

CLARIN – NL & CLARIN FLANDERS COOPERATION PROJECT

JAN ODIJK

CLARIN-NL MEETING 19 FEB 2010

NL-FL Cooperation Project



- Mar 2010- Sep 2012
- Funded by
 - EWI (Flanders-FL)
 - CLARIN-NL (Netherlands-NL)
- Budget:
 - Total: 2,376k Euro
 - 1,584k Euro NL (2/3)
 - 792k Euro FL (1/3)

NL-FL Cooperation Project



- Part I
 - 1,356k Euro, NL: 904k Euro, FL: 452k Euro
 - TST Tools voor het Nederlands als Webservices in een Workflow (TTNWW) project (NL+FL)
 - Infrastructure Implementation Plan (NL)
- Part II
 - 920k Euro; NL: 680k Euro; FL: 340k Euro
 - Call I CLARIN-NL (NL)
 - Projects on search in treebanks, stylometric tools and opinion / emotion mining in speech (FL)

NL-FL Cooperation Project



- TTNWW
 - Convert existing tools developed in STEVIN into webservices embedded in a workflow system
 - To run on servers van recognized CLARIN-centres
 - Aim: offer humanities researchers with no or little technical background facilities to
 - Address their research questions in a better and/or easier manner
 - Offer them opportunities for formulating completely new research questions
 - both for textual and for speech resources

NL-FL Cooperation Project



- TTNWW
 - Users (text):
 - Literary studies (Huygens Institute Den Haag)
 - Archaeologists (Sargasso KU Leuven)
 - Users (speech)
 - Social History / Political Science
 - KDC Nijmegen <http://www.ru.nl/kdc/>
 - Aletta Amsterdam <http://www.aletta.nu>
 - KADOC Leuven <http://kadoc.kuleuven.be/nl/index.php>
 - M2P Antwerp
<http://www.ua.ac.be/main.aspx?c=.ONDERZKDBE&n=40312&id=UA033&tid=21157>

NL-FL Cooperation Project



- TTNWW
 - Technology Providers (text)
 - Universiteit Utrecht, Universiteit van Tilburg, Universiteit van Groningen, Radboud Universiteit Nijmegen
 - KU Leuven, Hogeschool Gent, Universiteit van Antwerpen
 - Technology Providers (speech)
 - Universiteit van Twente, Radboud Universiteit Nijmegen
 - KU Leuven, Universiteit Gent
 - Infrastructure Expertise & Centres
 - MPI, INL
 - Coordinators: Marc Kemps-Snijders & Ineke Schuurman

NL-FL Cooperation Project



- Web services (text)
 - Text preprocessing
 - Lexical Unit Tagging
 - Shallow Parsing
 - Syntactic Parsing
 - Alignment
 - Named Entity Recognition
 - Coreference resolution
 - Semantic Role Assignment
 - Spatiotemporal Analysis

NL-FL Cooperation Project



- Web services (Speech)
 - Segmentation
 - Speech segments in Dutch
 - Segments in a different language
 - Non-speech segments
 - Transcription
 - Alignment
 - Speaker assignment and characterization
 - Indexing and Search

NL-FL Cooperation Project



- Why this?
 - Webservices in workflows will form a core functionality in CLARIN and NL has to play a role in developing this for CLARIN
 - The STEVIN tools offer an excellent test set for this
 - The STEVIN tools can offer excellent functionality for humanities researchers
- Why with Flanders?
 - The tools and data are for the (common) Dutch language and have been developed together with Flanders in the STEVIN programme

NL-FL Cooperation Project



Thanks for your attention!