Linguistics with CLARIN Introduction

Jan Odijk

LOT Winterschool

Amsterdam, 2015-01-12

1

Overview

- CLARIN: Introduction
- CLARIN Infrastructure (focus on NL part)
  - CLARIN Portal
  - Find data and tools
  - Apply tools to data
  - Store data and tools
- Goals and Outline of this course

Overview

➢ **CLARIN: Introduction**

- CLARIN Infrastructure (focus on NL part)
  - CLARIN Portal
  - Find data and tools
  - Apply tools to data
  - Store data and tools
- Goals and Outline of this course
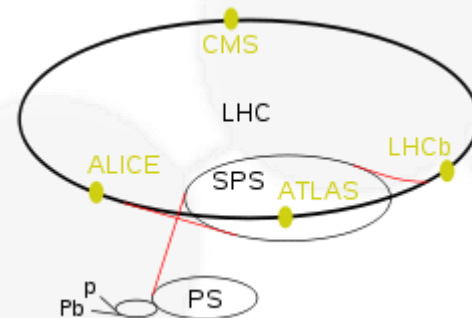
CLARIN Infrastructure

- A **research infrastructure** for **humanities researchers** who work with digital **language-related resources**

CLARIN Infrastructure

- Infrastructure:
  - (Usually large-scale) basic physical and organizational resources, structures and services needed for the operation of a society or enterprise
    - Railway network, road network, electricity network, …
    - eduroam

CLARIN Infrastructure

- Research infrastructure
  - Infrastructure intended for carrying out research: facilities, resources and related services used by the scientific community to conduct top-level research
  - Famous ones: Chile large telescope, CERN Large Hadron Collider

CLARIN Infrastructure

- humanities researcher
  - Linguists, historians, literary scholars, philosophers, religion scholars, ….
  - And a little bit in the social sciences: e.g. political sciences researchers

- Focus here on **linguists**

CLARIN Infrastructure

- Digital language-related resources
  - Data in natural language (texts, lexicons, grammars)
  - Databases about natural language (typological databases, dialect databases, lexical databases, …)
  - Audio-visual data containing (written, spoken, signed) language (e.g. pictures of manuscripts, av-data for language description, description of sign language, interviews, radio and tv programmes, …)

CLARIN Infrastructure

- Language in various functions
  - As object of inquiry
  - As carrier of cultural content
  - As means of communication
  - As component of identity

CLARIN Infrastructure

- CLARIN
  - Has not created any new data
  - Has mainly adapted existing data and tools
  - Has created new easy and user-friendly tools for searching, analysing and visualising data

CLARIN Infrastructure

- The CLARIN infrastructure
  - Is distributed: implemented in a network of CLARIN centres
  - Is virtual: it provides services electronically (via the internet)
- The CLARIN infrastructure
  - Is still under construction
    - Highly incomplete
    - Fragile in some respects
  - But you can use many parts already

CLARIN in Europe

- Prepared by [CLARIN preparatory project](#) (2008-2011)
  - coordinated by Utrecht University
- From Feb 2012 coordinated by the CLARIN-ERIC, hosted by the Netherlands
  - [ERIC](#): a legal entity at the European level specifically for research infrastructures
  - Other ERIC members: AT BG CZ DK DLU EE DE LT NO PL PT SE and growing

12

CLARIN in Europe

– CLARIN ERIC: www.clarin.eu
– [National Consortia](#)
  - Austria: [CLARIN+DARIAH Austria](#)
  - Bulgaria:
  - Czech Republic: [LINDAT CLARIN](#)
  - Denmark: [CLARIN Denmark](#)
  - Dutch Language Union: [CLARIN DLU](#) [CLARIN Flanders](#)
  - Estonia: [CLARIN Estonia](#)
  - Germany: [CLARIN-D](#)
  - Netherlands: [CLARIN-NL](#)
  - Norway: [CLARINO](#)
  - Poland: [CLARIN Poland](#)
  - Sweden: [CLARIN Sweden](#)
  - Lithuania:
  - Portugal:
  - …

Overview

- CLARIN: Introduction
- **CLARIN Infrastructure (focus on NL part)**
  - CLARIN Portal
  - Find data and tools
  - Apply tools to data
  - Store data and tools
- Goals and Outline of this course

CLARIN Infrastructure

- The CLARIN infrastructure offers services so that a researcher
  - Can find all data and tools relevant for the research
  - Can apply the tools and services to the data without any technical background or ad-hoc adaptations
  - Can store data and tools resulting from the research

  via one portal

Overview

- CLARIN: Introduction
- CLARIN Infrastructure (focus on NL part)
  - ➢ **CLARIN Portal**
    - Find data and tools
    - Apply tools to data
    - Store data and tools
- Goals and Outline of this course

CLARIN Infrastructure

- Portal
  - See next presentation

Overview

- CLARIN: Introduction
- CLARIN Infrastructure (focus on NL part)
  - CLARIN Portal
  - ➤ **Find data and tools**
  - – Apply tools to data
  - – Store data and tools
- Goals and Outline of this course

CLARIN Infrastructure
'*Can find all data and Tools*'

- Searching for Data: See next Presentation

Overview

- CLARIN: Introduction
- CLARIN Infrastructure (focus on NL part)
  - CLARIN Portal
  - Find data and tools
  - ➢ **Apply tools to data**
  - Store data and tools
- Goals and Outline of this course

Apply Tools to Data

- Illustration 1 (Search): Today, 2nd hour
- Illustration 2 (Search & Analysis): Tue 13th (OpenSONAR)
- Illustration 3 (Search, Dialect databases): Wed 14th (MIMORE)
- Illustration 4 (Enrichment & Search): Thu 15th (TTNWW, PaQu)

Overview

- CLARIN: Introduction
- CLARIN Infrastructure (focus on NL part)
  - CLARIN Portal
  - Find data and tools
  - Apply tools to data
  - ➢ **Store data and tools**
- Goals and Outline of this course

- What about your research data /software?
  - Make them CLARIN-compatible
  - →CLARIN tools and services apply to them
    - For analysis, improvement, creation
    - Others can use them more easily
  - Store them at a CLARIN Centre
    - For long term preservation
    - For easy access by you and others (e.g. via the VLO)
    - For verifiability and replicability of your research
  - More details: Fri 16th.

Overview

- CLARIN: Introduction
- CLARIN Infrastructure (focus on NL part)
  - CLARIN Portal
  - Find data and tools
  - Apply tools to data
  - Store data and tools

➢ **Goals and Outline of this course**

Goals of the Course

- Introduction to the CLARIN infrastructure
- Introduction to selected functionality within the CLARIN infrastructure
  - Focus on linguistics
  - Focus on results of CLARIN-NL
- Hopefully make you enthusiastic to use the CLARIN infrastructure and functionality in it
  - Because it improves your research
- If you use it, provide feedback
  - Helpdesk: helpdesk@clarin.nl
  - Service-specific user group lists

Course programme

Day	Who	What
Mon 12th	Jan Odijk	Introduction
		Illustrative Usage Case
Tue 13th	Jan Odijk	Search and Analysis with OpenSONAR
Wed 14th	Sjef Barbiers	Search and Analysis in micro-comparative databases with MIMORE
Thu 15th	Jan Odijk	Enrichment and Search (TTNWW, PaQu)
Fri 16th	Jan Odijk	Storing data in CLARIN
		Concluding Overview

26

Course programme

- For more details, see
  - http://www.clarin.nl/node/2044
  - http://www.lotschool.nl/files/schools/2015_Winterschool_Amsterdam_UvA/

Thanks for your Attention

And

Enjoy the Course!