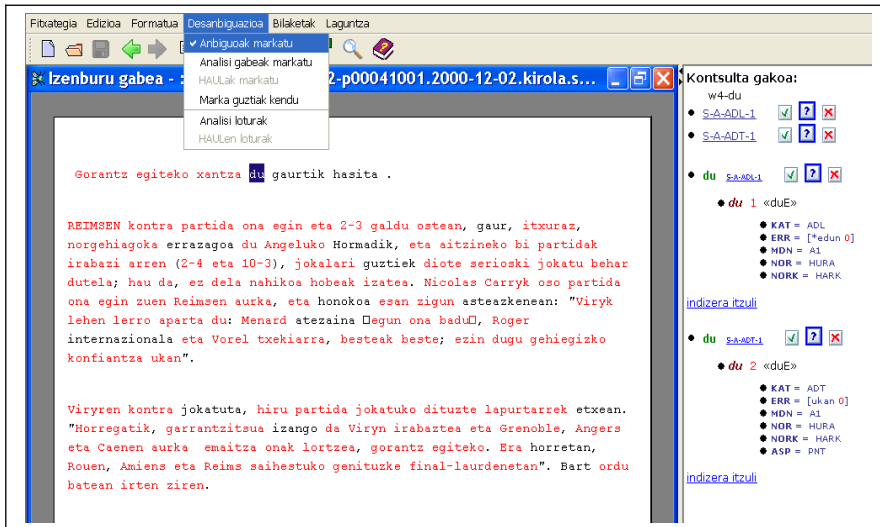
## 2.3. Fitxategia etiketatu

Fitxategi bat aukeratu ostean, corpora ezkerreko leihoan agertuko zaigu, eta aukeraturik dugun hitzaren analisia(k) eskuinekoan. Honela, testu horretako hitzak banan-banan klikatuz joan behar dugu, bakoitzari testuinguru horretan behar duen analisia esleituz.

Dagoeneko etiketaturik dauden hitzak zeintzuk diren jakiteko aukera badago. Tresna-barran “**Desanbiguzioa**”-ren gainean klik egin, eta “**Anbiguoak markatu**” aukeratu behar da. Bi izan daitezke hitz horiek etiketaturik egotearen arrazoiak:

- 1) Etiketazaileak berak lehenago etiketatu izana.
- 2) Hitzak analisi bakarra izatea aukeran. Kasu honetan automatikoki esleitzen zaio etiketa. Dena dela, hauek errebisatzea komenigarria da, izan ere, gerta liteke analisi

hori zuzena ez izatea eta egokia sortu behar izatea (leku-izen eta izen bereziak diren kasuetan batik bat).



### 3. irudia

3. irudiko eskuin leihoari erreparatuta, aukeraturik dugun hitzaren ('du' kasu honetan) analisi posibleak agertzen direla ikus dezakegu. Ematen dituen aukera horietatik, etiketatzaileak testuinguru horri dagokiona aukeratu behar du.





4. irudia

### Analisi posibleen laburpena.

### Analisi bakoitzari dagokion informazio guztia.

Bata edo bestea, bietatik zein aukeratu jakiteko, corpusera jo behar dugu. Kasu honetan **'du'** horrek aditz trinko gisa jotzen du, beraz, bi aukera hauetakoren batean klikatu beharko dugu:

Bai laburpenean, bai informazio zehatzagoa ematen den horretan,  ikurraren gainean klikatu. Honelako bat aukeratzekoan, aukeratu ez ditugune  ikurra jarriko zaie automatikoki.

Behin fitxategia etiketatzen bukatu dugula, egindakoa gorde behar dugu. Dena dela, etiketatu ahala gordetzea komeni da, interfazearekin arazorik bagenu, egindakoa ez galtzeko.

## 2.4. Euliatik irten

Eulia-tik irteteko, lehiotxoak erabili behar dira (Fitxategia/Bukatu), eta ez goiko eskubiko Xa. Hau garrantzitsua da, bestela fitxategia irekita geratzen baita eta memoria okupatzen baitu.

## 3.- Analisi berriak nola sortu

EULIAk aukeran ematen dizkigun analisisetatik egokirik aurkitzen ez badugu, guk geuk sortu beharko dugu. Dena dela, zenbait kasutan, analisi egokia ematen ez bada ere, etiketatzailen artean hartutako irizpideak oso kontuan izan beharko dira, kasu askotan ez baita beharrezkoa izango sortzea, bestelako irtenbide bat eman diogulako<sup>2</sup>.

Bestalde, analisia sortzen hasi aurretik, analizatzera goazen hitzaren informazio guztia biltzea komeni da. Izan ere, sortzeko leihoak irekita ditugula, ez da komeni informazioa bilatzeko beste leihorik irekitzea, gero bueltatzean, EULIAk arazoak izan baititzake leihuetan agertzen den informazio guztia bistartzeko.

Honenbestez, sortzen hasi aurretik *EDBL*, *ANALI*<sup>3</sup>... kontsultatuko dugu behar dugun informazio guztia eskuratzeko. Har dezagun behe-rago sortuko dugun hitza hemen ere adibide gisa: ***EHri***.

Hitz hau, agertuko zaigun atal bakoitzean zein modutan eman behar dugun jakiteko, *ANALI*ra jo behar dugu. *ANALI*n datu-basean egon daitekeen eta aztertzen ari garen formaren

<sup>2</sup> Ikus 5. atala (Etiketatzeko oinarriak eta irizpideak/erabakiak)

<sup>3</sup> Ikus '*EDBL*, diakritikoak, Anali eta kategoria-sistema' atala

antzeko ezaugarriak (kategoria, hitzaren amaiera, letra larria...) dituen hitza idatzi behar dugu.

**Adibidez:** ‘Amir’ izen berezia sortu beharko bagenu, ANALIn ‘Aitor’ izen bereziaren analisia eska dezakegu honako arrazoiengatik:

- 1.- Litekeena da ‘Aitor’ datu-basean sartuta egotea (euskal izen berezi arruntenak bertan baitaude)
- 2.- ‘Aitor’, ‘Amir’ bezala, izen berezia da, ondorioz, jarraitze-klase berdinak izango ditu.
- 3.- ‘Aitor’, ‘Amir’ bezala, ‘r’ kontsonanteaz amaitzen da.

Analizatzen ari garen honetan, ‘EHri’ kasuan alegia, ‘Euskal Herriari’ adierazia duen forma iristen zaigu datu-basetik. Testuinguruari erreparatuta, ‘Euskal Herriarrok’ adierazia behar duenez, analisi berria sortu behar diogu. Baina biek dute funtzionatzeko modu bera, beraz, ANALIn ‘EHri’ idatzi eta bertan agertzen den analisiak balioko digu (adieraziari dagokiona salbu).

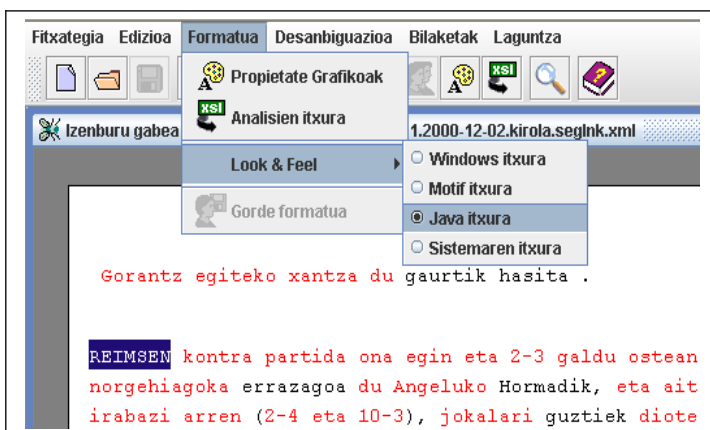
```
A1 9e9h3[[Sarrera_EH--0][KAT_IZE][AZP_LIB][PLU_-][MTKAT_SIG]
[ADZ_[Sarrera_Euskal__Herria--0]]]
+Ri[[Sarrera_i--2][KAT_DEK][KAS_DAT][MUG_MG][FSL_[FS1_@ZOBJ]]]
```

Puntu honetan ohar bat egin behar dugu, ANALIn letra larria adierazteko erabiltzen den ‘9’ zenbakiaren ordeaz, ‘\*’ jarri behar dugu.

Beharrezko informazioa eskuartean dugularik, analisia sortzen has gaitezke. Eta prozesu honetan zehar agertuko zaizkigun leihoetan informazioa sartzeko, ANALIk emandako datuetan oinarrituko gara. Jarraian sortze-prozesua pausoz pauso azalduko dugu:

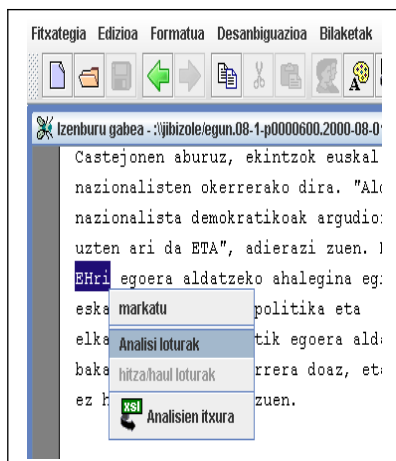
**1.-** Analisiak sortzeko leihoak egikaritzeko, inguruneak ‘**JAVA itxura**’ izan behar du:

**Formatua → Look & Feel → Java itxura**



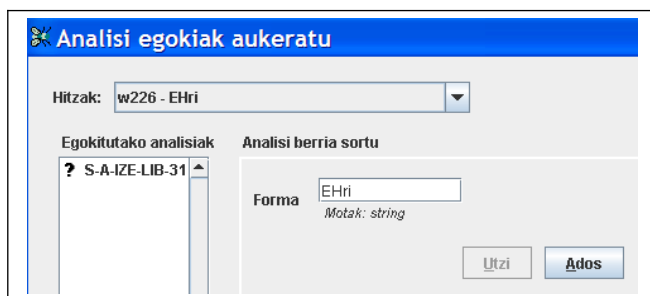
5. irudia

**2.-** Analisi berria sortu behar diogun hitzaren gainean jarri eta saguaren eskuineko botoiari emanda, ‘Analisi loturak’ agertuko da. Bertan klik egin.



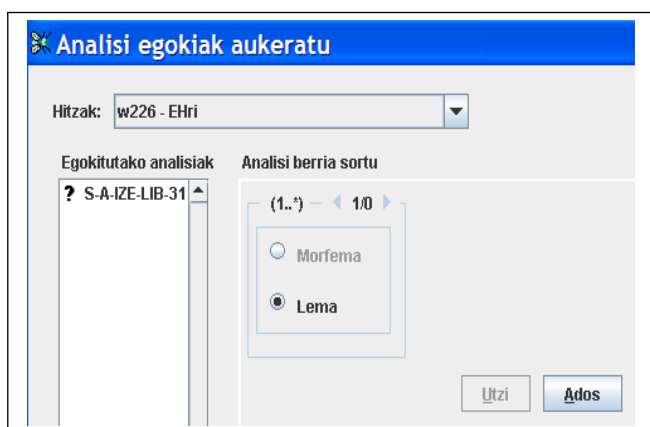
6. irudia

3.- ‘Forma’ jartzen duen horretan, hitza osorik sartu behar dugu, testuan agertzen den moduan (morfemarik badu, hau ere sartu).



7. irudia

4.- Hurrengo pausoa, lema edo morfema, aztertzeraz goazena zehaztu behar dugu. Lehenik, lema hautatuko dugu. Behin lema dagokion informazioa zehaztuta, berriro leiho honetara helduko gara. Hitzaren osagai guztiak definitu arte, kasuan kasuko aukera egin (jarraian beste lema bat edo morfema bat izan, bata edo bestea aukeratu). Hitzaren osagai guztiak definitu ditugunean ‘utzi’ sakatuko dugu.



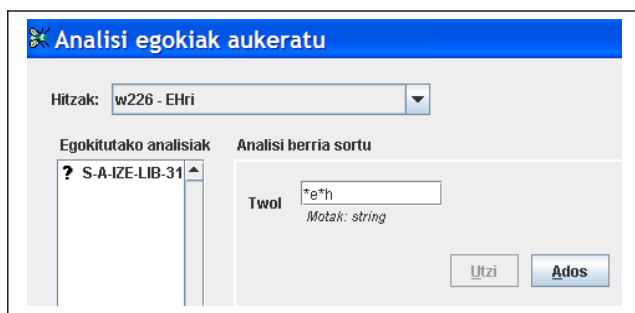
8. irudia

**5.-** ‘Twol’ horri dagokion zuriunean bigarren mailako forma idatzi behar dugu; hau da, XXX-k ezagutuko duen moduan idatzi behar dugu. Kasu honetan, esaterako, letra larria ageri da, eta bigarren mailako formalismoan letra larriak ‘\*’ ikurraren bidez adierazten dira. Osagai bakoitza zein modutan eman behar dugun jakiteko, *ANALIk* ematen digun informazioa jo behar dugu.

```
A1 9e9h3 [[Sarrera_EH--0] [KAT_IZE] [AZP_LIB] [PLU_-] [MTKAT_SIG]
[ADZ_[Sarrera_Euskal__Herria--0]]]
+Ri [[Sarrera_i--2] [KAT_DEK] [KAS_DAT] [MUG_MG] [FSL_[FS1_@ZOBJ]]]
```

Gogoratu, *ANALIn* letra larria adierazteko erabiltzen den ‘9’ zenbakiaren ordean, ‘\*’ jarri behar dugula.

Laugarren pausoa zehaztu dugun bezala, orain lema soilik ari gara zehazten. Morfemari dagokion bi mailako forma (Ri) aurrerago idatziko dugu, lema informazio guztia eman ostean.



9. irudia

**6.-** Sarreran, aztertzen ari garen osagaia (lema edo morfema) idatzi behar dugu, oraingo honetan ez diogu aurreko pausoa bezala bigarren mailako formari erreparatu behar. Lema nahiz morfema zein den zehaztu behar dugu hortaz, eta baita bakoitzari dagokion homografoa ere. Datu-basean ez dagoen lema sartu behar dugunez, homografoa **0** izango da. Baina morfema bakoitzak dagoeneko badu homografo bat eta hau ere *ANALIk* emango digu. Kontuan izan morfemaren numeroa, kasua etab.en arabera homografoa aldatu egiten dela, horregatik, ezinbestekoa da *ANALIn* sartu dugun forma horrek ezaugarri horiek izatea.

```
A1 9e9h3 [[Sarrera_EH--0] [KAT_IZE] [AZP_LIB] [PLU_-] [MTKAT_SIG]
[ADZ_[Sarrera_Euskal__Herria--0]]]
+Ri [[Sarrera_i--2] [KAT_DEK] [KAS_DAT] [MUG_MG] [FSL_[FS1_@ZOBJ]]]
```



10. irudia

7.- Hurrengo pausoa aldaera sartzera eskatuko zaigu, baina guk sortuko ditugun hauetan ez dugu hau zehaztuko. Beraz, zuzenean **'utzi'**-ren gainean egin klik.

11. irudia

8.- Ondoren, zein kategoria duen aukeratu behar dugu.

```
A1 9e9h3[[Sarrera_EH--0][KAT_IZE][AZP_LIB][PLU_-][MTKAT_SIG]
[ADZ_[Sarrera_Euskal_Herria--0]]]
+Ri[[Sarrera_i--2][KAT_DEK][KAS_DAT][MUG_MG][FSL_[FS1_@ZOBJ]]]
```

Lemari IZE dagokio, eta DEK morfemari.

12. irudia

**9.-** Aurreko pausoen aukeratutako kategoriaren arabera, informazio desberdina eskatuko zaigu orain. Aztertzen ari garen adibidean, esaterako, lema IZE markatu dugunez, honako aukera hauek ematen dizkigu:

**14. irudia**

Estandarrak (0..\*)

Sarrera   
*Motak: string*

Homografo-Id   
*Motak: nonNegativeInteger*

Jarraitu

**15. irudia**

Forma Flexionatua  
 Hitz-elkartua  
 Hitz-eratarria  
 Laburdura-Sigla

Utzi Ados

**16. irudia**



**13. irudia**

Hobetsiak (0..\*)

Sarrera   
*Motak: string*

Homografo-Id   
*Motak: nonNegativeInteger*

Jarraitu

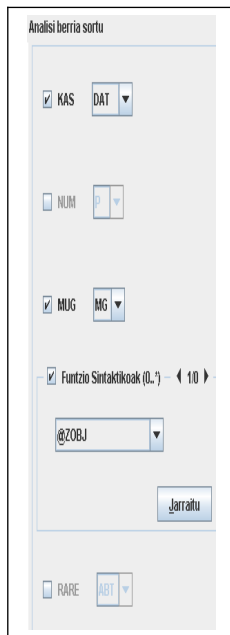


Lema aztertzen ari bagara, atal honi dagokion informazio hau aukeratu behar dugu (ikus 13-16 irudiak):

```
A1 9e9h3 [[Sarrera_EH--0] [KAT_IZE] [AZP_LIB] [PLU_-] [MTKAT_SIG]
[ADZ_[Sarrera_Euskal__Herria--0]]]
+Ri [[Sarrera_i--2] [KAT_DEK] [KAS_DAT] [MUG_MG] [FSL_[FS1_@ZOBJ]]]
```

Morfemaren kasuan, berriz (ikus 17.irudia):

```
A1 9e9h3 [[Sarrera_EH--0] [KAT_IZE] [AZP_LIB] [PLU_-] [MTKAT_SIG]
[ADZ_[Sarrera_Euskal__Herria--0]]]
+Ri [[Sarrera_i--2] [KAT_DEK] [KAS_DAT] [MUG_MG] [FSL_[FS1_@ZOBJ]]]
```



17. irudia

**10.-** Aurreko pausoan *laburdura-sigla* hautatu izanak hurrengo leiho honetara eramango gaitu:

18. irudia

‘**Sarrera**’ dioen atalean, siglaren adierazia sartu behar dugu, hitzen arteko tartekak ‘\_’ ikurraren bidez adieraziz. Hemen ere, datu-basean agertzen ez den lema izaki, homografoa **0** izango da.

**11.-** Lehen prozesu honetan lema definitu dugu, baina beharbada aztertzen ari garen hitzak morfema edo beste lemaren bat darama jarraian. Horregatik, berriro ere leiho honetara iritsiko gara, behar izanez gero, gainerako osagaiak definitu ahal izateko.

Adibide gisa hartu dugun honetan, morfemaren azterketa egin beharko dugu. Beraz, ‘morfema’ hautatu eta ‘ados’ botoiaren gainean klik egingo dugu. Berriro ere azaldu ditugun aurreko pausoetara itzuliz (5. pausotik aurrerakoak).

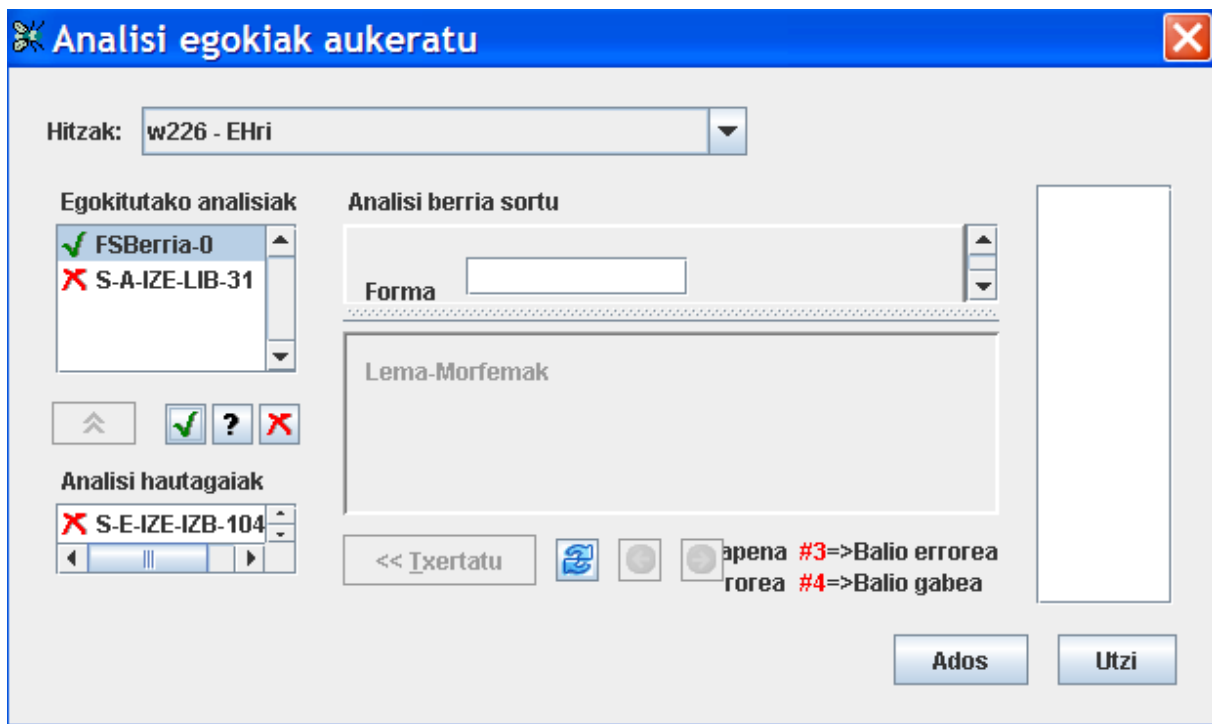
Hitzaren osagai guztien analisia egin badugu, ‘**utzi**’-ri eman beharko diogu, eta 12.pausoan azaltzen den lekura iritsiko gara.

**12.-** Heldu gara azken leihora. Aurreko urratsean esan bezala, ‘**utzi**’ klikatu eta honako leihora iritsiko gara. Sortu dugun analisia txertatzeko hiru botoi sakatu behar ditugu:

1.- ‘**Txertatu**’ lehenengo.

2.-  ikurraren gainean ondoren, proposatzen dituenen artean analisi berria aukera dezan.

3.- Azkenik, ‘**Ados**’ klikatu.



## EDBL, diakritikoak, Anali eta kategoria-sistema

### EDBL

Hitzen segmentazio-analisiak EDBLko bertsio bati dagozkio. Etiketatzeko prozesu hau 4.4.9 bertsioarekin egina dago, beraz, edozein kontsulta egiterakoan, hori izan behar dugu kontuan. Hala ere, EDBL etengabe eguneratzen da. Horrek esan nahi du EDBL zuzenean kontsultatzen dugunean, ez gaudela inongo bertsiotan, baizik eta momentu horretara arte egin diren azken aldaketak ikusten ditugula. EDBL-TIK EGITEN DIREN DATU-ESPORTAZIOAK DIRA BERTSIO BATI LOTUTA DAUDENAK. Hori horrela izanda, ANALI erabili behar dugu, bertan bertsioa zehaztu behar baita. EDBL zuzenean kontsultatu beharko dugu segmentazio-

analisian zer edo zer gaizki dagoela iruditu eta EDBLn aldatu beharko litzatekeela uste dugunean, ziurtatzeko jada horrela dagoen ala ez.

EDBL kontsultatzeko helbideak:

<http://ixa2.si.ehu.es/edbl/>

<http://ixa2.si.ehu.es/edblkontsulta>

## DIAKRITIKOAK

Diakritikoak hauexek dira:

\* maiuskula adierazteko. Adib. \*kepa Xeroxen 9 bihurtzen da.

A a itsasia gehienetan. Adib. \*donostiA

# a itsatsia 6 kasu berezietan. Adib. literatur#. Xeroxen 2 bihurtzen da.

& a itsatsi berezia tokietan. Adib. \*azpeiti&

@ bukaerako a aditz jokatueta. Adib. d@

a berezia erroan. Adib. ater@

E e epentetikoa. Adib. konpletiboa: +Ela

N n aditzen bukaeran. Adib. egiN

M n berezia deklinabidean. Adib. areM (genitibo mugatu singularra)

\ n berezia genitiboan ekin etab. egiteko. Adib. +e\

R r gogorra lemetan. Adib. zakuR

r epentetikoa atzizkietan. Adib. +Ri (datibo mugagabea)

~ r zaharra zenbakietan (hiru lau). Adib. hiru~

Q r berezia izenordainetan. Adib. haQ

^ hikaren eta 1. pertsonaren deklinabidea kontrolatzeko: Adib. duk^, dut^

%: deklinabide bereziko tokiak. Adib. \*usurbil%, \*gasteiz%. Xeroxen 1

: letra solteetan eta laburduretan erabilia. Adib. kg: Xeroxen 3

/ sigletan erabilia bokal/konts moduan ibiltzeko. Adib. \*e\*d\*b\*/

{ atziki lexikaletan. Adib. +{garri. ! bihurtzen da eta 4 Xeroxen

! atziki lexikaletan

arazi !taraz → "egitarazi" debekatzeko.

garren garren! → "hirugarrengo" debekatzeko.

## ANALI

Anali erabiltzeko pausuak:

```
sisx04
% setenv VERSION 4.4.9
% echo $VERSION
4.4.9
%anali
```

EDBL (eta beraz Anali) eraikitzeko, Euskaltzaindiaren gramatiketan oinarrituriko kategoria-sistema sortu zen, gure beharretara egokituta. Kategoria-sistema hori EDBL azpitaldearen dokumentuen artean dago:

<http://ixa.si.ehu.es/Ixa/Azpitaldeak/EDBL/dokumentuak/EDBL.kategoriak>

## Haul-en zerrenda, Postposizioen zerrenda eta Entitateak

Segmentazio mailan, hitz<sup>4</sup> bat baino gehiagoko unitate lexikalak (HAULak, POSTPOSIZIOAK, ENTITATE BATZUK) hitzez hitz (eta ez osorik) etiketatu behar dira. Horrek esan nahi du horiek nola landuta dauden / lantzen diren jakin behar dugula, gero etorkizun batean horiek EDBLn nahiz corpuseko analisisetan integratuta daudenean, analisi bera edukitzeko.

**HAULei dagokienez**, kontuan hartu behar da HAULen zerrenda ez dela behin betikotzat hartzen. Baina behintzat gaur egun dauden bezala analizatu behar direla.

HAULei buruzko edozein zalantza dugunean (behin zerrenda kontsultatuta, zalantzarik ez edukitzea, edo ez bururatzea izango da normalena), lehenik eta behin, HAULen zerrendan ageri dena azken erabakia dela ziurtatu behar dugu. Horretarako, Rubeni galdetuko diogu zuzenean. Gero berak aldaketak egingo ditu (hala uste badu), eta HAULen zerrenda eguneratuko du. HALA ERE, gerta daiteke aldaketa ez eguneratzea berehala zerrendan. Beraz, egokiena da, EPEC etiketatzearen ondorioz egiten diren aldaketa guztiak norberak apuntatzea ondorengo testuetan kontuan hartzeko.

### HAULen zerrendak:

[http://ixa.si.ehu.es/Ixa/Azpitaldeak/HAUL%20taldea/dokumentuak/HAUL\\_guztiak\\_gauzatze-eskemekin.txt](http://ixa.si.ehu.es/Ixa/Azpitaldeak/HAUL%20taldea/dokumentuak/HAUL_guztiak_gauzatze-eskemekin.txt)

**Postposizioei dagokienez**, postposizioei buruzko barne-txosten bat badago, oraindik bukatzeko ukitu batzuk (segmentaziorako nahiko garrantzitsuak) falta dituena. Bien bitartean, behin-behineko bertsio hori erabiliko dugu ahal dugun neurrian:

[http://ixa.si.ehu.es/Ixa/Taldearen\\_kudeaketa/Bileren\\_aktak/1068720449/Postposizioak%20taldeka](http://ixa.si.ehu.es/Ixa/Taldearen_kudeaketa/Bileren_aktak/1068720449/Postposizioak%20taldeka)

---

<sup>4</sup> Hitza, bi zuriguneren arteko karaktere-segida gisa hartuta.



**Entitateei dagokienez**, kontuan izan behar da *entitate* terminoan maiuskulaz hasten diren mota askotako hitzak sartzen ditugula: izen bereziak (pertsona-, leku- nahiz entitate-izenak), mota guztietako siglak (letra batekoak edo gehiagokoak)...

Osagai bakarreko entitate asko EDBLn bertan landuta daude; gehien bat pertsona-izen/abizenak, leku-izenak eta erakunde-izenak, baina batez ere euskarazkoak. Beste hizkuntzetakoak urriagoak dira, ez baitago argi horiek non/nola tratatu behar diren. Osagai bat baino gehiagokoak, berriz, gutxi daude datu-basean, eta gainera, HAUL lirarteke.

Kontua da Entitateen tratamendua beti egiten dela *a posteriori*, eta horretarako EDBLtik ematen den informazioa eta lexikorik gabeko moduluaren bidez sortzen dena hartzen dira kontuan. Horren ondorioz, osagai bat baino gehiagoko entitateak entitate oso bezala analizatuko dira bukaeran. Beraz, azken batean, entitatea den hitz baten analisia hautatzerakoan, oso gogoan izan behar dugu (HAUL eta POSTPOSIZIOEKIN bezalaxe), zertan (zein ezaugarritan) oinarritzen diren entitateak detektatzeko. Horrenbestez, edozein zalantza izanez gero, lehenik Izaskun F.ri galdetu behar zaio; askotan alferrikako zalantzak ekidingo dira-eta.

## Oinarrizko bibliografia

Badira gauza asko linguistikoki erabakita daudenak, hizkuntzalaritza mailan nahiz IXAn.

Dagoeneko erabakita daudenak Euskaltzaindiaren arau (<http://www.euskaltzaindia.net/arauak/>) eta gramatiketan (EGLU I, EGLU II, EGLab, EGLU III, EGLU IV, EGLU V, EGLU VI) topatutakoak dira gehienbat (hiztegi kontuetarako Elhuyar, UZEI eta abar ere asko baliatu dira). Hau da, bertan esaten direnak konputazionalki tratagarri diren modu batean daude jarrita, ahal den neurrian. Hala ere, esan behar da EGLU IVtik aurrerakoak ez direla sakonki aztertu, eta gerta daitekeela berez erabakita dagoen zerbait guk ez kontuan hartu izana. Beraz, corpora etiketatzerakoan edozein zalantza dugunean, lehenik eta behin saiatu behar gara ziurtatzen corpusaren analisiak ematen diguna Euskaltzaindiak esana duen zerbait den ala ez, eta gure zalantzek liburu horietan erantzunik baduten ala ez.

Bestalde, badira gauza asko Euskaltzaindiak ez dituenak oso esplizitu adierazi eta guk geuk nolabait jarri behar izan ditugunak.

Eta azkenik badira gauza asko ez daudenak inon esanda ez eta erabakita, eta guk geuk erabaki behar izan ditugunak.

Hori horrela izanda, 5. atalean (etiketatzeko irizpideak eta erabakiak) nabarmentzea merezi duten erabakiak jarri dira. Baten batek beste zerbait jartzea beharrezkoa dela uste badu, jarri egingo da.

Atal horretantxe joango gara sartzen EPEC landu ahala ateratzen zaizkigun irizpide berriak. Horrexegatik, berau izango da gehien aldatzen joango dena dokumentu elektronikoa honetan.

## 4. Kasu arraroak nola etiketatuko ditugun, eta zer ez dugun etiketatuko

- Errore sintaktikoak dauden esaldietan, elementu guztiak etiketatuko ditugu. Adib.

### **komunztadura erroreak:**

- \*Faboritoak asko zuten jokoan
- \*Kazetariak esan dute
- \*Mutilak lortu behar dute

Elementu bakoitzari dagokion analisia emango diegu, eta elementu erroredunen kasuan, oro har, garrantzia numeroari emango diogu; horrela, adibide hauetan subjektuari ABS PLUR etiketa emanaz (errorea ERGatiboa gaizki erabiltzea litzateke-eta).

- **Tokenizazio-kontuak:**

- Elkartutako hitzak: “astelenerarte”, “udaberrioro”, “46.metro” bezalako kasuetan, analisi berri bat sortuko dugu lema bakar gisa, dagokion analisia esleituz (“astelenera\_arte” ADB, “udaberri\_oro” ADB, “46.\_metro” IZE ARR). Adibide gehiago: “zuhaitzik-zuhaitz” (“zuhaitzik\_zuhaitz” ADB ARR ADLG), “urratsez-urrats” (“urratsez\_urrats” ADB ARR ADLG), “tarte-marteka” (“tarte\_marteka” ADB ARR ADLG)....
- Bestelako kasuak, (“gutxi gora behera”, “EPEC- ko”, “2000. garren”, “Eugi I.a-Olaizola, “ahal bait”...), ez ditugu etiketatuko; bere horretan utziko ditugu gero informatikariek, eskuz, ‘konpon’ ditzaten;

- **Errore lexikalak**, semantika ere sartzen denean. Adibidez: “isla”, “boza”, “karrera”... agertu zaizkigu corpusean, baina proposatutako analisietan adiera ezberdina dute (“isla = uharte”, “boza = poza”, “karrera = ikasketak”). Honelako kasuetan, hitzari dagokion adierarekin lema berria sortuko dugu (“isla = isla”, “boza = ahotsa, soinua”, “karrera = lasterketa”).

- **Aldaerak:**

EDBLn ez-estandarrek aldaerarekin lotura baldin badu, hitz horri dagozkion agerpen guztiak aldaera berarekin lotuko ditugu. Aldiz, EDBLn loturarik ez badago, nahiz eta badakigu aldaerarik baduela, ez dugu loturarik egingo, baina bai markatuko dugu + aldaera gisa. Adib. (*huan* ez-estandarra *huen* estandarra, baina EDBLn loturarik ez; beraz, *huan* “aldaera:+”. Batzuetan, estandarra zein den ere ez dakigu: dine, zaitan...). (Ikus EULIA tresnaren azalpena).

Kontuan izan momentuz ez dugula aukerarik erratatzat jo ditugunak markatzeko edo identifikatzeko (esakintzearen → errata; zuzendu eta kitto). Agian, etorkizunari begira, beste izen orokorrigo bat erabili beharko litzateke, edo bi eremu (aldaera +/-; errata: +/-) sortu, biak identifikatu edo desberdintzeko.

- **Ikurrak (+) (-)...** BST etiketa esleituko diegu; analisirik ez badute, noski, bere horretan utzi.

- **Zerrendatzeak:**

- 1) 2) 3) ... IZE ZKI
- 1. 2. 3. ... DET ABS
- A B C ... IZE ARR ABS MG

- **n solteak** utzi egingo ditugu.
- **Testuinguru sintaktikorik gabeko** elementuak:
  - **Zenbakiak sailkapenetan, puntuazioetan...** DET
  - **Kirolarien izenak...** IZB bezala sortu
  - **Kirol taldeak...** LIB
- **Erakunde** eta **pertsona izenak** berdinak diren kasuetan, ez ditugu ezberdindu. Adib.: Polizia.
- **Beste** hizkuntzatan dauden hitzak, IZE ARR. Adib.: “Avisu la exortacionea provetchosac bekhatorentz; at hitz lauz”, “Devoten Breviarioa”... Elementu guztiak IZE ARR eta azken elementua, deklinatuta badago, dagokion analisisa eta bestela ABS MG.
- “?” ikurrarekin utziko ditugu inola ere etiketatu ezin diren elementuak.

**OHARRA:** Etiketatu ezin dugunaren berri Butzilla erabiliz emango dugu, Morfologia (IEPA) taldeari zuzenduaz. Bertan zehaztu beharko ditugu: fitxategiaren izena eta hitzaren zenbakia.

## 5. Etiketatze oinarriak eta irizpideak/erabakiak

### Oinarriak

Etiketatzeko-lanean, zenbait oinarri linguistiko suposatuzat ematen dira (kategoria zeintzuk diren, zer den sintagma bat, zer den hitza, zer den kasua, zer den mugatasuna, zer den hitz eratorri bat, zer den hitz konposatu bat, zer den aposizio bat...). Hala ere, segmentazio mailan bereziki garrantzitsuak diren zenbait gauza aipatzea komeni da.

Hauexek dira ideiarik inportanteenak:

- Sintagma batean mugatasun morfologiko bakarra dugu (mugatasun lexikala beste gauza bat da, morfosintaxi-txostenean ageri den bezala ([http://ixa.si.ehu.es/Ixa/Argitalpenak/Barne\\_txostenak/1001000785/publikoak/TR-1-99.ps](http://ixa.si.ehu.es/Ixa/Argitalpenak/Barne_txostenak/1001000785/publikoak/TR-1-99.ps))).
- Mugatasun hori sintagma-bukaeran doa.

Adibidez, sintagma honetan ("bi alderdi abertzaleek") hauxe genuke:

bi DET DZH  
 alderdi IZE ARR  
 abertzale ADJ ARR MUG P ERG

Alegia, "alderdi" ez da MG, baizik eta EZER EZ. Beste era batera esanda NULL.

Salbuespen modukoak dira “hauek guztiak”, “nik neuk”, “Peruk berak” eta horrelako sintagmak, indargarri gisa-edo erabiltzen direnak. Horietan, jakina, bi osagai jarriko zaie kasua, mugatasuna eta numeroa, analizatzaileak ematen duenari jarraituz, alegia.

Koordinazioan gauzak ez dira hain sistematikoak. Kasuistika handiagoa dago (ikus Euskaltzaindiaren 111. araua), baina askotan analisiak berak adieraziko digu zer dagoen mugatua eta zer ez (Beste kontu bat da koordinazioa gaizki egina egotea).

Hori horrela izanda, hona hemen testua etiketatzerakoan agertu diren adibide argigarri batzuk:

Bilbo hiria → IZE LIB

720 legea → IZE ZKI

Mikel Arregi<sub>k</sub>, ikastolako zuzendari<sub>k</sub> → biak ERG

Mikel Arregi, ikastolako zuzendaria, → ABS MG + ABS M SING

Mikel Arregi ikastolako zuzendaria, → IZE IZB + ABS M SING

Mikel, nire anaia, → ABS MG + ABS M SING

aldaketa hauen guztien motorra zein da? → biak GEN

hitz eman IZE ARR ABS MG

bide eman IZE ARR ABS MG

aurre egin IZE ARR ABS MG

hitz egin IZE ARR ABS MG

gai izan ADJ ARR ABS MG

Jose Migel Beñaran da... IZE IZB + IZE IZB + IZE IZB ABS MG

Bartolome Leonardo da... IZE IZB + IZE IZB ABS MG

bi aulki → IZE ARR ABS MG

bi aulki → DET DZH

sinbolo bezala → IZE ARR<sup>5</sup> + ADB

Nereak bezala → IZE IZB M SING ERG + ADB

adierazten duteN bezala → ADOIN EZBU + ADL MOS + ADB

ahaztu gabe → PART + ADB

zattia salbu → IZE ARR M SING ABS + ADB

sinadura bidez → IZE ARR + IZE

zenbait gauza daude... → IZE ARR MG ABS

gauza bat → IZE ARR

etxe bat → IZE ARR

hainbat etxe → IZE ARR ABS MG (zero mugagabea)

EH apalari → IZE LIB SIG + ADJ DAT

Mikel jotzeko → ABS MG

Bat Barkoxen, bestea → DET DZH MG ABS +

Bata eta bestea → DET DZH M SING + DET DZG M SING

elektrizitate mota bi daude... → ABS MG

uretan, sutan, eguzkitan... (uretan disolbatu, eguzkitan ibili...) → IZE ARR MG INE

zenbait hizkuntzekin arazoak ditu // hizkuntzekin arazoak ditu

IZE ARR MG SOZ

IZE ARR M P

Euskara dute ama hizkuntza / Ama hizkuntza euskara dute

MG ABS

M S ABS

M S ABS MG ABS

ikastola eta kiroldegian → IZE ARR + IZE ARR M SING INE

ikastolan eta kiroldegian → IZE ARR M SING INE + IZE ARR M SING INE

beroa eta tenperatura → IZE ARR M SING ABS + IZE ARR M SING ABS

bero eta tenperatura → IZE ARR + IZE ARR M SING ABS

neke edo akidura handia → IZE ARR + IZE ARR + ADJ ARR M SING ABS

letra edo beste kreditu-abal batean → IZE ARR + IZE ARR + DET DZH M SING  
 INE

korronte elektrikoa → ADJ ARR M SING ABS  
 Valentziako artzapezpiku eta bikario nagusi, Gerako apezpiku (1952), Ferrolgo apezpi  
ku  
 (1957) eta Donostiako apezpiku izan zen (1968-78). → denak IZE ARR eta  
 azkena IZE ARR M ABS

bizi eta ezkututzen dira,  
 ADJ PART  
 MG ABS EZ-BURU  
 argazkiak errebelatzen eta saltzen diren lekua  
 PART PART  
 EZ-BURU EZ-BURU

apaintzeko eta edertzeko arropa  
 ADIZE ADIZE  
 M S GEL M S GEL

Gainontzekoan, lehenengo puntuan zehaztutakoak ere oinarriak izango dira

## Irizpideak/Erabakiak

Atal honetan bi azpiatal bereizi ditugu. Batetik *irizpide orokorrak* eta bestetik *kasu partikularrak*.

Batzuetan zail izango da irizpide bat azpiatal batekoa edo bestekoa den zehaztea; edota hasieran partikularra dena gero orokorra izatera pasako da. Hori, beti ere, kasuistikaren araberakoa izango da.

## Irizpide orokorrak

### Hitz-forma bera, segmentazio desberdina (*aurreko*, ‘aurre + ko’)

Hitz-forma bera lexikalizatua (*aurreko*; *hitz-erdi*; *arretaz*, *ikusgarri*...) edo morfemaz/lemaz osatua (‘aurre + ko’; ‘hitz + - + erdi’; ‘arreta + z’; *ikus + garri*) agertzen bazaigu, irizpide orokorra lexikalizatuaren alde egitea (*aurreko*, *hitz-erdi*, *arretaz*, *ikusgarri*) izango da.

Hala ere, zenbaitetan ezin da hori sistematikoki egin. Hori gehienetan gertatzen da forma lexikalizatuak egitura sintaktiko “arrunt” bat “bortxatzen”edo duelako, eta aldi berean esanahi-aldaketa dagoelako gehienetan.

Argia dugu, esaterako, hitz batek postposizio erabilera ere duen kasua. Horrelakoetan, postposizio balioa duenean gehien-gehienetan morfemaz osatutako analisia esleituko zaio, eta postposizio ez denean, forma lexikalizatuarena. Dena den, hori postposizioen zerrendan begiratzen ziurtatu behar da.

“Etxe aurreko denda” (postposizio-erabilera (IZE ARR kategoria)

“Aurreko egunean” (ADJ ARR gisa lexikalizatua)

sexuaren antzeko obsesioa → Postposizio-erabileran: antz (IZE ARR) + ko  
antzeko pelikula → lexikalizatua (ADJ ARR)

“buruz”en kasuan, postposizioen zerrendan bi aukerak eskaintzen dira:

buruz ADB: Liburu honi BURUZ mintzatu gara  
buru IZE + z: Irakasle gabe bere BURUZ ikasi zuen  
Hortaz, kasuan kasu etiketatu dugu.

Segidan, adibide gahiago:

“muturreko paretak”

*muturreko* (IZE ARR)

mutur + ko (IZE ARR + ko GEL) ✓

(Arrazoa: ez gaude ‘izen + izen’ moduko hitz elkartu baten aurrean, eta ‘izen + izen’ IS ez da egitura normala. Halaber, esanahi-aldaketa dago)

“XXI. mendeko”

*mendeko* (ADJ ARR)

mende + ko (IZE ARR + ko GEL) ✓

(Arrazoa: ‘det + adj’ IS ez da egitura normala. Halaber, esanahia guztiz desberdina da)

“Gizakien ibilian”

*ibilian* (ADB ARR)

ibili + an (IZE ARR + INE) ✓

(Arrazoa: ‘izlg + adb’ sintagma ez da egitura normala, ‘izlg + iz + INE, aldiz, bai)

“are gutxiago”

*gutxiago* (DET)

gutxi + ago (ADJ + ATZ GRAD) ✓

(Arrazoa: ‘adb + det’ sintagma ez da egitura normala, ‘adb + adj + grad’, aldiz, bai)

“-rik gorena”

*goren* (ADJ ARR)

gora + en (ADB + ATZ GRAD) ✓

(Arrazoa: -ik-ek graduatzaile superlatiboa behar du; adjektiboa hautatuz gero, ez genuke aukera hori)

Adi, ordea, aditzetan. Izan ere, juxtu alderantzizkoa gertatu ohi da: lexikalizatua gutxienetan da.

Adib.: “Egite handiko pertsona da”

egite: egin + te (atzizki lexikala)

egite (ize arr) ✓

egin + te (adize)

Adibide gehiago:

barne-egitekoa

sortzetiko bere ezagutza...

Hala ere, batzuetan ez dago oso garbi zer den komenigarriena. Horrelakoetan bakoitzaren intuizioak eta subjektibotasunak aginduko du. Adibidez:

data zehazte eta transkripzio neketsuetarako IZE ARR? ADIZE?

Gainontzean, gehienetan ADIZE izango da:

“Hori egiteak on egingo dio”

“Liburua jasotzeko esan dio”

jasotze + ko

jaso + tze + ko → ADIZE ✓

teko

teari

tea

Bide batez, gogoratu behar dugu hitz lexikalizatua adjektiboa denean, ez duela elipsiaren (ELI) analisia ematen aukeran; aldiz, morfologikoki banatutakoan bai (‘aurre + ko + ELI + a’ nahiz ‘aurre + ko + a’).

## Kategoria-bikoteak

### IZE/ADJ

Badira hitz asko izen nahiz adjektibo gisa jokatzen dutenak; besteak beste, -dun / -ari / -le/tzaile atzizkidunak eta jentilizioak.

Aspaldiko erabakiren batean horiek ez bikoiztea eta bakarrik ADJ gisa lantzea erabaki zen, baina gaur egun ez dirudi hori sistematikoki egina dagoenik EDBLn. Gainera, zenbait kasu ez dira oso argiak.

Horren aurrean:

- Testuan ADJ gisa jokatzen duenari, IZE proposamena bakarrik ematen bazaizu:

**ADJ sortuko** dugu, kasu hauetan IZE ez delako egokia eta seguru gaudelako ADJ dela:

pertsona moderno kultu bat

taila polikromatu erromanista bat

- ADJ bakarrik proposatu eta **IZE** egokiago dela uste izan arren, **ez sortu!** (IZEak ez ditugu sortuko):

Dirudunek

Esan dugu, hala ere, badirela kasu oso argiak. Adib.:

Paziente: ADJ gisa dago bakarrik, baina testuinguru batzuetan garbi dago IZE dela (paziente batekin zegoen); beraz, IZE sortu dugu.

- Bi aukerak ditugunean, jakina, testuan dagokiona. Adib.:

polizia islaerdar -> ADJ

bi palestinar hil dituzte -> IZE

75 magrebiar atzeman -> IZE

Musulmanen hilabete sakratua -> IZE

Palestinarrek egindako otoitza -> IZE

### ADB/ADJ

ADBak ere, ahal dela, ez ditugu sortuko. Adibidez:

zoriontsu bizi ADJ

egoki hitz egin ADJ

Hala ere, baldin balego hitz bat beti-beti adberbio erabilera duena eta analisisetan ADJ analisia bakarrik duena, orduan ADB sortu beharko litzateke. Corpusean ez dugu horrelakorik aurkitu.

## DET/IOR

Hirugarren pertsonako “izenordainak”, indartuak barne (*hura, haiek, bera, berau, berori*), berez determinante erakusle dira, izena isildu delarik IOR funtzioan daudenak.

Hortaz, benetako izenordainak hauexek dira:

ni, hi, zu, gu, zuek

## Izen berezi eta Leku-izenak

**IZB** pertsonen-izenak bakarrik dira:

Izen Berezi konposatuak:

Jose Migel Beñaran IZE IZB + IZE IZB + IZE IZB ABS MG

Bartolome Leonardo IZE IZB + IZE IZB ABS MG

Argala, (Jose Migel Beñaran) ETako buruzagia → **IZE IZB**

**LIB** leku-izenak, erakundeak, aldizkariak... (-ko/-ren hartzen dutelako)

BINA adimena,

BIC boligrafoak...

Norteko trena → LIB + ko

KAS alternatiba

Leku-izen konposatuak:

Ibn Tulun IZE LIB + IZE LIB ABS MG

Robert **MUSIL** → egokia ez den proposamena ematean, **sortu IZB!!**

Norteko trena → LIB + ko

Kontuz: San Migel → IZE ARR

Badago aipatzea merezi duten kasu batzuk: Tourra, Giroa...

Tour, leku-izen berezi gisa (IZE LIB) dago landuta EDBLn. Ondorioz, ez du “Tourrean” modukorik onartzen; hori izen arruntaren erabilera delako. Hala, lexikorik gabeak izen arruntaren (IZE ARR) analisia sortzen du, eta orduan ondo markatu dezakegu. Fenomeno hau (arrunt gisa deklinatzen den izen berezi (maiuskuladun)a) ez da oso logikoa, baina zenbaitetan gertatu egiten da, ikusten dugun bezala. Nahiz eta une honetan horrela egon EDBLn, badira JKak jokamolde hori onartzeko bereziki sortuak (hots, maiuskulaz idatzi baina izen arruntaren deklinabidea hartzen dutenak). Hala izanik, EDBLn JK egokia jarri beharko litzaioke Tour hitzari, baina momentuz corpusean agertzen den bezala etiketatuko da. Giro-ri dagokionez, ez dago EDBLn, beraz, sortu egin behar dugu. Behin sortzen jarrita, IZE LIB gisa bakarrik sortuko dugu, dakion deklinabide-atzizkiarekin.

- **Izen berezi bikoteak:**



Eugi-Olaizola  
EA-EAJ

Horrelakoetan, bat bezala etiketatuko ditugu: Eugi-Olaizola IZE IZB

Ohar bedi ondorengo adibidean ezin dugula hori egin, gaizki tokenizatutako hitzak direlako:

[Eugi I.][a- Olaizola]

## **DET DZH / IZE ZKI**

### **DET:**

. garren DET  
500 metro DET  
Orduak: Bederatziak arte DET  
5etan DET  
Kiroletako puntuazioak, sailkapenak....

### **IZE ZKI (DET bezala erabilia ez daudenean):**

5ak etorri dira (bost zenbakidunak) IZE ZKI  
11/83 legearen IZE ZKI  
2005 urtea beroa izan da  
Datak:  
2005-10-06 IZE ZKI  
martxoaren 1ean IZE ZKI + ean  
1963-80 bitartean IZE ZKI  
Ehunekoak: % 90 IZE ZKI  
IZE ZKI aukera bakarrik ematen digunean: IZE ZKI

## **Puntua**

Puntua ez da landuko. Hau da, testuan puntua falta bada edo analizatzaileak gaizki interpretatu badu (amaierako puntua izan beharrean, “garren”en sinbolotzat hartu badu, esaterako), guk ez dugu ez zuzenduko ez sortuko. Besterik gabe, “?” markarekin utziko dugu. Hori aurrerago aztertzeke geratzen da. Dena den, komeni da horrelakoekin topatzean fitxategiaren eta hitzaren erreferentzia apuntatzea.

## **Laburtzapenak: siglak, laburdurak, sinboloak**

EDBLtik sigla/laburdura/sinbolo gisa datozenak halaxe markatuko dira. Sigla batzuk IZB bezala eta beste batzuk LIB bezala izango ditugu, datozkigun analisien arabera. Siglen kasuan, oro har, ez dago koherentzia handirik, datu-basean ez daudelako guztiak eta daudenak ere ez direlako berdin-berdin aztertu. Adibide batzuk:

EH:

LIB SIG bezala dator, Euskal Herria adieraziarekin

EA:

IZB SIG bezala dator, Eusko Alkartasuna adieraziarekin

PP:

LIB bezala dator, SIG dela esan gabe

Datu basetik SIG bezala datozenak, horrelaxe etiketatu ditugu. Batzuetan, gure ustez sigla hori LIB izan arren, analisiak IZB eman badigu, IZB etiketa esleitu diogu analisi berri bat sortzen ibili gabe. Datu basetik SIG bezala ez datozenak, LIB SIG bezala sortzen ditugu adiera dakigun kasuetan, adib. PAN – Palestinako Aginte Nazionala. Adiera ez dakigun kasuetan, LIB bezala sortuko dugu SIG zehaztu gabe, adibidez BEA.

Laburdura eta sinboloei dagokienez, gehienak ez dira datu basetik etortzen. Aurreko irizpide berdina jarraitzea erabaki dugu: Laburduraren adierazia zein den dakigun kasuetan, zehaztu egingo dugu; bestela ustez hitzari dagokion kategoría jarriko dugu (LIB edo IZB edo IZE ARR). Adibide batzuk:

54. zenb. IZE ARR LAB  
Th Mann → IZB  
L. del Rio → IZE ARR  
Vit. → LIB  
Kel. (Kelme) → LIB LAB

### **-ko DESK, -ko BNK, -ko GEL (ikus Euskaltzaindiaren 34. araua)**

deskribatzailea → dun.ekin ordezkari daitekeena:  
3 urteKO ardoa / 3 urtedun ardoa  
diru gutxiKO pertsona / diru gutxidun pertsona  
mami gutxiKOa / mami gutxiduna  
leku-genitiboa → gehienetan ‘nongo’ galderari erantzuten diona  
kaxaKO zapatak  
unitateKO ikasgai nagusiak  
3 haurreARIKO bat  
banakaria: bost hurreKO, begirale bat; litroKO, 100 pta

Ohar bedi EDBLko 4.4.9 bertsio honetan -ko BNK-k ez duela 0 mugagabea ematen, eta beraz, sintagma amaieran doanean ez dagoela aukera hori:

2 milioi laguneKO [ø mugagaberik ez] banatu dute

Hau EDBLn zuzendu beharra dago, eta beraz, hurrengo bertsioetako analisisetan agertuko da. Etiketatzeari begira, berriz, analisi egokia emango dugu, eta beraz, ø mugagabea (ø 8) erantsiko dugu.

### **Elipsia**

Gehienetan elipsidun analisisia markatzea tokatuko da:

“Etxekoa ez da etorri” GEL **ELI** ABS MUG SING  
“kargak ikur berdinekoak badira” **ELI**  
“beti mendikoa izan den jendea” **ELI**  
“Karga hau bi motatakoa izan daiteke” → mota / ta / ko / **ELI** / a

Elipsirik gabeko kasuak horrelakoak dira:

“aingeru goardakoa” GEL ABS MUG SING (elipsirik gabe)

-KO, lemaren zati bada, **ELI**rik ez, ADJ ondoren ez baitator elipsirik analisisetan:

Argiaren hedapena nolakoa den aztertzea  
Istorio hori benetakoa da

**Gainontzekoak** ondorengo ikasturtean  
Liburu handiak han daude eta **txikiak** hemen  
Hainbat berrikuntza jasan ondoren, **azkena** 1999an, ...

DETetan ere ondoren ez dator elipsirik:  
XV. eta XVI.**etan**

Erlatibozkoen ondoren beti ELI:

Egin zena → zen / n erlt / **ELI** / a  
Deritzona → deritzo / n / **ELI** / a  
Hark esaten duenAZ / dioenAZ, ez dakit ezer → ELI

Dakigunez  
Dioenez  
Dagokionez  
Esaten denez  
Konturatuko zinenez

} Elipsirik ez!!  
eta KAUS erlazioa

Etortzen denean, → elipsirik ez; eta DENB erlazioa  
egiterakoan  
duenean

dena dela: ADT ELI + ADT MOD } HAULen zerrendako kategori-analisiekin bat  
dena den: ADT ELI + ADT MOS } etorri behar du

Delakoa → ELI ez, *delako* sarrera delako

## Parentesi artekoa

Parentesi barrukoa sintagma kontsideratuko dugu!!

Natur Baliabideen Antolamendu Plana (**NBAP**) → ABS MG (Ø = 8)  
Natur Baliabideen Antolamendu Planetik (**NBAP**) ondorioztatzen da... → ABS MG (Ø = 8)

## Entitateak (maiuskulaz idatziriko hainbat izen)

Gauza asko dira entitateak. Baina guri batez ere EDBLn landuta ez dauden horiek interesatzen zaizkigu: izen berezien siglak (adib. J.M Artetxe); euskarazkoak ez diren pertsona-, leku- eta entitate-izenak... (maiuskuladun foreing asko, seguruenez).

Horrelakoak askotan aposizioan ( “X (deitzen) den Y” ) azaltzen dira. Hala izanik, ez dira testuinguruan “txertatuta” kontsideratzen; hau da, deklinabide-atzizkirik badu, izenaren berezko marka da, ez esaldian dagokiona (adibide garbia: Ameriketatik txalupa).

Entitateak lokalizatzeko, Izaskun F.ek lehenik eta behin maiuskularik badagoen begiratzen du, eta hala denean, hitzaren kategoria IZE den. Alegia, ez zaio axola arrunta edo berezia den.

Hori horrela izanik:

- Testu-hitza EDBLtik LIB, IZB edo ARR gisa analizatzen bada, dagokiona hautatu behar da. Adib.:

Nikolas Txiki oporretan eleberria -> IZE IZB

- Testu-hitza lexikorik gabe analizatzen bada eta LIB, IZB edo ARR analisia ematen badira, dagokiona hautatu behar da. Adib.:

Windows → LIB

Txanogorritxu → ipuina: LIB / pertsonaia: IZB

- Testu-hitza lexikorik gabe analizatzen bada eta ARR analisia bakarrik baldin badago, LIB etiketa sortuko dugu. Adibidez:

Korapiloak eleberrian → IZE LIB

- Lexikorik gabekoak ez badu IZE ARRIk ematen, dagokion azpikategoriarekin sortuko dugu izen berezia. Adib.:

Argala, (Jose Migel Beñaran) ETako buruzagia -> ADJ analisia bakarrik, beraz, IZB sortu.

Hau, gehienetan ondorengo kasuetan gertatzen da: izen berezia, berez, arrunki erabiltzen diren hitzez osatuta dago. Hori horrela izanik, analizatzaileak arrunki duen analisi hori eskaintzen du, eta, jakina, ez du lexikorik gabekorik proposatzen. Adibide gehiago:

Negu Gorriak: IZE ARR, ADJ ARR. Baina guk IZE LIB IZE LIB sortuko dugu.

Ameriketatik (txalupa): IZE LIB. Baina guk IZE LIB sortuko dugu.

Ur Kirolak: IZE ARR IZE ARR. Baina guk IZE LIB IZE LIB sortuko dugu

Santiagotarrak: ADJ ARR. Baian guk IZE LIB sortuko dugu.

Ikertu (aldizkaria): ADI SIN. Baina guk IZE LIB sortuko dugu.

Mendi Magikoa: IZE ARR ADJ ARR. Baina guk IZE LIB IZE LIB sortuko dugu.

Sortzerakoan, leman kasu eta guzti jarriko dugu:

Negu Gorriak -> lema: Negu; lema: Gorriak

Ameriketatik (txalupa) -> lema: Ameriketatik

Ur Kirolak -> lema: Ur; lema: Kirolak

Santiagotarrak -> lema: Santiagotarrak

Ikertu -> lema: Ikertu

Mendi Magikoa -> lema: Mendi; lema: Magikoa

Aposizioan agertzen ez direnean (Negu Gorriak-ek, Santiagotarrak-ekoek...), dagokion lemaren analisia (IZE LIB, IZE ARR...) eta gero dagokion deklinabide-atzizkiaren analisia (-ek, ekoek...).

Adibide luze bat:

Herrera	de	la	Mancha
LIB (A)	BST (A)	BST (E)	LIB (E)

EDBLn dauden izen bereziak IZB edo LIB agertuko dira; ez daudenak, lexikorik gabe IZE ARR ematen badu, IZE ARR jarriko ditugu, eta bestela dagokion azpikategiaz sortuko. *la* BST-tzat hartzen da entitateetan; EDBLtik ez datorrenez, *la* BST sortuko dugu.

Oharra:

Aipatutako izenak minuskulaz eta arrunki deklinaturik baldin badaude, guk ere horrela kontsideratuko ditugu. Adib.: “Donostiara etorriko direla esan dute santiagotarrek”; “santiagotarrek etorri dira” -> ADJ ARR. Gure asmoa ez da horiek ondo edo gaizki erabiliak dauden auzian sartzea.

## Hitz arrotzak/zaharrak/arraroak/ezezagunak

Hitz arrotzei eta hitz zaharrei IZE ARR analisia jarriko diegu. Adibidez:

Bi lan idatzi zituen euskaraz: Avisu la exortacionea provetchosac bekhatorentz;at hitz lauz (Bordele, 1641) eta Devoten Breviarioa bertso-liburua (1665).

Sistemaren (green car)... IZE ARR IZE ARR

Mecauen Satanas IZE ARR

Enfin

Hitz arraroei (inongo esanahirik ez dutenei), adib.: aba, ean, an.... IZE ARR edo BST jarri izan diegu.

Hitz ezezagunekin, azkenik, lehehengo Google-en begiratu eta maiztasun handiena duena aukera daiteke. Ez badugu gustukorik aurkitzen, IZE ARR jarriko diogu, gainontzekoetan bezala.

Adib., Google-en begiratu:

hartmane + k

hartman + ek

brullete + k

brullet + ek

Paleolito, Ordotriziko, Deboriar

## Anbiguotasuna

Anbiguotasuna dagoenean, norberak subjektiboki markatuko du iruditzen zaiona.

Hala ere, badira kasu batzuk erabaki ditugunak:

Anbiguotasuna askotan dator kasu instrumentalean eta inesiboan mugagabea, mugatu plurala eta mugatu singularra bereizteko zailtasunetik. Eta batez ere “a organikodun” hitzak direnean. Horrelakoetan, a-z amaitzen ez den beste hitz batekin probatzea da egokiena. Adib.:

Euskaraz badaki → IZE ARR MG INS edo IZE ARR M SING?

Ez dugu esaten “katalanaz, ingelesaz... badaki”, baizik eta: katalanez, ingelesez...

Beraz: IZE ARR MG INS

Ordea: pilaz dabilen jostailua: IZE ARR MG INS edo IZE ARR M SING???

Beste erabaki bat:

uretan murgildu: IZE ARR INE MGG

BAINA:

Laida eta Lagako uretan murgildu ginen: IZE ARR INE M P

Aditz-izenen koordinazioan:

- ADIZE + deklinabide marka dagoen kasuetan, bi osagaiak hartuko dute sintagma bukaeraren marka

Itxura apaintzeKO eta edertzeKO (helburuzkoak)

ADIZE ADIZE

KO 6 KO 6

Ø 19 Ø 19

-Deklinabide-markarik ez badago, jakina, sintagma-bukaeran doanak hartuko du dagokiona:

Itxura apaindu (ADI PART) eta edertzeko (HELB)

· Forma ez-burutu bera izan baina aditzoina desberdina dutenak:

nahasten → nahasi / nahastu

bereizten → bereizi / bereiztu

estaltzen → estali / estaldu

}

1. aren alde egingo dugu

## Metahizkuntza

Metahizkuntzaren aurrean gaude hizkuntzaren operatzaileei egiten zaienean erreferentzia. Adibidez:

eta juntagailua gehienetan erabili, hala ere, ordea, arren eta horrelakoen ordeaz.

Yong hitzak “ausarta” esan nahi du, eta Yung-ek “salbatu”

alkar ik. elkar

Hauetan IZE ARR jarri, nahiz eta analisisien artean bere zeregin linguistikoko analisia agertu (eta -> JNT).

Azken batean, hemen ere aposizio-egitura baten aurrean gaude, askotan bigarren elementua eliditu egiten delarik (eta [juntagailua] erabili). Entitateetan esan dugun bezala, aposizioetan aurreko elementua hutsik etiketatzen da (horregatik IZE ARR), esaldian bere ohiko funtzioan ez dagoelako. Metahizkuntza-elementua deklinatua baldin badago (adib. Yung-ek “ausarta” esan nahi du), IZE ARR eta dagozkion mugatasuna eta deklinabide-atzizkia jarri (-e)k: MGG ERG)

Sortzerakoan, leman komatxo bikoitzak jarri behar da, ez, ordea, forman.

Forma	Lema
<u>eta</u>	“eta”
<u>alkar</u>	“alkar”
<u>Yong</u>	“Yong”

Komatxo bikoitzen bidez, elementu linguistiko horrek metahizkuntzako erabilera duela jakingo dugu, eta EDBL aberasteko garaian erraz multzokatu eta baztertu ahal izango ditugu.

**Liburuak, testuen... izenburuak, eta titularrak, atalak, azpiatalak...**

Horrelako testu-zatietan, hitzok, hain zuzen ere, funtzio hori dute: izenburu izatearena, titularrak/atalak/azpiatalak... izatearena. Orduan, delako funtzio horretan duen analisisiaz markatu behar da. Adib:

Bi anai -> DET IZE MGG ABS

Hiru txerritxoak -> DET IZE M S ABS

1. jarduera → ABS MUG SING

Leire Agirre, Espainiako txapeldun → MGG ABS MGG ABS

Aldiz, liburuen, testuen... izenburuak testuan txertatuta agertzen direnean (adib., Hiru txerritxoak ipuina mundu guztian zabaldua dago...), orduan entitateak bezala etiketatuko ditugu. Ikus, beraz, "Entitateak" atala.

## **Aldaeretan bi balizko aukera baldin badaude**

Marratxoa:

partehartzaile / parte-hartzaile  
nonahi / non-nahi

Bi aukerak dira zuzenak, baina marratxodunaren alde egitea erabaki dugu:

- corpus teknikoetan marratxoa gomendatzen delako
- EDBLn sarrera marratxoarekin dagoelako (eta aukeran, marratxorik gabea)

## **Determinatzaile konplexuak**

dozena bat lagun  
makina bat liburu

Hauetan sintagma bakarra markatuko dugu. Hots:

dozena bat lagun: IZE DET IZE MG ABS

## **Partizipio hutsarekin egiten diren erlatiboak**

Iragan astean

ADI PART

egin bilera

ADI PART

...egunkariak Paul Laxalti eskaini azalean

ADI PART

## **Kasu partikularrak**

. **Ohi, berri:**

Erakutsi	berri	dut
BURU	ADJ MGG	ADL

Erakutsi ohi dut  
BURU BAST MGG ADL

Idazkari nagusi berraukeratu berria izango da  
IZE ADJ PART ADJ M S ABS

### . Tourra, Giroa....

Ikus “izen berezi eta leku-izenak” atala. Hango azalpen batzuk ona ekarri ditugu.

Tour, leku-izen berezi gisa (IZE LIB) dago landuta EDBLn. Ondorioz, ez du “Tourrean” modukorik onartzen; hori izen arruntaren erabilera delako. Hala, lexikorik gabeak izen arruntaren (IZE ARR) analisia sortzen du, eta orduan ondo markatu dezakegu. Fenomeno hau (arrunt gisa deklinatzen den izen berezi (maiuskuladun)a) ez da oso logikoa, baina zenbaitetan gertatu egiten da, ikusten dugun bezala. Nahiz eta une honetan horrela egon EDBLn, badira JKak jokamolde hori onartzeko bereziki sortuak (hots, maiuskulaz idatzi baina izen arruntaren deklinabidea hartzen dutenak). Hala izanik, EDBLn JK egokia jarri beharko litzaioke Tour hitzari, baina momentuz corpusean agertzen den bezala etiketatuko da. Giro-ri dagokionez, ez dago EDBLn, beraz, sortu egin behar dugu. Behin sortzen jarrita, IZE LIB gisa bakarrik sortuko dugu, dakion deklinabide-atzizkiarekin.

### . gabeak

Postposizioen zerrendan ez da aurreikusten ‘-ik + gabe + deklinabide-atzizkia’, *gabe* hutsa baizik, ADB kategorian. Beraz, “akatsik gabea” moduko egituretan, *gaber* ADJren analisia ematea erabaki dugu. Aztertu beharko litzateke ea *gaberen* postposio-erabileran ADJ kategoria onartu/jarri behar den.

### . Iparraldeko “objektu-genitibo”aren erabilera (*etxearen egiteko*)

etxearen egiteko  
arauen betetzeko:  
GEN ADOIN + IZE + GEL ABS Ø8 → izen sintagma

### . Gutxi, gehienak....:

mami gutxikoa, indar gutxikoa... → DET DESK  
pronostiko gehietak → DET

### . -N:

erlatibozkoa  
zehir-galdera  
MOS (ERLT eta ZH ez diren beste guztiak)  
non... deN  
nahiz eta... deN  
konturatuko zineN bezala  
deN bitartean



ahal deN azkarrena  
zeintzuetan frogatua izan deN

zaitezeN:

helburuzkoa: zu etor zaitezen, opilak egingo ditut  
konpletiba: zu etor zaitezen nahi dut  
erlatiboa: etor zaitezeN nahia dut

· **-neko (direneko, duteneko...)**

Ondoko erabileratan (ondo edo gaizki egon...):

“Materialak pitzatzen hasten direneko deformazioen balioak ezagutu daitezke”  
“Nukleo atomikoek parte hartzen duteneko elkarrekintza”

‘-(e)n erlatiboa + ko’ genitiboa markatuko dugu.

· **Helburuzkoak, konpletiboak:** joTZEKO, maitaTZEKO...

**Helburuzkoak:**

- izenaren aurretik agertzen denean (festara joateko arropa erosi du) → ADI ADIZE GEL IZLG  
- izenaren aurretik agertzen ez denean (arropa erosi du festara joateko) → Ø 19 ADLG OBJ  
(gero, morfosintaxian emango zaio helburuzkoei dagokien informazioa)

**Konpletibak:** esan, eskatu, agindu.... -teko/-tzeko

· **Geroa:** aplikatuko zaiolarik, egingo du, ikusiko dugu... kenduko bagenu  
MG (mugagabea) ez!!! **PART GERO ADL**

· **Agintera:** PARTIZIPIOan → **ADI SIN PART**  
ikus eranskinak, ikus 20. orria.... → **ADOIN**

· **ADOIN:**

agintera → etor zaitez  
ahalera → etor naiteke  
subjuntiboa → etor nadin\* esan dit

\*Aditz laguntzaile honek KONPL erlazioa beteko du esaldian

· **ADIZE INE (obj):**

behatu beroTZEN ari den bitartean  
jarraitzu egiTEN  
jarraitu tenperaturak harTZEN  
irakiTEN hasi arte

· **-t(z)en ahal da**

‘Ahalerazko perifrasi honetan -t(z)en duen formari objektu analisia emango diogu, nahiz eta ez dugun inon horretarako arrazoi zehatzik aurkitu.

**-t(z)en** ahal da  
OBJ IZE ADT

· **Horrela izaki...**

Momentu honetan EDBLn dagoen informazioarekin ondoko analisiak agertzen zaizkigu:

A 1 jarraiki ADI SIN + 0[10] ADOIN

- A 2 jarraiki ADI SIN + 0[10] PART
- A3 jarraiki ADI SIN + 0[10] PART + 0[13] BURU
- A4 jarraiki ADI SIN + 0[10] PART + 0[13] BURU + 0[8] ABS MG @OBJ @PRED @SUBJ
- A 5 jarraiki ADJ ARR IZAUR\_-
- A 6 jarraiki ADJ ARR IZAUR\_- + 0[8] ABS MG @OBJ @PRED @SUBJ

Eta horietako bat bera ere ez da egokia moduzko balioa emateko. Berez, horrelako zerbait behar genuke:

- A 3 izan ADI SIN ADOIN + ki ERL MOD @-JADLAG\_\_MP @-JADNAG\_\_MP

Alegia, -ki ERL MOD beharko luke “jarraiki”k eta gero erregelen bidez-edo ‘ki+ki=ki’ egin. Hori EDBLn konpontzeko geratzen da. Guk momentuz sortu egingo dugu analisi hori.

· **PNT** (puntuaria): guk bagenu (trinkoetan, ADLtan ez)

· **ari / aritu:**

ari da berotzen → HAOS + ADT + ADIZE INE (ari izan HAULen zerrendan dagoenez: HAOS)

ari izan da → HAOS + PART BURU + ADL

aritu da → PART BURU + ADL

aritu da ikasten → PART BURU + ADL + ADIZE

aritu izan da ikasten → PART + PART BURU + ADL + ADIZE OBJ INE

· **PART BURU:** izan zen, izan zituen, egin zuen, eraiki zen, ikusi omen dute....

· **PART:** bizi eta ezkututzen dira,

· **Aditz-modalak:** nahi izan, ezin izan, ahal izan, behar izan...**IZE ARR** (obj modukoa) / **ADJ ARR** (pred modukoa)

oharra: gauzatze eskemetan “-“ dagoenean, ABS MG da beti !!!

<b>behar</b>	<b>dut</b>	<b>bizi</b>	<b>behar</b>	<b>zuten</b> (bizi izan HAUL)
IZE	ADT	ADJ	IZE	ADT
MG		ARR	MG	

<b>joan</b>	<b>behar</b>	<b>dut</b>	<b>etorri</b>	<b>nahi</b>	<b>dut</b>
ADI	IZE	ADT	ADI	IZE	ADT
SIN	MG		SIN	MG	
PART			PART		

<b>behar</b>	<b>izan</b>	<b>dut</b>
IZE	ADI	ADL
MG	SIN	
	BURU	

<b>joan</b>	<b>behar</b>	<b>izan</b>	<b>dut</b>	<b>egin</b>	<b>beharko</b>	<b>dut</b>	<b>nahiko</b>	<b>dut</b>
ADI	IZE	ADI	ADL	ADI	ADI	ADL	ADI	ADL
SIN	MG	SIN		SIN	GERO		GERO	
PART		BURU		PART				

<b>nahi</b>	<b>izango</b>	<b>dut</b>	<b>egin</b>	<b>behar</b>	<b>izango</b>	<b>dut</b>
-------------	---------------	------------	-------------	--------------	---------------	------------

IZE ADI ADL  
MG SIN  
GERO

ADI IZE ADI ADL  
SIN MG SIN  
PART GERO

**sufritu beharra dagoen mundua dago ukatu beharra**  
ADI IZE ADT IZE ADT ADI IZE  
SIN ABS PNT SIN PNT SIN ABS  
PART SING ERL PART SIN

**utzi egin behar zaio**  
ADI ADI IZE ADT (aurretik izena duelako)  
SIN SIN MG  
PART PART

**utzi egin behar izan zaio**  
ADI ADI IZE PART ADL (aurretik aditza duelako)  
SIN SIN MG BURU  
PART PART

**esan behar dio**  
ADI IZE ADT (aurretik izena duelako)  
SIN MG  
PART

**Izandun aditz konposatuak:** maite izan, bizi izan, berdur izan, kosta izan... IZE ARR (obj modukoa) / ADJ ARR (pred modukoa)

oharra: gauzatze eskemetan “–“ dagoenean, ABS MG da beti !!!

**bizi behar zuten** (bizi izan HAUL)  
ADJ IZE ADT  
ARR MG

**beldurra dio**  
IZE ADT  
MG

#### . Kosta/Pasa/Falta:

‘Kosta’ eta ‘pasa’ aditzekin ADOIN eta BURU aukerak ematen zaizkigu, ‘falta’-ren kasuan ADOIN bakarrik. Horrela izanda, laguntzailerik gabe azaltzen direnean, eta PART hutsaren analisirik ez daukatenez, ADOIN analisia ematea erabaki dugu.

Urteak pasa ahala...  
ADOIN

Hurrengoetan berriz, PART BURU analisia emango diegu (laguntzaile bat badagoelako):

pasa da  
ADI PART BURU

% 17 pasa egin ditu  
ADI PART BURU

asteburu pasa etorri da  
ADI PART BURU

-

. **Berbizi**: Kasu honetan ere ADOIN da ematen zaigun aukera bakarra. 'Bizi izan' aditz modala izanik, lehen osagaiari (bizi) aurreko puntuan zehaztutako kategoria esleitu diogu (ADJ). 'Berbizi' agertu denean baina, ADOIN soilik ematen zaigunez, hura aukeratzea erabaki dugu.

- Sua une batzuetan **berbizi** egiten da eta behin betiko itzaltzen da.  
(ADOIN aukera bakarra eta hori markatu)

### **Izan aspektuala, egin indargarria.**

'Egin' indargarriaren eta 'izan' aspektualaren aurreko aditzei PART jartzen diegu.

Adibideak:

**joan izan da**  
PART ADI ADL  
SIN  
BURU

**piztu egin denez**  
PART PART ADL  
ASP  
BURU

**piztu egiten denez**  
PART ADOIN ADL  
ASP  
EZBU

landarea **hazi egin da**  
PART PART ADL  
ASP  
BURU

'Kosta'-k, ordea, ez du aukera hori ematen (ADOIN edo BURU ematen du). Beraz, ADOIN analisisia ematea erabaki dugu.

Oraindik **kosta** egiten zaidak, benetan, ulertzea...  
ADOIN EZBURU

### **Perifrasiak: PART BURU MUG S/P + ADT**

**esana da** } PART ABS SING + ADT

**idatzia den**  
**hartua da**  
**lapurteratua den**

· **ohi:** BST

Kapitela errematatzen jarri ohi den pieza prismatikoa  
ADIZE INE PART BST ADL ERL

· **ahalera:**

Aginterak, subjuntiboak eta ahalerak laguntzailearen aurreko adizkia aditzoin izatea eskatzen dute.

Baina badira zenbait kontu hemen zehaztu beharrekoak. Batetik, aditzaren arabera, aditzoina nahiz partizipioa, biak dira zilegi laguntzaile hauen aurretik. (Ikus Euskaltzaindiaren 28.araua)  
argi ditzakegu  
argitu ditzakegu

Baina corpusean gertatzen da aditzoin beharko lukeena partizipio gisa agertzea.  
Ikusi daiteke ('ikus daiteke' izan beharko litzateke)  
Horrelakoetan PART analisia onartu dugu.

· **Predikatiboak:** ABS MG

etxea puskatua ikusi dut  
ABS MG  
haurdun geratu  
ABS MG  
gaizkile bihurtu  
ABS MG

· **PRT:**

Erroma salbatu omen zuten (PART BURU + PART + ADL)

Ez hau, ez beste hau

Bai zu eta bai ni

· ezin hobe, ezin jarraitu → IZE ARR

(ezin du -> IZE ARR MG)

· dela ekialdeko, dela mendebaldeko.... Konpletiboa ere izan zitekeen, baina aditz bereziak dira eta  
MOD/DENB MOD/DENB MOD/DENB aditz guztiek onar dezakete

· Hori dela eta → MOD/DENB + KAUS

· bata beira igurtziz sortzen zena eta bestea anbar izeneko erretxin fosiletik sortutakoa  
DET (elipsirik ez, izenordain funtzioa) -KO deskribatzailea - IZLG

· batez beste: DET INS + DET MGG

· euskal eta espainiar IZE ARR ; italiar.... ADJ

· nekazal IZE ARR (euskal-en tankerakoa)

- **Euskal Herria:** IZE ARR + IZE ARR (→ euskal beti IZE; ez dugu LIB bezala etiketatuko)
- **Hego Euskal Herria** eta **Iparraldea:** denak IZE ARR (ez LIB)
- **Eusko Jaurlaritza:** IZE ARR + IZE ARR
- **konparaziozkoak** (baino....) → BST
- Piztu egin **denEZ**, erre egin **denEZ** → KAUS
- ren **omenez** → IZE (postposizioen txostenean ez da agertzen!)
- gauza **askotaz**, **beretaz**... → erroreak direnez, utzi (ondo: gauza askoz, beraz....)
- **berbera...**

‘Berbera’ determinatzailea singular gisa bakarrik tratatzen da EDBLn. Orotariko Euskal Hiztegian, ordea, ‘berberak’ pluraleko adibide batzuk agertzen dira. Eta guri corpusean pluraleko erabilerak agertu zaizkigu, beraz, pluralean sortuko dugu.