

# CLARIN FOR LINGUISTS INTRODUCTION

Jan Odijk

LOT Summerschool

Nijmegen, 2014-06-23

# 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: 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

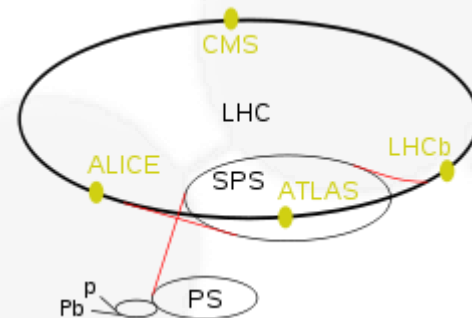
- **A research infrastructure for humanities researchers who work with digital language-related resources**

# 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



- 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**

- 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 NO PL (SV) and growing

# CLARIN IN EUROPE

- CLARIN ERIC: [www.clarin.eu](http://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)
  - ...

# 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

*'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, 2<sup>nd</sup> hour
- Illustration 2 (Search, Syntax): Tue 24<sup>th</sup> (GrETEL)
- Illustration 3 (Enrichment): Wed 25<sup>th</sup> (TTNWW)
- Illustration 4 (Search, Lexical Semantics): Thur 26<sup>th</sup> (Cornetto)

# 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

*‘CAN STORE THE DATA & TOOLS’*

- 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 27<sup>th</sup>.

# 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](mailto:helpdesk@clarin.nl)
  - Service-specific user group lists


# COURSE PROGRAMME

Day	Who	What
Mon 23rd	Jan Odijk	Introduction
		Illustrative Usage Case
Tue 24th	Liesbeth Augustinus	Treebank Mining By Example
	Frank Van Eynde	Number agreement in copular constructions
Wed 25th	Marc Kemps-Snijders	Enriching your data with grammatical information using TTNWW
Thu 26th	Piek Vossen & Isa Maks	Cornetto
Fri 26th	Jan Odijk	Storing data in CLARIN
		Concluding Overview

# COURSE PROGRAMME

- For more details, see
  - <http://www.clarin.nl/node/2016>
  - [http://www.lotschool.nl/link.php?url=http://www.lotschool.nl/files/schools/2014 Summerschool Nijmegen/](http://www.lotschool.nl/link.php?url=http://www.lotschool.nl/files/schools/2014_Summerschool_Nijmegen/)

Thanks for your Attention  
And  
Enjoy the Course!

A large, faint background graphic consisting of a network of light gray circular nodes connected by thin, curved lines, resembling a molecular or network structure. The nodes are arranged in a roughly circular pattern, with some nodes being larger than others.