EUROPEAN
LANGUAGE
EQUALITY

META-NET
META-FORUM 2022

EUROPEAN
LANGUAGE
GRID

The Dutch Language in the Digital Age

Frieda Steurs (Instituut voor de Nederlandse Taal, Leiden)
Frieda.Steurs@ivdnt.org

The Dutch Language in the Digital Age

- Dutch: 25 million speakers in the Low Countries.
  - Native language in other parts of the world: Suriname, Caribbean
  - Minorities in France, Germany, Indonesia
- Language skills: Key qualification in education (a very multilingual society)
- Language technology: CALL, educational games, subtitling, search & text categorization
- The Low Countries: rich, strong economy, strong activity on the internet
- Websites, internet use, e-commerce is extremely well developed.
- Twitter, social platforms, new linguistic trends, language variation, subcultures, language of the youth, ….

Language Technology for Dutch

- European Infrastructures include Dutch as a major language
  - ELG, Clarin and ELRC
  - ELDA and LDC (paid membership services)
  - Clarin Knowledge Centre (K-Dutch) at the Dutch Language Institute
- Large number of corpora (SoNaR, DPC, Subcorpora from OPUS, ELRC, Open Dutch Wordnet)
  - INT develops a computational lexicon of the Dutch Language from the 6th century up to now
- Lots of tools are available for linguistic text analysis, parsing, text-to-speech processing (both for Belgian Dutch and Netherlandic Dutch)
- Dutch is present in most commercial online translation services (Google, Bing, DeepL) and also in eTranslation.

Projects & Initiatives

- Strong NLP community in NL and Flanders, but **no overarching research programme**
- Dutch AI Coalition (Dutch AI for Dutch), Dutch Speech Coalition, NoTAS (Dutch organi-zation for Language and Speech Technology), Flanders AI programme (research track NLP)
- Clarin.nl and Clarin.be – Clariah NL and Clariah VL
- Dutch Language Institute: Clarin technical centre
  - Catalogue: Taalmaterialen (INT) https://ivdnt.org/taalmaterialen/
- Language Services LaMachine from Radboud University Nijmegen

Conclusions

- Dutch is not in bad shape digitally, however, public corpora become quickly outdated (often made before 2011).

- The former Stevin programme ran from 2004 till 2010. The aim was to realize an appropriate digital language infrastructure for Dutch and to stimulate strategic research in the domains of language and speech technology. However, since 2010, there is no further possibility for cooperation between the two research communities.

- A great number of important resources and tools is lacking or needs updating.

- Language use in recently emerged domains needs to be collected, such as social media, and the construction of corpora that document recent language in written, spoken form and material from blogs.

- The future is uncertain.

- Urgent need for a new research programme for cooperation between the Netherlands and Flanders.

European Language Grid

European Language Equality

Thank you!

Frieda Steurs (Instituut voor de Nederlandse Taal, Leiden)
Frieda.Steurs@ivdnt.org

08/09-06-2022 META-FORUM 2022 – Joining the European Language Grid (hybrid conference)
http://www.european-language-grid.eu – http://european-language-equality.eu

EUROPEAN
LANGUAGE
GRID

EUROPEAN
LANGUAGE
EQUALITY