



A Common Danish Language Resource

November 2021

Speaker

Christian Plaschke

**Head of office, Center for Technology and Data part of the
Danish Agency for Digitisation**

New Technologies in the public sector

Data in the public sector

Lead on the public initiative "sprogteknologi.dk"



The initiative sprogteknologi.dk

2019



2019



- Joint government cooperation (Central, regional and local)
- **Goal – enhance the development of AI-solutions that works in the context of Danish**
- Enable SMV's and smaller public organisations to work with Danish LT.
- €2,6 million allocated between 2019 – 2026
- Working with private and public sector organizations

Current and future work

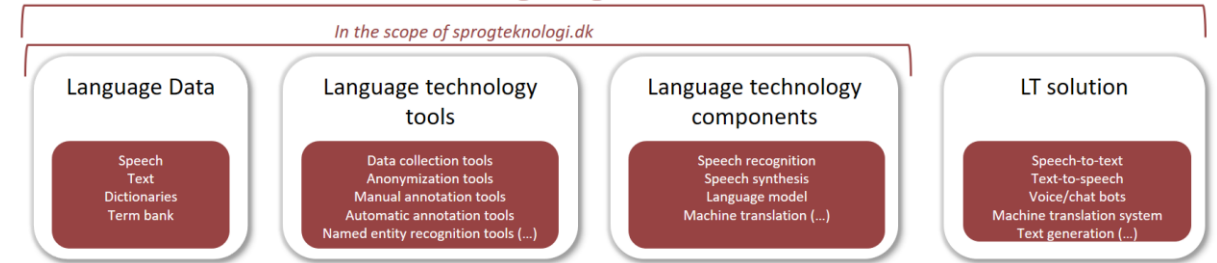


Gather

- There is already a good amount of language resource available.
- Gather existing Danish language resource and display them on our portal.
- Make it easier for small organizations to identify, evaluate and use available resources.

Common Danish Language Resource

Language Resources



- www.sprogteknologi.dk
- 127 ressource atm.
- CKAN
- DCAT-AP-DK
- We display metadata on and link to existing ressourceces
- We write news on latest ressource and activities for the community.

Current and future work



Gather



Create

- Central word register, lagrer speech corpus, make interesting language ressource public available etc.



Educate

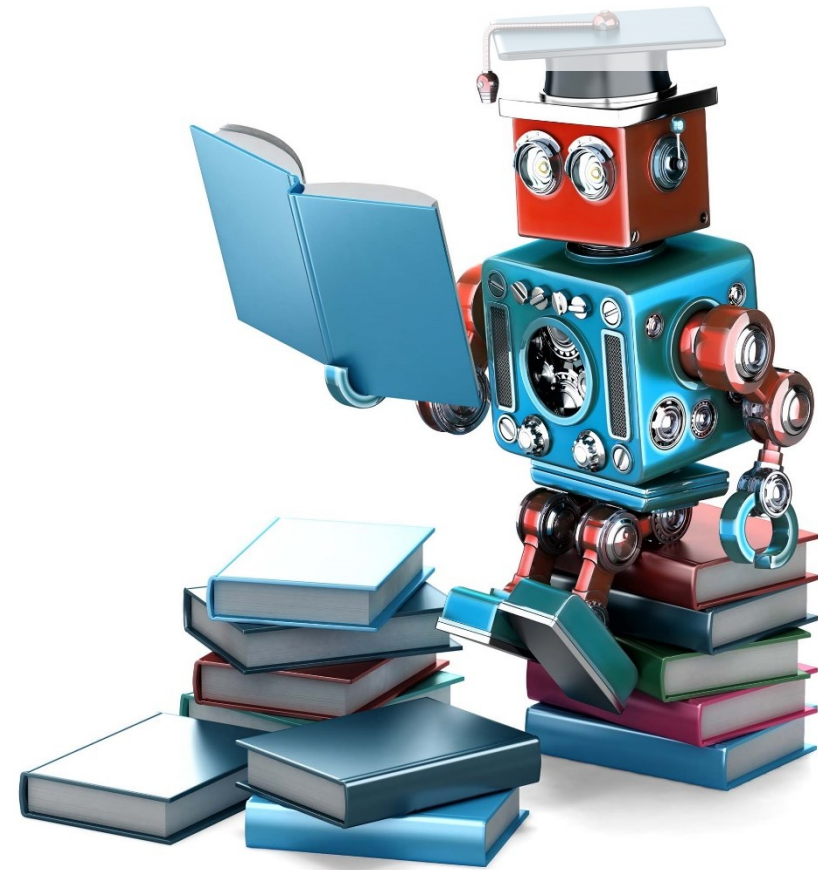
- Inform relevant actors on EU projects
- Gathering experience from LT projects
- Arrange events where different actors can meet and network
- Put emphasis on the importance of LT in other policy initiatives, if related.

Lessons learned so far

Growing demand for LT and we are able to pitch LT in other projects and initiatives

GDPR and Copyright-rules makes developing language-resources difficult

Data is key – we try to limit our work from going into building actual end user solutions – as they will go old fast





DIGITALISERINGSSTYRELSEN

Visit www.sprogteknologi.dk
and follow us on [LinkedIn](#)

Christian Plaschke

chpl@digst.dk or info@sprogteknologi.dk

A Central Word Register for Danish AI

- COR

Bolette Sandford Pedersen, University of Copenhagen



DET DANSKE SPROG- OG
LITTERATURSELSKAB

Dansk
Sprognævn



DIGITALISERINGSSTYRELSEN

Background and needs

- Companies in Denmark are right now entering the field of language-centered AI – and are therefore working intensively with Danish language data from an NLP perspective
- An increasing request for a standard machine readable lexicon of Danish with basic morphology and semantics (core senses, sentiment etc.)
- COR is funded by the ***Danish Agency for Digitisation*** under the AI initiative

Why not use what already exists?

- Time-consuming for the companies to get an overview of existing resources and their availability
- Coverage and validity is often unclear
- Lack of in house language experts in the companies who can estimate quality and relevance
- Coordination with existing terminology in the company is often complex

Aims (phase 1)

- COR aims at meeting these requests in three ways:
 - A coordinated and standardized framework for machine readable lexical resources for Danish where all lemmas (including terminology) are assigned ***a unique identifier***
 - A lexicon of the general language vocabulary with basic morphology and semantics (syntax and phonology is foreseen in phase 2)
 - Ability to merge domain-specific vocabularies

Who

- COR is developed by three core language and language technology institutions in Denmark and is based on *high-quality, locally-anchored knowledge about the Danish language and society* as depicted in national dictionaries:
- The Danish Language Council, (Retskrivningsordbogen)
- The Danish Society for Language and Literature, (Den Danske Ordbog)
- The Centre for Language Technology, University of Copenhagen (PAROLE etc.)

The semantic component of COR

- Several computational lexicons already exist for Danish (WordNet, FrameNet, Sentiment Lexicon) but they need to be integrated and somewhat simplified
- Task 1: Merge of info types from the different sources (e.g. ontological types, synonymy, polysemy, semantic relations, semantic frames and roles, positive-negative polarity)
- Task 2: Simplification of the sense inventory from Den Danske Ordbog by clustering and deletions of senses
- Task 2: Extension of coverage