



CLARIN-Flanders

New prospects

Vincent Vandeghinste
Els Lefever
Walter Daelemans
Tim Van de Cruys
Sally Chambers



A new member to CLARIN.eu: CLARIN-BE

Given that:

- CLARIN members need to be countries or international organisations
- Flanders is not a country

Ergo:

- Flanders can not become a member of CLARIN.eu

Solution:

- Apply for political support from Flanders (no funding) for the formal founding of CLARIN-BE and the membership fee
- Support was granted, membership is now a fact (since September 2021)



CLARIN-VL Consortium

OPEN FOR NEW RESEARCH GROUPS

/instituut voor
de Nederlandse
taal/



- CLARIN-B centre for CLARIN-BE and CLARIN-VL
- CLARIN-K Centre K-Dutch for the Dutch language
- Vincent Vandeghinste, national coordinator CLARIN-BE
- Language and Translation Technology Team (LT3), Els Lefever, CLARIN-VL principal investigator
- Internet Technology and Data Science (imec-IDLab)



- Centre for Computational Linguistics and Psycholinguistics (CLIPS), Walter Daelemans



KU LEUVEN

- Centre for Computational Linguistics (CCL), Tim Van de Cruys
- Language Intelligence and Information Retrieval Lab (LIIR), Sien Moens
- ESAT Processing Speech and Images (ESAT-PSI), Patrick Wambacq



CLARIAH-VL: Plans

- **Digital Text Analysis Dashboard and Pipeline for**
 - Dutch texts
 - Parallel texts
 - re-using existing CLARIN tools
 - extended with variation of models and NLU analysis layers
- **Benchmarking existing tools and new models**
 - Linguistic processing tools (pos, lemma, ner, ...): LeTs Preprocess, Frog, DeepFrog, SpaCy, Stanza
 - Natural Language Understanding: development of a suite of NLU evaluation tasks (cf. GLUE) for Dutch
 - Explore application of neural network models for search and extraction of linguistics structures
- **Integrate NLU tools in pipeline**
 - Sentiment analysis, Emotion detection, Document similarity clustering, Topic modelling, Stylometry
- **Integrate Tools for multilingual or parallel data**
 - Sentence alignment, Word alignment
- **Setup of Use Cases**
- **Setup of a Help Desk**