

CLARIN Europako sarea: eHumanitateak eta zientzia sozialak lankidetzarako behar duten hizkuntza-azpiegitura sortzen

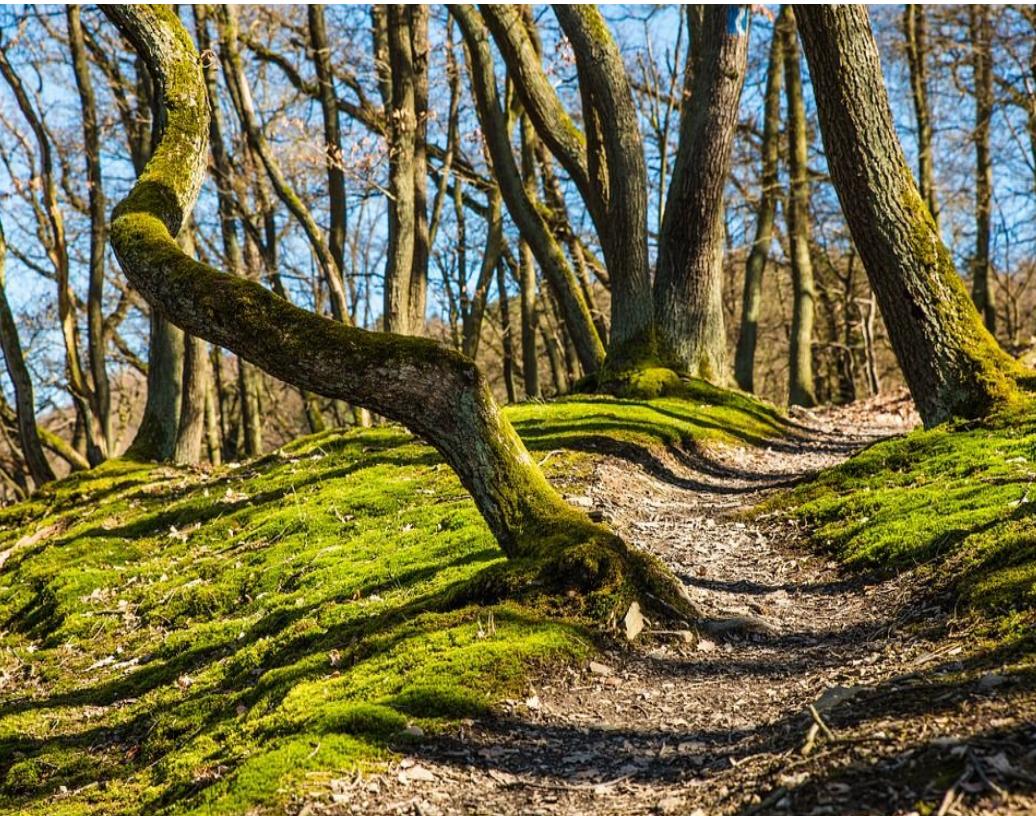
www.clarin.eu

www.clarin-es.org

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

Mikel Iruskieta
Ixa Taldea - HiTZ zentroa
UPV/EHU





Hitzaldiaren gidoia

- a) Azpiegituraz: CLARIN
- b) Helburua
- c) Testuingurua
- d) Zenbait proiektu



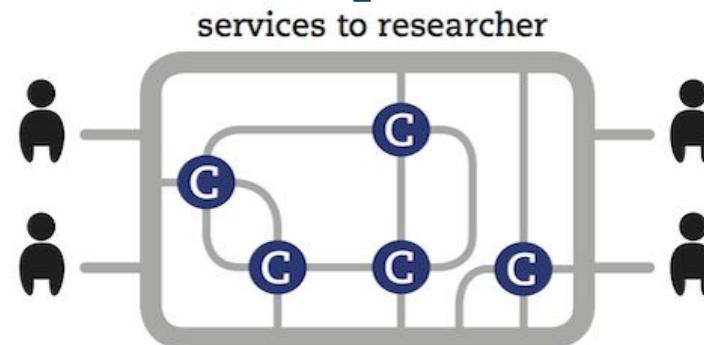
- Common LAnguage Resources and Technology INfraestructure
 - ESFRI ERIC (2012) eta ESFRI Landmark (2016)
- Ikerketa azpiegitura honen helburua da Europako hizkuntza guztien **baliabide digitalak webgune bakarrean** eskuragarri egitea, Humanitate eta Gizarte Zientzietako ikertzaileei laguntzeko
 - Datu linguistiko digitalak (testuak, ahozkoak, multimodalak)
 - Tresna aurreratuak: deskribatzeko, aztertzeko, konparatzeko



- Hizkuntza-teknologiaren **erabilera erraztu** eta aholkatu
- HD egin eta HDkoekin egin
- Norekin? Proiektu zehatz eta corpus landuak dituenarekin

Humanitateak eta Gizarte Zientzietako sarbidea eman

- Webgune bakarretik
 - **Datuak edonon egonik**
- Hizkuntza-datu digitalak
 - Idatziz, ahoz, bideoz
- Tresna aurreratuak
 - bilatu, aztertu, konbinatu, sortu...





K-zentroa euskararentzat

USA



- ERIC members
- Observers
- Countries with participating centres
- Centre Providing Data
- Centre Providing Metadata
- Knowledge Centre

EUROPE





Helburua

[Yann, Matxitxakoko itsasargia, CC BY 2.0](#)

Helburua

- Hausnartzeko eHumanitateek azpiegitura zabalik behar duten EHn
 - CLARINen adibidea azaldu



Bide batez, CLARIN norbaiti interesatuko balitzao,
horren berri emateko eta gure zerbitzuak eskaintzeko
edota lankidetza aztertzeko



Parte hartu
INKESTAN
<https://labur.eus/6PZzT>

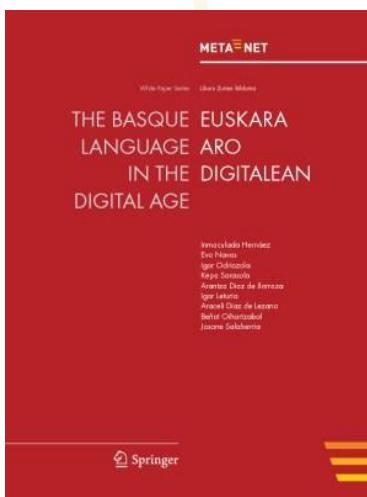


Testuingurua

Pixabay: [cstibi](#)

Hizkuntza-teknologia eta garapena

- 5 egoera nagusi 2012an
 - Egoera bikaina:
 - Egoera ona: ingelesa
 - Egoera ertaina: alemana, espaniera, frantsesa, holandesa...
 - Egoera osagabea: **euskara**, esloveniera, suediera, daniera...
 - Egoera apala: irlandera, islandiera, errumaniera...

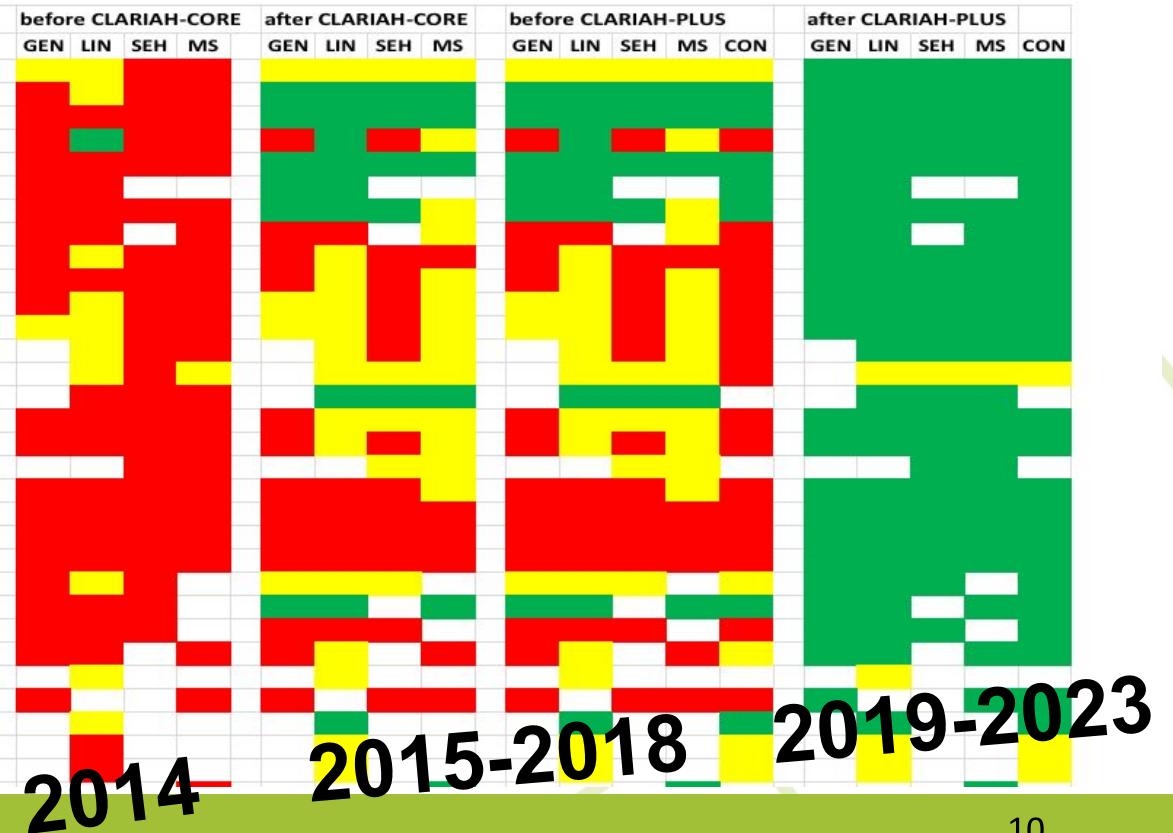


Egoera ertaineko Herbereetako eredua

GENeric functionality, LINguistics, Socio-Economic
History and Media Studies, CONtent of texts: history, literary
ITURRIA:
<https://www.clariah.nl/over/bestanden/downloads/send/10-folders/166-clariah-plus>

Figure 1. Progress Matrix CLARIAH

planning item
Persistent Identifier Framework
Federated Authentication
Federated Authorisation
Metadata Framework (formats, tools etc)
Ingestion, storage, linking of Linked Data (Persons, locations, Concepts)
Dutch Diachronic Thesaurus
procedure for standards, guidelines for software, data sustainability, etc
Secure Processing Environment
High Performance Computing
Virtual Collections
tools for browsing and search for data
select resource (fragment)s for Virtual Collection
integrate KB holdings
research data curation
Tools for the Conversion of (meta)data to linked data
tool metadata
Facilitating Crowdsourcing
Text to Structured data
Open Annotation Environment
Off-line Annotation Store
Sand Box
image analysis
OCR and OCR-correction
OCR in a linguistic enrichment workflow
HTR
Manual Annotation Tools text
Manual Annotation Tools Linguistics
Manual Annotation Tools Images
automatic token annotation Modern Dutch
automatic token annotator generation
automatic token annotation Historical Variants of Dutch



Egoera osagabeen dauden ereduak

	% PIB 2016	3.12	0.27	1.87
	%	Suecia	%	Dinam.
National coordination	17.00	537,653	21.00	8,510
Support/admin work	39.00	1.233,440	10.64	4,255
Technical support, repository	21.00	664,160	23.40	9,361
Technical support, ling. and com.	10.00	316,266	28.37	11,347
Outreach, curation or resources	10.00	316,266	16.31	6,524
CLARIN fee, approx	3.00	94,880	3.00	13,000
Total	100.00	1,929,226		53,000
Computers				4,000
Co-financing				28,000
special activities: works.				22,107
Research projects			20,000	
Total		1,929,226		73,000
				250,107

Honako iturrian oinarritutako estimazioak:

<https://office.clarin.eu/v/CE-2017-1052-CLARINPLUS-D5-3-4-costs.pdf>

CLARIN technology - an introduction

Intro to technology

The technical pillars of the CLARIN infrastructure are:

- **Federated identity** - letting users login to protected data and services with their own institutional login and password
- **Persistent identifiers** - enabling sustainable citations of electronic resources
- **Sustainable repositories** - digital archives where language resources can be stored, accessed and shared
- **Flexible metadata and concept definitions** - to ensure semantic interoperability when describing language resources
- **Content search** - offering a search engine for a wide range of language resources
- **Web service chaining** - giving users the possibility to freely combine language processing services

The [Services section](#) gives an impression of how all these technological components are combined into ready-to-use packages for our scientific community.

Organization

CLARIN is based on a distributed network of organizations ('centres') that host language resources and related services. Currently there are [38 of these centres](#) – mostly in Europe – each with its own expertise. Within a single country these centres are grouped into a [national consortium](#).

Each consortium has appointed one centre as a representative in CLARIN's technical body, [the Centre Committee](#). That is were most of the technical work happens: writing [specifications](#), planning [software development](#) and organizing the quality control for each of the centre candidates. The independent [Center Assessment Committee](#) will analyze each of the candidate centres and provide feedback with regards to compliancy to the technical and organizational requirements.

Learning more

If you want to learn more about the technology behind CLARIN, there are several sources of information:

- the [web page of the centre committee](#) contains documents about the planning and implementation;
- the [formal specification documents](#) are available from a single page;
- the [Trac and Subversion](#) are used for most of the software development;
- the dev and all-centers [mailinglists](#) are handy to stay informed about recent developments.

[Component Metadata](#)

[Content Search](#)

[Federated Identity](#)

[Persistent Identifiers](#)

[Repositories](#)

[Specification Documents](#)

[Development Information](#)

[Status: service monitor](#)

Zerbitzuak ez dira soilik
ikertzaileentzat, baizik eta bestelako
azterketatxoak egiten dituzten HDko
lagunentzat. Kasu batzuetan, ordea,
sarbide mugatua dago.

Knowledge Sharing

The Knowledge Sharing Infrastructure

Research infrastructures based on distributed data have three main components:

- At the heart is the technical infrastructure, comprising both the technical facilities that provide users access to data and tools, and the people that operate those facilities.
- A set of commonly agreed upon organizational rules, measures and conventions that should ensure seamless interaction between infrastructure users, operators and components, including e.g., standards, access, licences, quality assurance etc.
- A set of measures and facilities aimed at securing a continuous transfer of knowledge between all players involved in the construction, operation and use of the infrastructure. Each of them needs knowledge and expertise to do their jobs, and they generate new knowledge and expertise as they go along. This we will call the Knowledge Sharing Infrastructure, abbreviated KSI.

All players in CLARIN, whatever their role is, need knowledge and expertise to do their jobs. Infrastructures such as CLARIN, as well as its users, are distributed all over Europe (and even beyond), in dozens or maybe even hundreds of locations. The mission of the CLARIN KSI is to ensure that the available knowledge and expertise does not exist as a fragmented collection of unconnected bits and pieces, but is made accessible in an organized way to the CLARIN community and to the Social Sciences and Humanities research community at large.

FIND OUT MORE ABOUT...



Knowledge centres



Workshops



Digital Humanities Course Registry



CLARIN Mobility Grants



Oral history & technology

Projektu
eredugarriak
pausoaz pauso



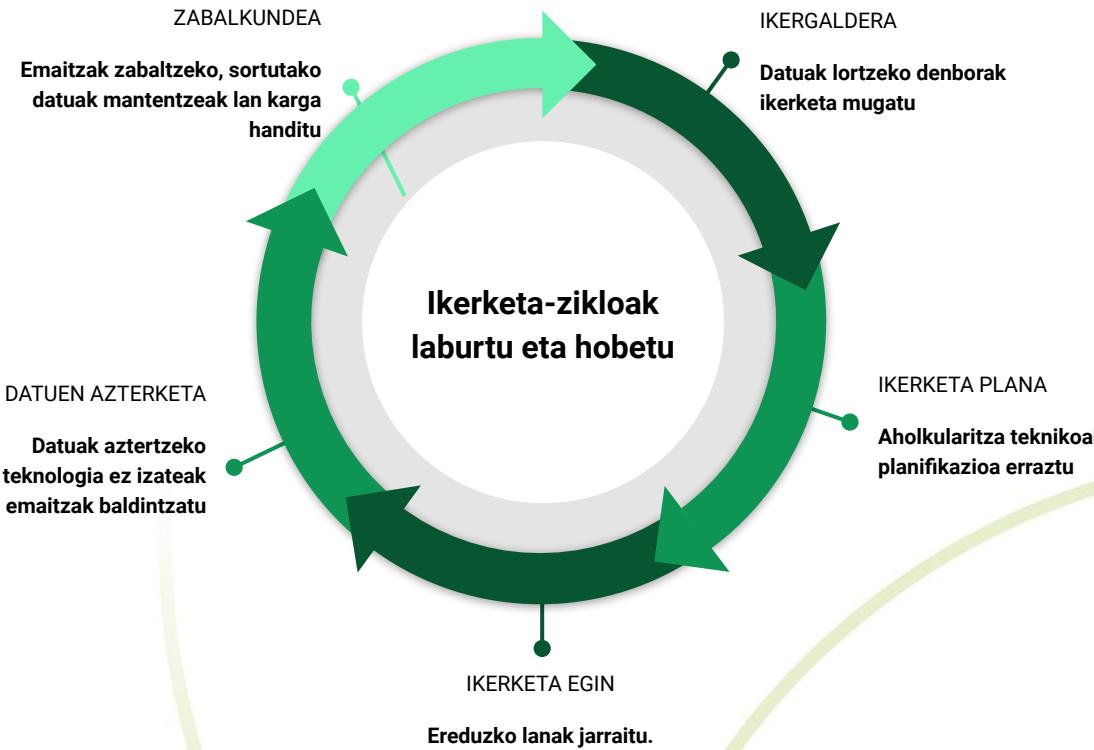
Videolectures

Ikastaroak



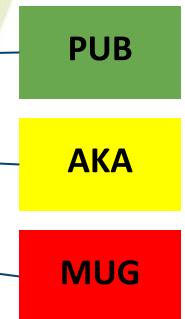
Annual Conference

Azpiegitura bat ez da proiektu bat gehiago



- Corpus irekiak eta **publikoak**
- Testuinguru **akademikorako** corpusak
- Eskabidea bete duten erabiltzaile **mugatuentzako**

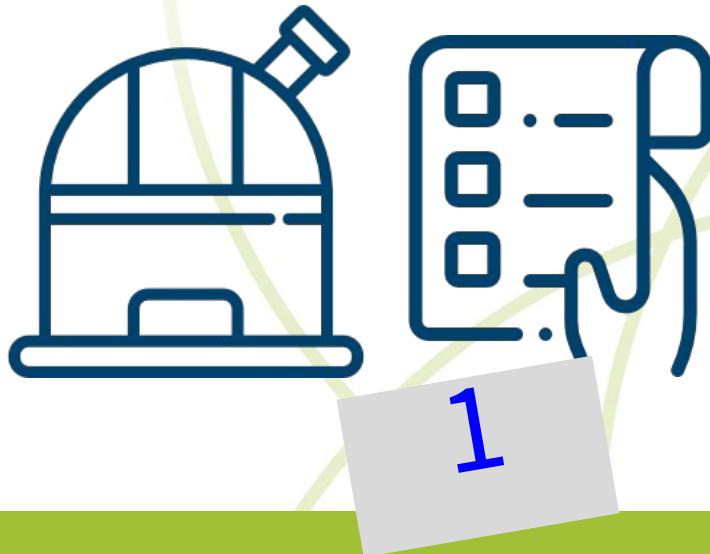
- Euskarazko baliabideak webgune bakarrean jaso
- **Estandarrak** eskatuz, baliabideak jaso eta mantendu
- Paradigma aldatzeko aholkularitza eman
- Behar diren baliabide **estrategikoak** landu
- Sarbide konfederatua



Language resources

Corpus eta metadatu kopuru handietan eta ezberdinetan bilaketa azkarrak egiteko aukera:

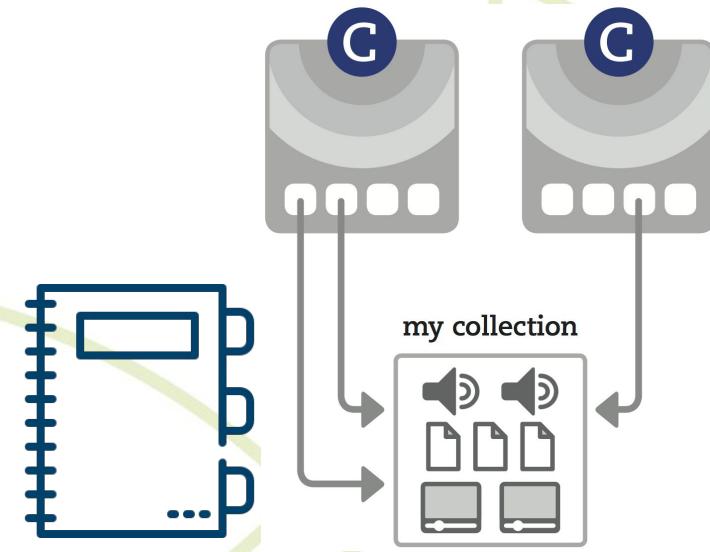
- vlo.clarin.eu/#tour
- contentsearch.clarin.eu
- <https://labur.eus/gZ1ld>



Virtual collections

Besteen lanetatik sortutako ikerketa-datuak aipatzeko modua:

- <https://collections.clarin.eu/app/public?16>



2

Language Resource Switchboard

Hizkuntza-datuak aztertzeko
behar duzun tresna bilatu eta
aztertu

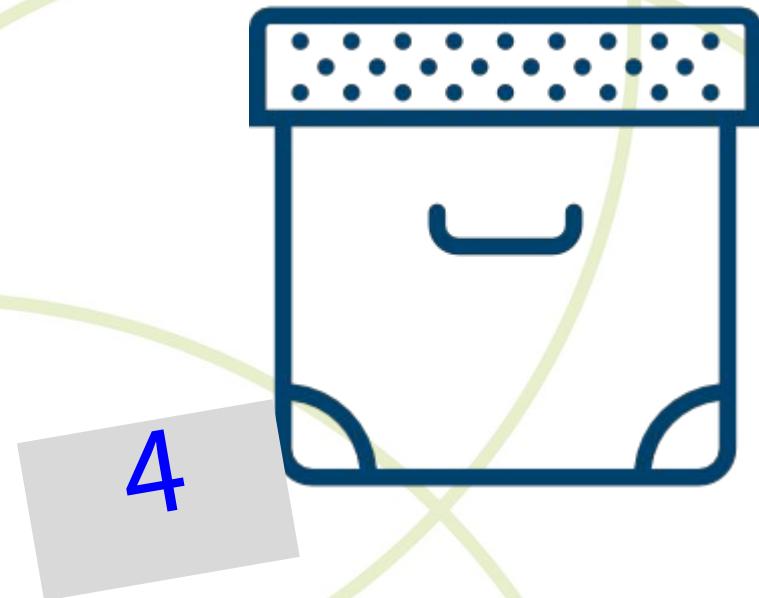
- <https://switchboard.clarin.eu/>



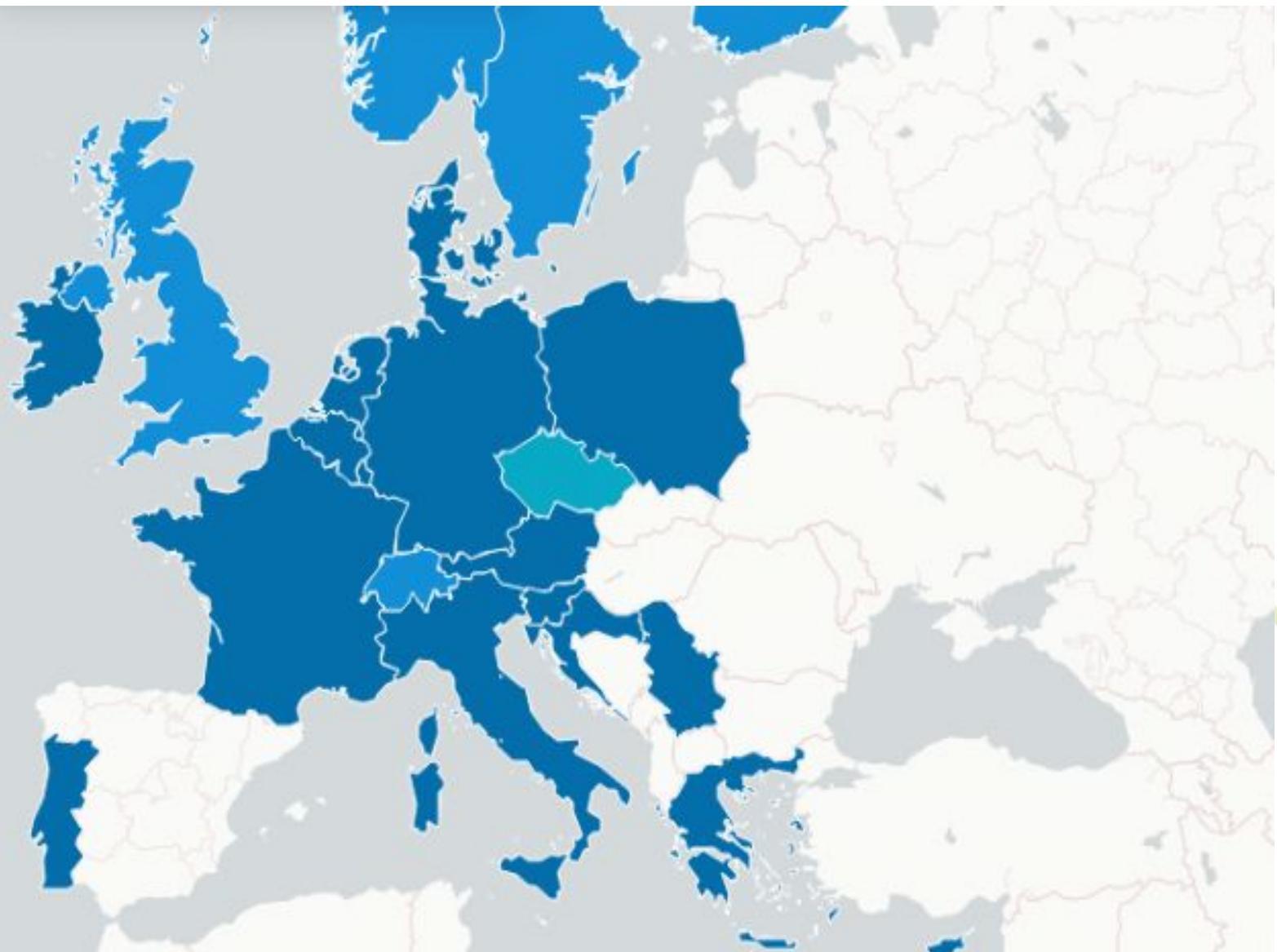
Depositing services

Corpusak eta baliabideak
gordetzeko aukera:

- [www.clarin.eu/content/d
epositing-services](http://www.clarin.eu/content/depositing-services)



Baina, zein da euskararen egoera CLARINen?



CLARIN-K zirrikitua... CLARINen parte hartzeko?



Algunas herramientas



Universitat
Pompeu Fabra
Barcelona



CLARIN-K zentroko helburuak

- Europa-mailako lankidetza bultzatzea, eHumanitateetan eta Giza Zientzietan behar diren tresna eta baliabideak eskaini ahal izateko
- CLARIN azpiegitura izan daiteke eHumanitateak eta Giza Zientziei laguntzeko moduetako bat, honela:
 - **Sarrera** bateratua, elkarbanatua eta lankidetzazkoa hizkuntza baliabide sakabanatuera
 - **Datu prozesamendu** esanguratsua paradigma aldatzeko eta ikerketa-zikloak laburtzeko...
 - **Datu linguistiko digitalak** (ahozkoan, idatzizkoan edo bideoan) lortzea, txukuntzea...
 - **Tresnak** aztertzeko, esplotatzeko, etiketatzeko eta konbinatzeko

CLARIN-K zentroko ebaluazioa (2015-7)

- Webgunetik zerbitzuak (2016tik): www.clarin-es.org
 - 2 egunetan lehen erantzuna
- Tresna errazen garapena
 - [Analhitza](#): testu-analisia
 - [Compress-eus](#): laburpenak kudeatzeko
- Akordioak:
 - i) EH Ikastolen Elkartea
 - ii) EGA (Hezkuntza saila)
 - iii) Ikertzaileak
- Bi tesitan laguntza
- Argitalpenak (9+):
 - Bel, N. Gonzalez-Blanco, E. Iruskieta, M. (2016). [CLARIN Centro-K-español](#). Procesamiento del Lenguaje Natural 57: 151-154. ISSN: 1135-5948.
 - Iruskieta, M. Bel, N. (2017). [CLARIN-K Centre Spain: una infraestructura orientada usuario](#). LINHD-UNED. Escuela de Verano HD.

Hasiera besterik ez da, dauden
baliabide murritzakin egina...
egun dauden aukerak
azpimarratzeko.





Zenbait proiektu
martxan
(bukatu gabe)

Pixabay: [cstibi](#)

IXA-CLARIN: euskararako CLARIN zentroa



Hizkuntza-teknologia garatzeko eta zabaltzeko jakintza zentroa
CLARIN-ERICek onartua

Hasiera Gure produktuak Bideoak Ekimenak Zerbitzuak Kontak

CLARIN CENTRE K

EU EN ES

Testu analisi orokorra:



Analhitza erabili

- Hitzak eta letrak kontatu
- Hitz eta lema desberdinak ateratzen
- Hitzak kategoriaka (estatistikak)
- Izen bereziak eta bestelako entitateak



Laburpenak egiteko tresna:

Compress eu

Compress eu erabili

- Eskolako laburpenak jasotzeko tresna, diskurtso-egiturako informazioan oinarritu:
 - Estrakziozko laburpenak
 - Abstrakziozko laburpenak

○ Campaña de divulgación de CLARIN-K

- Humanitate digitaletan aukerak (sarean.eus, Alegria eta Agirre, 2018-11-22)
 - CORPUS HISTORIKOEN PROZESAKETA (Jardunaldi irekia, 2018-06-11)
 - Atro Voutilainen-en hitzaldia. Zuhaitz-bankua finlandierarako.(2011/06/08)
 - Euskararen ingurune digitala

Ikusi albiste gehiago

twitter

CLARIN Tutorial at AISV2019 in Arezzo: blog post written by Fabio Ardolino, Silvia Calamai, Letizia Cirillo, Duccio... <https://t.co/fwAw5W8R1u>

RT @SSHOpenCloud: Thanks for taking part in the #SSHOCKickOff! We had two days of interesting discussions and work sessions. The #SSHOpenCI...

RT @scorpiommma: Great reminder that we kind of celebrate the 30th anniversary of the Web at #SSHOCKickOff making more scientific clouds pl...

Gehiago irakurri...

Clarin K-centre taldearen helburuekin bat eginez, giza-zientzia eta zientzia sozialetako ikerketan lagungarri diren hizkuntza-teknologietako hainbat tresna, lan egiteko modu berri, baliabide, artikulu, bideo eta bestelako informazioa aurkituko dituzu webgune honetan. Ikergaldera eta corpora izanez gero, hizkuntza-teknologiko tresnaren bat moldatu edota elkarlanean berria sortzea aukera ere eskaintzen dizugu.

ANALHITZA: testu azterketarako tresna erraza

- Erabilerraza eta euskararako Ixa Taldeak garatu dituen zenbait teknologia integratuta
- Emaitza moldagarria: erabiltzaileak kalkulu-orria moldatu



ANALHITZA
Powered by CLARIN

Ikas-elkarreko ipuineta n-gramak
otsoaren pertsonaia nolakoa den
jakiteko

<http://ixa2.si.ehu.es/clarink/analhitza.php?lang=eu>

Maizt.	Ezker.	Kat.	Eskum.	Kat.
5	otso	N	gaizto	G
1	otso	N	beltz	G
1	otso	N	handi	G
1	otso	N	maltzur	G
1	otso	N	goseti	G

BIM projektua



Sintaktikoki etiketatutako
euskalgorako corpus historikoa

Bilaketa: gutxi Mota: Lema Q Bilatu

17 emaitza, 3 testutan

Axular Gero [5]

...gutigatik ere bortitzki gatzigatzen zuela (Plutarc. in vita Caton. Censor.). Katon ...gora ekkarriko du, burupe izanen du eta nehoren **guti** beldurrik, iendarera, b ...zure aita launa, bere etxeaz, onez eta biziaz ere kontu **guti** eginik, ioan zen Do ...gauza **guti** edireiten den euskalgoraz eskiribaturik, gogain behartu naiz eta beldu ...heriotzea bezain gauza segurik, eta orena noiz izanen den bezain **guti** gerthuri?

« 1 »

Escualdun Cocinera [11]

...eltçian trempatu, tipula **guti** eta echalota pochi batekin egunice : hachicha frita ...chehatuac, olio fina bianan **guti**. Ohe hortan, anthola çatçu makallau puchcac e ...onxa chehatuic, charbola pochi bat, biper belitcha, gatça, citroin yusa, orotari ...denian. Ahatia yusian. Cheha çatçu hirur tipula, hirur carrota, celeria **guti** bat ...tipi batian chingar puchca batekin eta canela **guti** batekin egosi denian pasa

« 1 2 3 » + Ikusi guziak

Igarkizunak (R.M. Azkue) [1]

...batek bildurikoak ere **guti** batzuk baditut. Gainerako guziak eztute orain

« 1 »

Euskalgorazko testuak (ez-estandarrak)

- Testamentu berria (Leizarraga, lap. eta behenap.)
- Refranes y Sentencias (ezezaguna, biz.)

16 IRAC VRTÇAILLEARI:

tu-eragotz carri beçala, liburutto baten, bi partean partituric, guero, hunen gañican eguitara.

Eta nahi nituzquçyen bi parteac elccarrequin, eta batetan athera Baiña iccuñiric cein gauça guti edireiten den euskalgoraz esquiribaturic, gogain behartu naiz eta veldurtu, ez-tiren bideac asego segur eta garbi, baden bi-tartean, cembai trabu edo behaztopa-harri. Eta halatan hartutud gogo, lehenbicico parte hunen, lehenile benturatzeko, eta berri iakitera bezala aitzinerat igortzeko. Hunek zer irragaiten den, zer begitarte izalten duen, eta nor nola mintzo den, abisu eman diazadan. Gero abisu haren arauaz ethorkizunerat gobernatzeako: eta bigarren partearen kanporat atheratzeko, edo barrenean gelditzeko eta estaltzeko. Badakit anhitzek miretsiko duela eta ediren arrotz eta estraiñio, ni lan huni lotzea. Zeren anhitz izan baita orainokoan, eta baita orai ere, ni ez bezalakorik, ni bañio hunetako gaiagorik, eta anzatsuagorik, ezpaitute guziarekin ere, orainokoan, hunelako materiatan, hunela ausartziarik eta eskudantziarik harta. Baitirudi ezen asko behar lizatekeiela arrazoin haur ene gibelatzeako eta geldi aratzeko ere. Baiña ene kontra dela dirudien arrazoin hunek beronek, ni esportzatzen eta aitzinatzent nau, hu-

Badaquit anhitzeko miretsicoduela eta ediren arrotz eta estraiñio, ni lan huni lotcea. Ceten anhitz içan baita orainokoan, eta baita orai ere, ni ez bezalakorik, ni bañio hunetako gaiagorik, eta anzatsuagorik, ezpaitute guziarequin ere, orainokoan, hunelako materiatan, hunela ausartziarik eta esku dantziarik harta. Baitirudi ezen afo behar liçatequeyela arraçoin haur ene gibelatzeako eta gueldi araciteco ere. Baiña ene contra dela dirudien arraçoin hunec beronek, ni esportçatzen eta aitzinatzent nau, hu-

Laburpenen sailkapen automatikoa: Compress-eus (Atutxa et al. 2017)

Compress eus

Compress eus laburpenak egiteko tresna bat da.

Tresna hau erabiltzeko derrigorrezkoa da erregistratzea. Mesedez dokumentua laburtzeko erregistratu, [hemen](#) sakatuz
Compress eus nola erabili ikusteko ikusi ondorengo dokumentua: [pdf formatuan](#) edo [Power point formatuan](#)

Oharra:

Aplikazioak onto funtzionatzeko erabili **chrome** edo **firefox**. **Explorer**-ekin ez duela onto funtzionatzen detektatu dugu.
Barkatu eragozpenak.

[Gida azkarra erakutsi](#)

Testua berrasi

Laburpena ikusi eta editatu

Testua bukatu

Berrikusi duzu 0 orainarte eginiko testuak

GIZA-GORPUTZA ETA ELIKAGAIAK.

GIZA-GORPUTZA ETA ELIKAGAIAK.

Janari edo edari bakoitza, hau da elikagai bakoitza, substantzia edo
mantenugai jakin batzuez osaturik dago.

Mantenugai bakoitzak zeregin jakin bat du gure gorputzaren osaketan eta
funtzionamenduan.

Horregatik, janari batek beste batek baino hobeto erantzun diezai oke
gorputzaren behar jakin bat.

Betetzen duten funtzioren arabera, elikagaiak hiru multzotan sailkatzen dira.

Gorputzak, hazten ari denean, zelula eta ehun gehiago sortzen lagunduko
duten mantenugaiak beharko ditu.

Testu konprimitura

GIZA-GORPUTZA ETA ELIKAGAIAK.

Janari edo edari bakoitza, hau da elikagai bakoitza, substantzia edo
mantenugai jakin batzuez osaturik dago.

Mantenugai bakoitzak zeregin jakin bat du gure gorputzaren osaketan eta
funtzionamenduan.

Horregatik, janari batek beste batek baino hobeto erantzun diezai oke
gorputzaren behar jakin bat.

Betetzen duten funtzioren arabera, elikagaiak hiru multzotan sailkatzen dira.

Gorputzak, hazten ari denean, zelula eta ehun gehiago sortzen lagunduko
duten mantenugaiak beharko ditu.

Eginkizun hori, zelula berriak sortzea, alegia, betetzen
mantenugaiak dituzten elikagaien eraikitzaile deritze.

Ezin badira testu guztiak zuzendu,
zuzenketa automatikoa eskertuko
al du ikasleak?



Ikastolen Elkartarekin lankidetzan, 100
ikasleren 4 testuren laburpenak jaso eta
irakasleak egindako laburpenaren
araberako sailkapen automatikoa egin da

<http://ixa2.si.ehu.es/clarink/tresnak/compress-eus/>

Ikasgai ez-formalak lankidetzen teknologiez

HAP-LAP master ofiziala

(UPV/EHU)



HiTZ

Hizkuntza Teknologiako Zentroa
Basque Center for Language Technology



#hezikt graduondokoan
(UPV/EHU eta UEU)
hezkuntza- eta
hizkuntza-teknologiei buruzko
ikasgaiak parte hartu nahi
duen edonorekin lankidetzen

Ikastaroak laster!

<https://hezikt.eus/hizkuntza-teknologiak/>

IKTak eta konpetenzia digitalak hezkuntzan

UEUk eta UPV/EHUk elkarlanean antolatutako graduondokoa

HEZIKT ▾

ALBISTEAK

TEKNOLOGIAK ▾

IKAS-MATERIALAK ▾

IKASLEEN LANAK

Hizkuntza-teknologiak

2019-03-04

Hizkuntzaren prozesamenduan oinarritutako edo erabiltzeko aplikazioei deritze hizkuntza-teknologiak.

Gaur egun Hezkuntzan edukiak sortzeko edota kudeatzeko teknologien erabilera edo aipamen handia bada ere, hizkuntzazko edukiak baliatzeko teknologiaren erabilera urria da oso. Zer dela-eta?

Ondoko arrazoia aipa daitezke:

- Datuetan oinarritutako ikerketa gutxi
- Hurbilpen teorikoen eraginez
- Diziplinen arteko lotura eskasa: pedagogia eta hizkuntzaren prozesamendua
- Tresna batzuk erabiltzeko konplexutasuna edota kalitatea



Ikuspegi orokorra

Bilatu

Bilatu ...

Azken artikuluak

Humanitate digitalak: aukerak, erakundeen rol berriak eta elkarlana

Eman izena HEZikt graduondoko 4. edizioan

Mikel Iruskieta eta Abel Camacho Bikaintzuna Euskal Ikasketetan IX. Euskara eta euskal kultura eremu digitalean ikastaroan izango dira

Eta zuon ustez euskarak edota Euskal Herriak
behar al dute Humanitate Digitalak
bultzatzeko erakunde edo azpiegiturarik?

Erreferentziak

Atutxa, U. Iruskieta, M. Ansa, O. Molina, A. (2017). [COMPRESS-EUS: I\(ra\)kasleen laburpenak lortzeko tresna](#). EUDIA: Euskararen bariazioa eta bariazioaren irakaskuntza-III. 87-98

Bel, N. Gonzalez-Blanco, E. Iruskieta, M. (2016). [CLARIN Centro-K-español](#). Procesamiento del Lenguaje Natural 57: 151-154. ISSN: 1135-5948.

Etxeberria, I. Alegria, I. Uria, L. 2019. Weighted finite-state transducers for normalization of historical texts. *Natural Language Engineering* 25 (2), 307–321. <https://doi.org/10.1017/S1351324918000505>

Iruskieta, Mikel.; Isasi, Jennifer. (argitaratzeko). Humanitate Digitalak eta ikerketa ezberdinen azterketa. In *IKTak eta konpetentzia digitalak hezkuntzan*. UEU-UPV/EHU. Eibar.

Iruskieta, M. Bel, N. (2017). [CLARIN-K Centre Spain: una infraestructura orientada usuario](#). LINHD-UNED. Escuela de Verano HD.

Krauwer, S., & Hinrichs, E. (2014). The CLARIN research infrastructure: resources and tools for e-humanities scholars. In *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC-2014)* (pp. 1525-1531). European Language Resources Association (ELRA).

Váradi, T., Wittenburg, P., Krauwer, S., Wynne, M., & Koskenniemi, K. (2008). CLARIN: Common language resources and technology infrastructure. In *6th International Conference on Language Resources and Evaluation (LREC 2008)*.

Otegi, A. Imaz, O. Díaz de Ilarza, A. Iruskieta, M. Uria, L. (2017). [ANALHITZA: a tool to extract linguistic information from large corpora in Humanities research](#). Procesamiento del Lenguaje Natural 58, pp. 77-84.

CLARIN Europako sarea: eHumanitateak eta zientzia sozialak lankidetzarako behar duten hizkuntza-azpiegitura sortzen

www.clarin.eu

www.clarin-es.org

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

Mikel Iruskieta
Ixa Taldea - HiTZ zentroa
UPV/EHU

