

# **CLARIN-NL: CONTRIBUTIONS BY NLP & CURRENT STATUS**

Jan Odijk

CLIN-dag

Antwerp, 2015-02-06

# OVERVIEW

- CLARIN
- CLARIN-NL and NLP
- CLARIN-NL Portal
- CLARIN-NL: Current Status and Near Future

# CLARIN INFRASTRUCTURE

- **A research infrastructure for humanities researchers who work with digital language resources**
  - **Research infrastructure:** facilities, resources and related services used by the scientific community to conduct top-level research
  - **Humanities researchers:** linguists, historians, literary scholars, ...
  - **Language resources:** lexicons, corpora, databases, ...; text, audio, video, ....

# CLARIN-NL AND NLP

- NLP data and software developed by you now available to humanities researchers in general thanks to CLARIN-NL
  - By specifically designed search engines and interfaces
  - By user friendly interfaces to software
  - Mostly as web applications (and web services)
  - Most by NL, some by FL, some together

- Easy Access to data
  - OpenSONAR (see **Reynaert's presentation**)
  - GrETEL (Flanders) (see **Augustinus et al.'s poster** )
  - Cornetto (lexico-semantic database)
  - DuELME (multiword expressions database)
  - PaQu (see **Odijk's poster** )
  - Autosearch (as of March 2015)

- Easy use of software
  - INPOLDER/Adelheid (syntactic analysis of 13<sup>th</sup> century Dutch )
  - TTNWW (orthographic normalisation, pos-tagging, parsing, NER, co-reference assignment, speech transcription, ....) (NL+FL)
  - TiCCLOps (orthographic normalisation)
  - PICCL (see **Reynaert's poster**)
  - Stylene (Flanders) – stylistic analysis

# CLARIN-NL AND NLP

- Facilitated by
  - CLAM wrapper to easily build web services
  - Increased *syntactic interoperability*: FoLiA increasingly a *de facto* standard format in NL for linguistically annotated corpora
  - Increased *semantic interoperability* through ISOCAT and currently the CLARIN Concept Registry (poster by **Schuurman et al.**)

- CLARIN-NL Portal: <http://portal.clarin.nl>
  - Faceted search in data and services created by CLARIN-NL
    - Search by [research domain](#), [language](#), [tool task](#), linguistic annotation and several other facets
  - Links to search facilities in the whole CLARIN infrastructure
  - Educational packages: <https://dev.clarin.nl/node/CLARIN%20Educational%20Packages>
  - And much more....



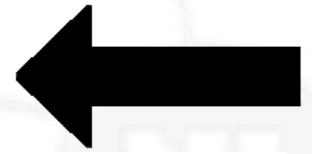
# CURRENT STATUS

- CLARIN-NL finishes by April 1<sup>st</sup>, 2015
- Final event:
  - Friday March 13, 2015, (morning)
  - Netherlands Institute for Sound & Vision, Hilversum
  - Register via the CLARIN-NL website:  
<http://www.clarin.nl/node/2074>

# NEAR FUTURE

- CLARIAH-CORE successor project
- Jan 1, 2015 – Dec 31, 2018
- 12 m euro
- Kick-off:
  - Friday March 13, 2015, (afternoon)
  - Netherlands Institute for Sound & Vision, Hilversum
  - Register via the CLARIAH-website  
<http://www.clariah.nl/activiteiten/clariah-kick-off>

**Thanks for Attention!**



## CLARIN-NL – Linguistics subdisciplines

|                        |                                    |                           |
|------------------------|------------------------------------|---------------------------|
| Historical linguistics | Dialectology                       | Discourse Studies         |
| Language Acquisition   | Language Documentation             | Lexicology / Lexicography |
| Morphology             | Morpho-syntax                      | Phonetics                 |
| Phonology              | Pidgin & Creole Studies            | Semantics                 |
| Sign Language          | Specific Language Impairment (SLI) | Syntax                    |
| Typology               | ...                                |                           |

# LANGUAGES COVERED

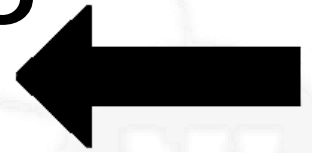


## CLARIN-NL – Languages

|         |   |                 |
|---------|---|-----------------|
| Dutch   | English   | German          |
| Frisian | Dutch Sign Language   | Classical Greek |
| French  | Hebrew  | Aramaic         |
| Syriac  | > 50 languages from Insular South East Asia and West New Guinea | And more...     |

- And many more (> 2,000) from all CLARIN countries

# WHAT YOU CAN DO



## CLARIN-NL – Functionality Offered

|                                 |                           |                            |
|---------------------------------|---------------------------|----------------------------|
| Searching                       | Browsing                  | Analysis                   |
| Corpus Exploration              | Annotation                | Tokenization               |
| Pos-tagging                     | Lemmatization             | Orthographic normalisation |
| Grammatical relation assignment | Named entity recognition  | Chunking                   |
| Co-reference assignment         | Multiword unit assignment | Parsing                    |
| Visualisation                   | Diarisation               | Speech recognition         |