

Presentació dels resultats del projecte CLARIN-CAT

El Departament d'Innovació, Universitats i Empresa de la Generalitat de Catalunya i la Universitat Pompeu Fabra van signar un conveni l'any 2008 per tal d'integrar recursos i eines lingüístiques en i per a la llengua catalana des de la fase inicial de construcció de la infraestructura europea CLARIN: *Common Language Resources and Technology Infrastructure*. Aquesta infraestructura ha d'oferir accés a dades i serveis de suport a la recerca en els àmbits de les humanitats i les ciències socials que treballen amb textos, i, específicament, donar suport a l'exploració de grans quantitats de textos. Els primers elements de la infraestructura CLARIN ja són visibles i accessibles a www.clarin.eu.



A CLARIN, l'Observatori Virtual de les Llengües (*Virtual Language Observatory*), l'Inventari de Recursos CLARIN (*CLARIN Language Resource Inventory*), l'Inventari d'Eines Lingüístiques CLARIN (*CLARIN Language Tool Inventory*) i el Catàleg de Metadades (*Catalogue of Harvested Data*) contenen informació sobre els recursos i tecnologies en català i per al català.

CLARIN
Common Language Resources and Technology Infrastructure

Home >

Tools inventory

Please [click here](#) to add a tool to the inventory. Number of tools that match the current filter criteria: 12
Please [click here](#) to add a resource to the inventory.

Type: <Any> Country: <Any> Input language: Catalan Output language: <Any> Name: Contains

Organisation: Contains Distribution Type: <Any>

Organisation (CLARIN member): <Any>

Apply

Name	Type	Country	Description	Organisation (not a CLARIN member)	Distrib Type
Alinea	toolbox	Spain	A tool for parallelizing translated texts, which has been specially designed for specialized corpora and also as a translation validator.	IULA - UPF	
freeing	annotation tool web service	Spain	Web service consisting of the Freeing open source language analysis tool suite.		
freeing_dependency	written language web service	Spain	Freeing-based dependency parser.		
freeing_morpho	annotation tool web service	Spain	Freeing-based morphological analyzer.		
freeing_parsed	written language web service	Spain	Freeing-based shallow parser.		
freeing_tagging	annotation tool web service	Spain	Freeing-based part-of-speech tagger.		

A més d'oferir informació sobre dades i recursos, CLARIN està construint la infraestructura necessària per facilitar l'ús i explotació de les dades i les tecnologies mitjançant dues innovacions crucials:

- 1) Compilació automàtica de corpus d'estudi ocasionals mitjançant un accés únic a dades textuais distribuïdes.
- 2) Explotació de dades textuais mitjançant l'accés per Servei Web a components tecnològics.

CLARIN-CAT ha participat també en la construcció d'aquesta infraestructura i ha posat ja a disposició dels usuaris una col·lecció de serveis lingüístics. La novetat rau en el fet que l'usuari pot enviar els textos que necessita analitzar per tal que, remotament, es facin les operacions que tria des del catàleg, sense necessitat d'instal·lar-se els programes. Aquestes operacions són, per exemple: comptar paraules d'uns textos, extreure seqüències de dues paraules (bigrames) més freqüents d'uns textos, lematitzar i desambiguar uns textos, etc.

CLARIN-CAT ja ha posat a disposició dels usuaris de la comunitat d'investigadors que treballen amb textos catalans la possibilitat d'usar aquests serveis. L'usuari ha d'enviar les dades segons les instruccions que trobarà al CLARIN-CAT-LAB, el laboratori virtual, i rebre'n els resultats.

CLARIN-CAT-LAB ofereix també accés a un nombre de repositoris de diferents característiques per permetre als usuaris disposar immediatament de dades per fer experiments. Gràcies a la col·laboració amb l'Institut d'Estudis Catalans, l'Arxiu de l'Ajuntament de Girona, el Centre de Documentació Ramon Llull i el Corpus Informatitzat del Català Antic, des del CLARIN-CAT-LAB hom pot accedir a les dades textuais d'aquestes institucions i recuperar-les per poder fer cerques i operacions com les descrites abans.

Per tal d'il·lustrar com els investigadors poden beneficiar-se d'aquestes facilitats CLARIN-CAT-LAB ofereix exemples d'investigacions que s'han dut a terme amb la infraestructura CLARIN. Aquests exemples mostren també l'ús dels serveis com una cadena de processos a mida.

Els serveis que ja es poden trobar a CLARIN-CAT-LAB donen accés a programes ja existents i reconeguts: etiquetadors morfosintàctics del FreeLing i TreeTagger, lematitzador de català antic de la Universitat d'Alacant, conversors de formats (de .pdf, .doc i .html a .txt), eines estadístiques per a l'anàlisi del vocabulari (freqüències, Informació Mútua, paraules clau, etc.), entre d'altres. CLARIN-CAT-LAB es posa ara a disposició de la comunitat d'investigadors catalana per tal de continuar donant informació sobre recursos i integrant accés a dades i a serveis de tots i per a tots els investigadors que treballen amb textos en català.

The screenshot shows a web interface for searching distributed corpora. At the top, the title "Accés a corpus distribuïts" is displayed in a large, bold font. Below it, a sub-header "Accés a corpus distribuïts" is followed by a link "» [Escriu la Consulta](#)".

The main search area is titled "Consulta:" and contains a search box with the query "dc.date > 1400". Below the search box, there is a "Resultats per pàgina:" label and a dropdown menu set to "4".

Below the search area, there is a button labeled "ENVIAR" and a link labeled "Ajuda".

At the bottom of the search area, there is a horizontal bar with four tabs: "GROA", "EC", "CDA", and "LB".

The footer of the interface features several logos: the Universitat de Girona logo, the Ajuntament de Girona logo, and the CICA (Corpus Informàtic del Català Antic) logo. A small "Powered by" logo is visible in the bottom right corner.



CLARIN-CAT
**Col·laborant en la construcció d'una infraestructura de
suport a la recerca
en Humanitats i Ciències Socials**

19-01-2011
Sala Coromines, Institut d'Estudis Catalans

- 18:00 Benvinguda. Isidor Marí, president de la Secció Filològica de l'IEC
18:10 Presentació CLARIN i CLARIN-CAT. Núria Bel (UPF)
18:30 Taula Rodona. Coordina: Mercè Lorente (UPF)

Els col·laboradors de CLARIN-CAT: els recursos i serveis integrats al demostrador

Dades:

Joan Soler – Institut d'Estudis Catalans

Joan Boadas – Arxiu de l'Ajuntament de Girona

Albert Soler – Centre de Documentació Ramon Llull (UB)

Joan Torruella – Corpus Informatitzat del Català Antic (UAB)

Eines:

Juan Antonio Pérez – Lematitzador de Català Antic de la Universitat d'Alacant (UA)

Lluís Padró – FreeLing (UPC)

Jorge Vivaldi – Eines de l'IULA (UPF)

- 19:30 CLARIN-CAT-LAB.

Els resultats de CLARIN-CAT ja a disposició de la comunitat. Marta Villegas (UPF).

- 20:00 La construcció de CLARIN a Europa. Steven Krawer, coordinador de CLARIN Europa (*en anglès*)
20:30 Clausura de l'acte.

En acabar, se servirà un refrigeri.