

Hizkuntzaren alderdi formalak lantzeko aurkezpena

Euskara eta euskal kultura eremu digitalean
2019 Donostia

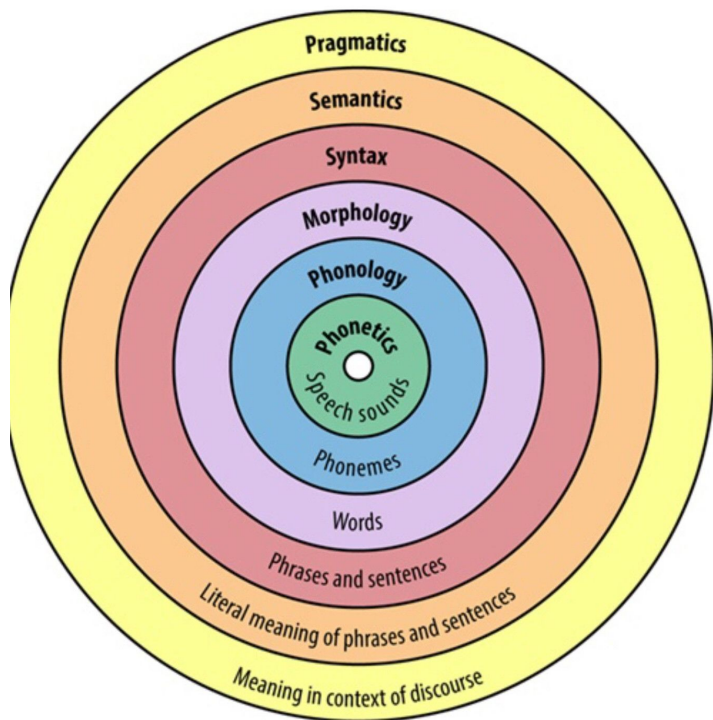
Mikel Iruskietta
Abel Camacho



Hizkuntza-teknologia eta hizkuntza-mailak

	Hizkuntzaren prozesamendua
Fonetika	PRAAT eta WaveSurfer: fonetika aztertzeko AhoSR: ahotsa testu bihurtzeko AhoTTS eta ElhuyarTSS testua ahots bihurtzeko
Morfologia	EDBL : datu-base lexikala Morfeus : analizatzaile morfologikoa
Lexiko-semantika	EuskalWordNet : euskarako sortutako ezagutza-base lexikal eleanitza ANALHITZA , Voyant : aztertzaile morfo-lexikalak
Sintaxia	UD corpora eta Arborator: euskarazko zuhaitz sintaktikoak manipulatzeko MALTIXA : analizatzaile sintaktikoa
Testua eta diskurtsoa	Lextutor : baliabideak sortzeko, ikuspegi lexikaletik Coh-metrix eta Textinspector: testuen informazioa lortzeko tresnak Compress-Eus : laburpenak kudeatzeko eta ebaluatzeko tresna SENEKO : galdera-erantzun sistemak garatzeko ingurunea Euskal RST Treebanka : koherentziazko erlazioen datu-basea

Hizkuntza-mailak eta teknologia



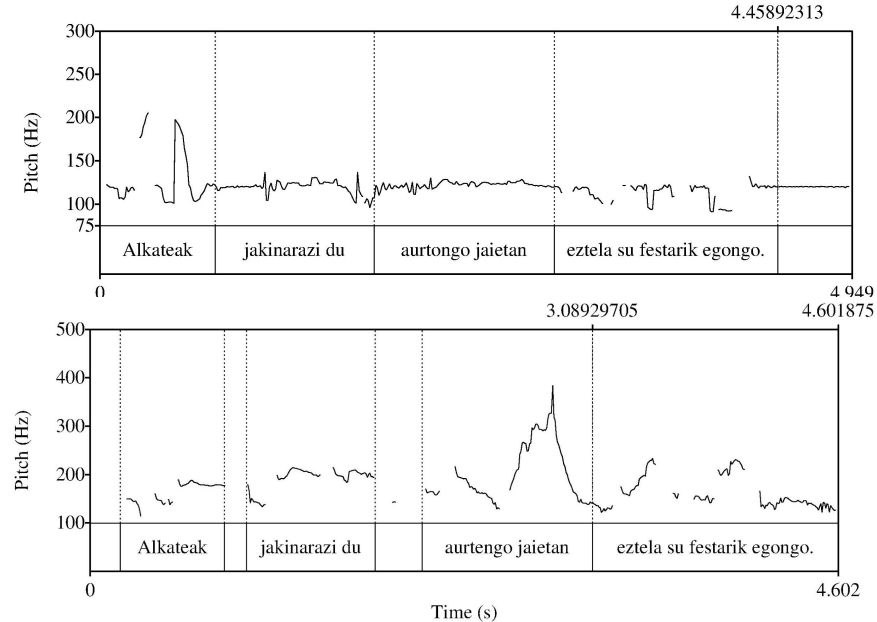
Hizkuntza mailak				
Fonetika Fonologia	Morfologia	Sintaxia	Semantika	Pragmatika
Soinuak, soinu-sistema	Formak eta hitzak	Perpausak eta esaldiak	Esanahi-motak eta erlazio semantikoak	Hizkuntza erabileran
PRAAT, WaveSurfer, AhoSR, AhoTTS eta ElhuyarTSS	EDBL eta Morfeus , ANALHITZA eta Voyant	MALTIXA eta Arborator	ANALHITZA , Voyant eta EuskalWordNet	Euskal RST Treebanka

Fonetika ikasteko



Fonetika ikertzeko edota irakurketa ozena aztertzeko tresna: PRATT

- Nabarmendu intonazioarekin NOIZ eta NORK, esaldi honetan eta aztertu PRAATEkin
- Fonetika denbora errealean aztertzeko tresna:



Esaldi bateko hitzen kategoriak

Erabili Morfeus hizkuntzaren anbiguotasuna edota aukerak zein diren ikusteko

- Zapata berriak erosi nituen, Zaragozan. Polit askoak, gainera!
- Non dago zuk aipatu duzun zentro komertziala?
- Nire lagun batek biolina jotzen du, oso belarri fina baitu.
- Mila esker! Bihotz handia duzu, eta, gainera atsegina zara.

ADIBIDEEN ITURRIA: <https://vimeo.com/11489099>

Erabili Eustagger hitzen kategoriak zein diren jakiteko:

- Mila esker! Bihotz handia duzu, eta, gainera atsegina zara

Kategorien esanahia ikusteko clickatu [hemen](#).

Mila	esker	!	Bihotz	handia	duzu	,	eta	,	gainera	,
<i>mila</i>	<i>esker</i>	<i>!</i>	<i>bihotz</i>	<i>handi</i>	<i>ukan</i>	<i>,</i>	<i>eta</i>	<i>,</i>	<i>gain</i>	<i>,</i>
DET	IZEARR	PUNT_ESKL	IZEARR	ADJIZO	ADT	PUNT_KOMA	LOTJNT	PUNT_KOMA	IZEARR	PUN

Morfeus:

<http://ixa2.si.ehu.es/demo/analisianali.jsp>

Eustagger:

<http://ixa2.si.ehu.es/demo/analisisimorf.jsp>

EDBL: euskarazko datu-base lexikala

- Zer kategoria eskatzen du -tasun atzizkiak eta zer kategoria ematen du?
 - Gorritasun, edertasun, *etxetasun, *egontasun, *gaizkitasun
- Eta -garri atzizkiak?
- Eta -tzar atzizkiak?

- Zer erlazio adierazten du -lako erlazio morfemak?
- Eta -bait erlazio morfemak?
- Eta -la erlazio morfema anbiguoak?

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

Hiztegiak

- [Elhuyar Hiztegiak](#) erabiliz, eman euskaraz ondorengo hitzak eta entzun euskarazkoak nola esaten dituen gizakiak eta makinak:
 - Darse por vencido, excelente, rima, manjar
 - Inapetencia, rechazar...
- Erabili [Euskalbar](#) eta bestelako hiztegiak:
 - <http://www.hiztegia.net/>
- Ontologia eleanitza: EuskalWordNet

<input type="text" value="ume"/>	<input type="button" value="Look up"/>	<input checked="" type="checkbox"/> Gloss	<input checked="" type="checkbox"/> English_3.0	<input type="checkbox"/> Catalan_3.0
<input type="text" value="Word"/>	<input type="text" value="Nouns"/>	<input type="checkbox"/> Score	<input checked="" type="checkbox"/> Basque_3.0	<input type="checkbox"/> Portuguese_3.0
<input type="text" value="has_hyperonym"/>	<input type="text" value="Basque_3.0"/>	<input type="checkbox"/> Rels	<input checked="" type="checkbox"/> Spanish_3.0	
	<input type="text" value="Basque_3.0"/>	<input type="checkbox"/> Full	<input type="checkbox"/> Galician_3.0	

EuskalWordNet: hiperonimoak

ume

Word Nouns Basque_3.0

has_hyponym Basque_3.0

Gloss English_3.0 Catalan_3.0
 Score Basque_3.0 Portuguese_3.0
 Rels Spanish_3.0
 Full Galician_3.0

Multilingual Central Repository (ILI 3.0) - [WikiMCR](#)

ili-30-01321579-n

[animals](#)
[biology](#)
[Animals](#)
[animal_1](#)
[animal](#)

[eng-30-01321579-n](#) 46 [young_1](#) [offspring_3](#) any immature animal
[eus-30-01321579-n](#) 46 [kume_1](#) [ume_1](#) animalia jaio berria: *hamabi hilabetetik gora behar ditu zerri emeak umea hazteko;*
[spa-30-01321579-n](#) 46 [cria_2](#) [vástago_1](#)

[NonFullyFormed+](#)

[Animal=](#)
[Object=](#)

ili-30-00015388-n mn

99

[animals](#)
[biology](#)
[Biology](#)

base concept

[object_1](#)

[Tops](#)

[Animal=](#)

[Animal=](#)

[Object=](#)

[eng-30-00015388-n](#) 4017 [animal_1](#) [animate_being_1](#)
[beast_1](#) [brute_2](#) [creature_1](#) [fauna_2](#)
[eus-30-00015388-n](#) 4017 [fauna_1](#) [animalia_1](#) [abere_1](#)
[spa-30-00015388-n](#) 4017 [animal_1](#) [bestia_1](#) [criatura_1](#)
[fauna_1](#) [res_1](#) [animales_1](#)

a living organism characterized by voluntary movement

sentitzeko eta berez mugitzeko gauza den izaki bizia: *jalkitako material horiekin batera arro edo haran sedimentarioetan bizitako animalia eta landare batzuen hondarrak pilatzen dira, harri bereziak sortaraziz;*

EuskalWordNet: hiperonimoak eta adierak

ili-30-09917593-n

[person](#)
[Animals](#)

base concept

[person_1](#)
[person](#)

[NonFullyFormed+](#)

[Function+](#)
[Human=](#)
[Object=](#)

[eng-30-09917593-n](#) 27 [child_1](#) [kid_1](#) [youngster_1](#) [minor_1](#) [shaver_2](#) [nipper_1](#)
[small_fry_2](#) [tiddler_1](#) [tike_2](#) [tyke_3](#) [fry_3](#) [nestling_2](#)

[eus-30-09917593-n](#) 27 [sein_1](#) [hur_4](#) [ume_3](#)

[spa-30-09917593-n](#) 27 [alevín_1](#) [chaval_1](#) [chico_1](#) [chiquillo_1](#) [chiquiuelo_1](#) [criatura_3](#)
[cria_1](#) [crío_1](#) [jovenzuelo_1](#) [menor_de_edad_1](#) [mozalbeta_1](#) [muchachuelo_1](#) [nene_1](#)
[niño_1](#) [pececillo_1](#) [tike_1](#) [menor_1](#) [joven_1](#) [chango_1](#)

a young person of either sex: *she writes books for children; they're just kids; 'tiddler' is a British term for youngster;*

gizaki oso gazte eta txikia, nerabezarora heldu ez dena: gaur egungo umeak beste joku batzuekin dibertituko dira, baina garai hartakoak ere ez zuten aspertzeko astirik; eragiketa matematikoak haurrei irakasteko modu grafiko ohikoena marrubi-pastelen marrazkien bidezkoa izan omen da; persona joven de cualquiera de los dos sexos

ili-30-09622049-n mn 99

[person](#)
[Animals](#)

[person_1](#)
[person](#)

[NonFullyFormed+](#)

[Function+](#)
[Human+](#)
[Object+](#)

[eng-30-09622049-n](#) 49 [juvenile_1](#) [juvenile_person_1](#)

[eus-30-09622049-n](#) 49 [gazte_1](#)

[spa-30-09622049-n](#) 49 [menor_2](#) [juvenil_1](#)

a young person, not fully developed

haurtzaroaren eta adin nagusiaren artean dagoena: gazte bat ikusi nuen sartzan, nire adinekoa gutxi gorabehera;

ili-30-09918248-n

[person](#)
[Animals](#)

base concept

[person_1](#)
[person](#)

[Human+](#)

[Function+](#)
[Human=](#)
[Object=](#)

[eng-30-09918248-n](#) 32 [child_2](#) [kid_4](#)

[eus-30-09918248-n](#) 32 [seme-alaba_1](#) [hur_3](#) [ume_4](#)

[spa-30-09918248-n](#) 32 [hija_2](#) [hijo_3](#) [ninzzo_1](#) [niña_5](#)
[niño_4](#)

a human offspring (son or daughter) of any age: *they had three children; they were able to send their kids to college;*

edozein adin duen giza ondorengoa (seme edo alaba): familia bitzita normala nahi dugu, hobeki biziti nahi dugu, eta gure haurrak handitzen ikusi nahi ditugu;

ili-30-10373998-n mn 99

[factotum](#)
[Animals](#)

[person_1](#)
[person](#)

[familyRelation+](#)

[Function+](#)
[Human=](#)
[Object=](#)

[eng-30-10373998-n](#) 43 [offspring_1](#) [progeny_1](#) [issue_8](#)

[eus-30-10373998-n](#) 43 [ondorengo_4](#)

[spa-30-10373998-n](#) 43 [descendiente_1](#) [vástago_3](#) [descendencia_1](#)

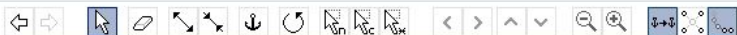
the immediate descendants of a person: *she was the mother of many offspring; he died without issue;*

Hiztegiak eta interakzioa

WordVis [About](#)

WordVis
the visual dictionary

+ options



basque

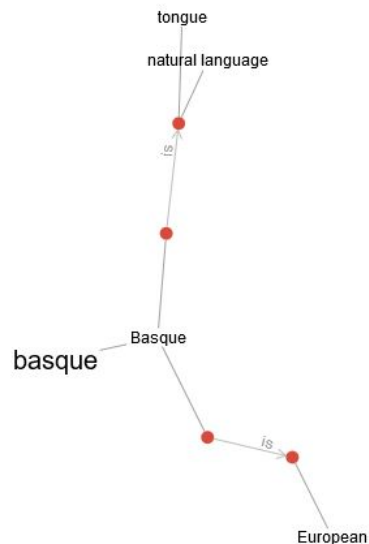
Noun

a member of a people of unknown origin living in the western Pyrenees in France and Spain

the language of the Basque people; of no known relation to any other language

a human written or spoken language used by a community; opposed to e.g. a computer language

a native or inhabitant of Europe



Konbinatoria

antzeko lemak

leizor [63] alderdi [164,783] ihardokitzaille [27] urteroko [3,311] kultura [100,259]
 gonazale [72] euskara [196,687] gazte [163,681] pollit [2,854]
 abenturazale [378] euskalzale [194] peto [2,252]
 jokozale [46] euskaltzain [3,854] apez [7,904] idazle [57,993]
 athleticzale [148] dantzari [8,619] eskulangile [979] elebakar [1,198]
 emakumezale [36] pilotazale [1,184] jakintsun [323] umezale [34] subiranista [1,201]

konbinazioak beste lemekin

	pisua	ager.
euskaltzale biltzar	12,45	405 ▶
euskaltzale topagune	11,96	143 ▶
euskaltzale euskaltzale	11,63	50 ▶
inistor euskaltzale	11,39	4 ▶
euskaltzale suhar	10,84	62 ▶
euskaltzale begilabur	10,62	4 ▶
euskaltzale ekargune	10,01	39 ▶
euskaltzale bilgune	9,61	47 ▶
euskaltzale amorratu	9,53	55 ▶
euskaltzale porrokatu	9,44	22 ▶
bage euskaltzale	8,96	8 ▶
euskaltzale ekidistante	8,91	2 ▶

Sintaxia automatikoki aztertzeko tresnak

- Aztertu honako esaldiko izen sintagmak eta aditz sintagmak, azentuak eta intonazioa markatzeko: lxati
 - Nire lagun [

Izen Sintagma Aditz Sintagma

Eraitza:

Nire lagun batek biolina jotzen du, oso belarri fina baitu .

Nire lagun batek BIOLINA jotzen du, oso belarri fina baitu.

oO-oo-oO

+2-1

oOoo ooO,

+2-1 ,

oO-ooo-oo oO.

+2-1.

BIOLINA

ooOo

-2

belarri

oOo

-2

Ereduak ikasteko eta jolasteko erabil daiteke

Sintaxia automatikoki aztertzeako tresnak

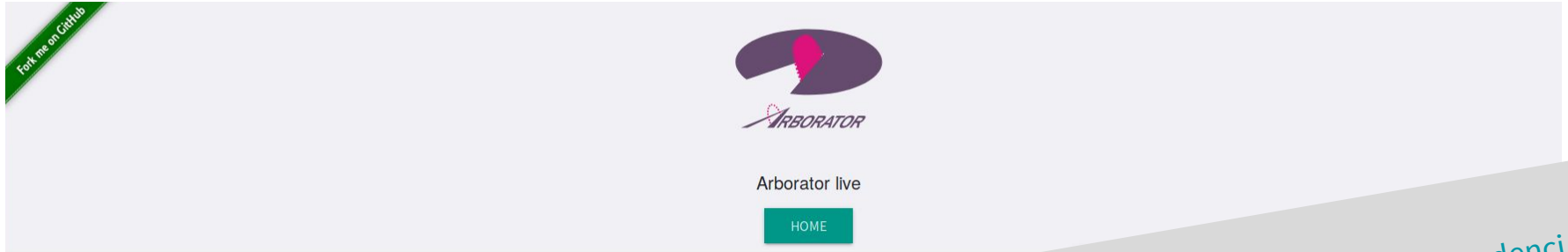
- Aztertu honako esaldiaren zuhaitz sintaktikoa [Maltixa](#)rekin
 - Nire lagun batek biolina jotzen du, oso belarri fina baitu.

Syntax analysis of: Nire lagun batek biolina jotzen du, oso belarri fina baitu. is:


Index	Word	Head	Relation	Lemma	Features	Category	Subcategory
1	Nire	2	ncmod	ni	KAS:GEN NUM:S	IOR	IOR_PERARR
2	lagun	5	ncsubj	lagun	_	IZE	IZE_ARR
3	batek	2	detmod	bat	KAS:ERG	DET	DET_DZH
4	biolina	5	ncmod	biolin	KAS:ABS NUM:S	IZE	IZE_ARR
5	jotzen	0	ROOT	jo	ASP:EZBU	ADI	ADI_SIN
6	du	5	auxmod	*edun	_	ADL	ADL
7	,	6	PUNC	,	_	PUNT	PUNT_KOMA
8	oso	10	gradmod	oso	_	ADB	ADB_ARR
9	belarri	11	ncobj	belarri	_	IZE	IZE_ARR
10	fina	9	ncmod	fin	KAS:ABS NUM:S	ADJ	ADJ_ARR
11	baitu	5	cmod	ukan	ERL:KAUS ASP:PNT	ADT	ADT
12	.	11	PUNC	.	_	PUNT	PUNT_PUNT

Arborator: zuhaitz sintaktikoak ikusteko

<https://arborator.github.io/live.html>



Fork me on GitHub

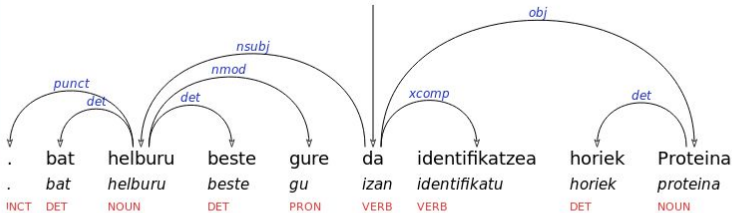


Arborator live

HOME

VIEW CONLL SAVE PNG SAVE SVG

. bat helburu beste gure da identifikatzea horiek Proteina



Euskarazko adibideak:

https://raw.githubusercontent.com/UniversalDependencies/UD_Basque-BDT/master/eu_bdt-ud-dev.conllu

Iturria:

https://universaldependencies.org/treebanks/eu_bdt/index.html

Stanfordeko parser sintaktikoa

Stanford CoreNLP 3.9.2 (updated 2018-11-29)

Iturria: <http://corenlp.run/>

— Text to annotate —

El rápido zorro marrón saltó sobre el perro perezoso.

— Annotations —

dependency parse ✕

— Language —

Spanish

Submit

Basic Dependencies:



Stanford eta ANCORA

<http://clic.ub.edu/corpus/es/busquedas>

- Downloads
- Documentation
- Publicaciones
- Competitions
- IARG-AnCora
 - CesCa
 - CitInt
 - Nident
 - Araknion Project
 - Paraphrasing
 - SFU Review-NEG
- Hopinion
- HistoCat
- DialCat
- EsPal
- CoMIT
- Lexicons
 - AnCoraVerb_ES
 - AnCoraVerb_CA
 - AnCoraNom_ES
 - AncoraNet

User Login

Username: *

Password: *

- Create New Account
- Request New Password

sentenceno: 3 filename: eu-xml_eu-xml_eu-xml_eu_egun.07-2-p0001500.2000-07-02.poltika.tbf.xml Node name: <sentence>

[view sentence](#) [view file](#) [view graphic](#) [download file](#)

```
graph TD
    sentence[S] --> s1[s]
    s1 --> s1_s[s]
    s1_s --> s1_p[p]
    s1_p --> s1_n1[n]
    s1_n1 --> s1_n2[n]
    s1 --> gv1[gv]
    gv1 --> v1[v]
    v1 --> s2[S]
    s2 --> s2_s[s]
    s2_s --> s2_gn[grup.nom]
    s2_gn --> s2_n[n]
    s2_n --> s2_es[espec]
    s2_es --> s2_d[d]
    s2 --> sn1[sn]
    sn1 --> s2_gn
    s2_gn --> s2_n
    s2 --> coord[coord]
    coord --> c[c]
    c --> sn2[sn]
    sn2 --> s2_gn
    s2_gn --> s2_n
    s2 --> s3[s]
    s3 --> r[r]
    r --> s4[s]
    s4 --> s4_gn[grup.nom]
    s4_gn --> s4_n[n]
    s4 --> s5[s]
    s5 --> word[word]
    word --> gv2[gv]
    gv2 --> v2[v]
    v2 --> s6[S]
    s6 --> s6_s[s]
    s6_s --> s6_gn[grup.nom]
```

func="suj" | func="cc" | "Euren" | lem="eurak" | grup.nom | "helburua" | lem="helburu" | gv | "da" | lem="izan" | func="atr" | func="cd" | func="cc" | grup.nom | "ekimena" | lem="ekimen" | espec | "hauen" | lem="hauek" | sn | grup.nom | "lekukoa" | lem="lekuko" | coord | c | "eta" | lem="eta" | sn | grup.nom | "ardura" | lem="ardura" | s | func="cc" | r | "ondoren" | lem="ondoren" | s | func="suj" | grup.nom | "Udabilitzak" | lem="Udabilitza" | s | func="atr" | word | "berak" | lem="bera" | gv | v | "egitea" | lem="egin" | s | func="cc" | S | func="cc" | s | func="cd" | grup.nom

Esanahia: proposizioak

- DRTn (Discourse Representation Theory) esaldien eta diskurtsoko esanahia deskribatzen da: anafora, denborazko adierazpenak eta auresuposizioak
 - Logikako atazak egiteko: [DRT demo](#)

DEMO

Sentence:

My friend plays the violin.

Parse

DRS(PERSON(X1) OF(X2 X1) friend(X2) violin(X3) play(E1) AGENT(E1 X2) CO-AGENT(E1 X3))

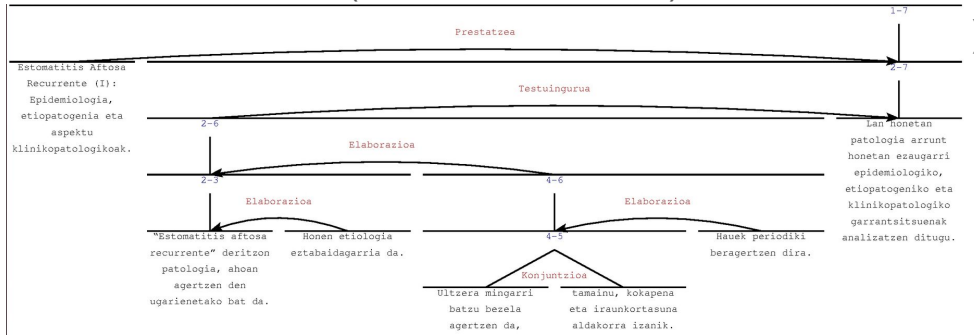
x_1, x_2, x_3, e_1
PERSON(x_1), OF(x_2, x_1), friend(x_2), violin(x_3), play(e_1), AGENT(e_1, x_2)
CO-AGENT(e_1, x_3)

DRS

PERSON(X1) OF(X2X1) friend(X2) violin(X3) play(E1) AGENT(E1X2) CO-AGENT(E1X3)

Esanahia: diskurtso-egitura

- [DiZer](#) analizatzaileak RST (Rhetorical Structure Theory) darabil testuen erlazio-egitura deskribatzeko
- Euskararako [erlazio-egitura bilatzailea](#):
 - [EusEduSeg](#): Diskurtso segmentatzailea
 - [UZ detektorea](#): unitate garrantzitsuenaren bilatzailea (testu zientifikoetan)



What is it?

DiZer 2.0 is a web interface for discourse parsing. It is based on DiZer (Pardo and Nunes, 2008), the first discourse parser for Brazilian Portuguese. The system aims at producing the discourse structure of a source text following the Rhetorical Structure Theory - RST (Mann and Thompson, 1987), one of the most used discourse theories in Computational Linguistics and Natural Language Processing.

DiZer 2.0 also allows the customization for other languages, being minimally necessary a discourse segmenter and a list of discourse patterns, which correlate text superficial markers to RST characteristics. DiZer 2.0 is currently customized for Brazilian Portuguese and Spanish.



Development

Núcleo Interinstitucional de Linguística Computacional (NILC)
 Instituto de Ciências Matemáticas e de Computação (ICMC)
 Universidade de São Paulo
 Erick G. Maziero, Thiago A. S. Pardo



Traitement Automatique de la Langue Naturelle Ecrite (TALNE)
 Laboratoire Informatique d'Avignon (LIA)
 Université d'Avignon et des Pays de Vaucluse
 Iria da Cunha Fanego, Eric SanJuan, Juan-Manuel Torres-Moreno



Grup Iulaterm
 Institut Universitari de Lingüística Aplicada (IULA)
 Universitat Pompeu Fabra
 Iria da Cunha Fanego

