

META FORUM 2016

HLT Language Policy for Dutch

Jan Odijk

Utrecht University, the Netherlands
j.odijk@uu.nl

META-FORUM 2016
Lisbon, Portugal – July 04/05, 2016



META-NET has received funding from the EU's Horizon 2020 research and innovation programme through the contract CRACKER (grant agreement no.: 645357). Formerly co-funded by FP7 and ICT PSP through the contracts T4ME (grant agreement no.: 249119), CESAR (grant agreement no.: 271022), METANET4U (grant agreement no.: 270893) and META-NORD (grant agreement no.: 270899).

Outline

- ❑ Dutch in the White Papers
- ❑ Causes
- ❑ And Now?
- ❑ New prospects
- ❑ Conclusions

Dutch in the White Papers

- ❑ Digital support for European languages
 - Data and basic software
- ❑ 21/30 languages: 'weak' or 'non-existent'
 - Big risk of digital extinction
- ❑ Dutch: 'limited support'
 - In the same class as Spanish, German, ..
- ❑ Least bad: (American) English

How come?

- ❑ EU pre-competitive data collection projects (1995-2003)
 - Esp. SpeechDat family of projects
 - The Dutch language was present in almost all projects
- ❑ Important HLT companies (at the time) had Netherlands (NL) or Flanders (FL) origin (Philips, Lernout & Hauspie) (1996- 2003)
- ❑ Joint NL - FL project STEVIN: resource creation and research (2004 – 2010)
 - intentional language policy aimed at ensuring the continued existence of the Dutch language in the information and communication society
 - Cooperation between NL and FL
 - Cooperation between academia and industry
 - Co-ordinated by the Dutch Language Union
 - 11.4 million euro budget

- ❑ Yielded basic data and software for R&D of the Dutch language
 - Monolingual
 - (richly annotated) text and speech corpora, lexical resources
 - Improved and new systems for text-to-speech, (noisy) speech recognition, pos tagging, parsing, NER, coreference resolution, sentiment mining, summarisation etc.
 - Multilingual
 - Parallel corpora, research pilot MT-systems
 - All data and many tools available through the HLT Agency:
 - <http://tst-centrale.org>
- ❑ Produced 17 real world applications (mainly by industry)

- ❑ Strengthened HLT research in NL+FL
- ❑ Strengthened cooperation (continued after the project)
 - NL and FL
 - academia – industry
- ❑ Many more details on policy and scientific results in the Open Access book:
 - Spyns & Odiijk (eds.). 2013. *Essential Speech and Language Technology for Dutch*, Springer. DOI 10.1007/978-3-642-30910-6, <http://link.springer.com/book/10.1007%2F978-3-642-30910-6>

Current Situation?

- ❑ EU pre-competitive data collection projects (1995-2003)
 - SpeechDat family of projects
 - The Dutch language was present in almost all projects
- ❑ Important HLT companies (at the time) had Netherlands or Flanders origin (Philips, Verbeet & Hauspie) (1996- 2003)
- ❑ Joint Netherlands - Flemish project STEVIN: resource creation and research (2004 – 2009)

HLT flourishes in NL+FL

- ❑ Yearly NL+FL conference: [CLIN](#)
 - > 100 participants, this year 27th edition
- ❑ Piek Vossen (VU, A'dam) [Spinoza laureate 2013](#)
- ❑ Good scores with project proposals
 - Speech recognition, language generation, (semi-) automatic translation, semantics, analysis of social media

HLT flourishes in NL+FL

- ❑ Active and organised industry
 - NOTaS
- ❑ Companies with good products / services, some also for Dutch
 - Speech recognition and synthesis, search engines, authoring systems, classification systems, ...
- ❑ Multiple Spin-offs
 - *inter alia* Antwerpen, Nijmegen

- ❑ HLT for Digital Humanities
 - CLARIN(-NL), CLARIAH
- ❑ Search through audio-visual material
 - Parliament, NISV
- ❑ Inclusive society
 - HLT used in Care, for low literacy persons, subtitling (VRT)
- ❑ Analysis of social media texts

Recent Developments

- ❑ **Oct 10, 2015:** Language Congress organized by the Joint Parliamentary Committee on the DLU
- ❑ I argued NL+FL should start up a new HLT programme [[pptx](#)]
 - Very well received, positive reactions
- ❑ **Dec 2015:** a [request](#) from the chair of the ministerial committee for the Dutch Language Union (Dutch minister Jet Bussemakers, ministry of Education, Culture and Science)

to inventory what a new language and speech technology programme would have to contain and what it would cost, however without making any financial commitments

- ❑ Currently working on that...

- ❑ Active language policy for the digital support of languages is possible (if the political will is present)
- ❑ At the national and at the European level
- ❑ Active policy can have real positive effects on
 - HLT R&D
 - cooperation between academia and industry
 - the availability of HLT-enabled applications
 - Leading to reduction of costs for many services

Q/A

META  NET

Thanks for your attention.

office@meta-net.eu

<http://www.meta-net.eu>

<http://www.facebook.com/META.Alliance>