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 webservice 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
NL-FL Cooperation Project

• TTNWW
  o Technology Providers (text)
    o Universiteit Utrecht, Universiteit van Tilburg, Universiteit van Groningen, Radboud Universiteit Nijmegen
    o KU Leuven, Hogeschool Gent, Universiteit van Antwerpen
  o Technology Providers (speech)
    o Universiteit van Twente, Radboud Universiteit Nijmegen
    o KU Leuven, Universiteit Gent
  o Infrastructure Expertise & Centres
    o MPI, INL
  o 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
• 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
• 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
Thanks for your attention!