

Country Profile  
Denmark

Sabine Kirchmeier  
sabine@dsn.dk

Dansk  
Sprognævn

---

# Overview

- Language Technology in Denmark – public and private actors
- Organisation of translation activities
- Political initiatives
- The work of the language technology committee

# Denmark an overview

- 5.8 million people
- One official language: Danish
  - German in the southern border region
  - Swedish around Copenhagen
- English dominates as second language
- Arabic, Turkish, Polish, Romanian as main immigrant languages
- Denmark is a highly digitized society

# Main public actors

- Danish Agency of Digitization (Ministry of Finance)
- Center for Language Technology/University of Copenhagen (Ministry of Education and Science)
- Danish Language Council (Ministry of Culture)

# Secondary public actors

- Danish Municipalities
- Other ministries and agencies
- Royal Library
- Technological service institutes and universities
- Danish State Auditors

# Private actors

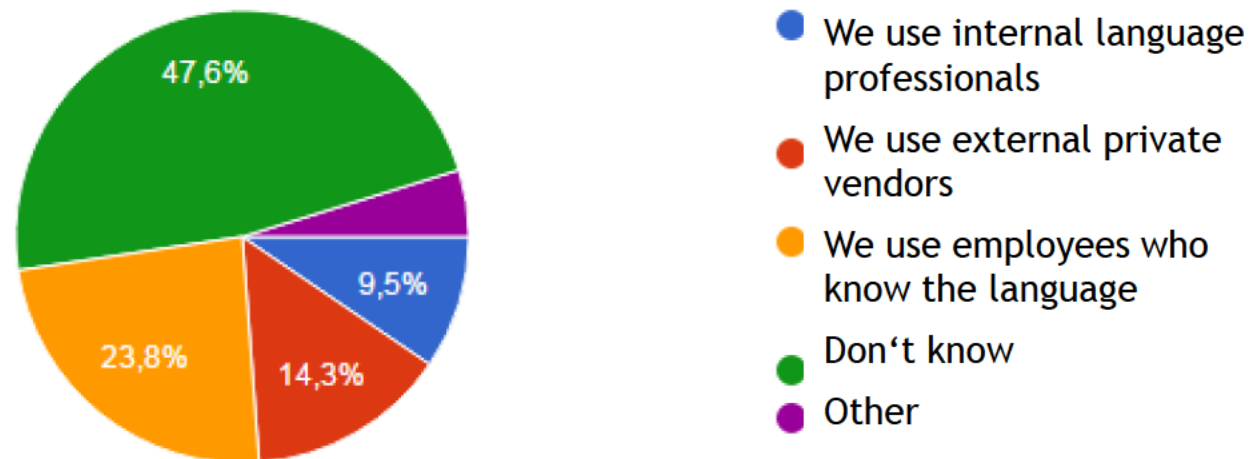
- Microsoft, IBM, Google. etc.
- Speech technology (Nuance, MaxManus, Dictus, Mirsk, etc.)
- Translation agencies (Semantix, Lionbridge, etc.)
- Language technology (Ankiro, Gramarssoft, etc. )
- Consultant agencies

# Translation services

- Few public in house translation services (Region South Jutland-Schleswig, Danish Tax Authorities, Nordic Council of Ministers)
- 80-90 % outsourced to private vendors
- Very little awareness in public institutions about the importance of resources

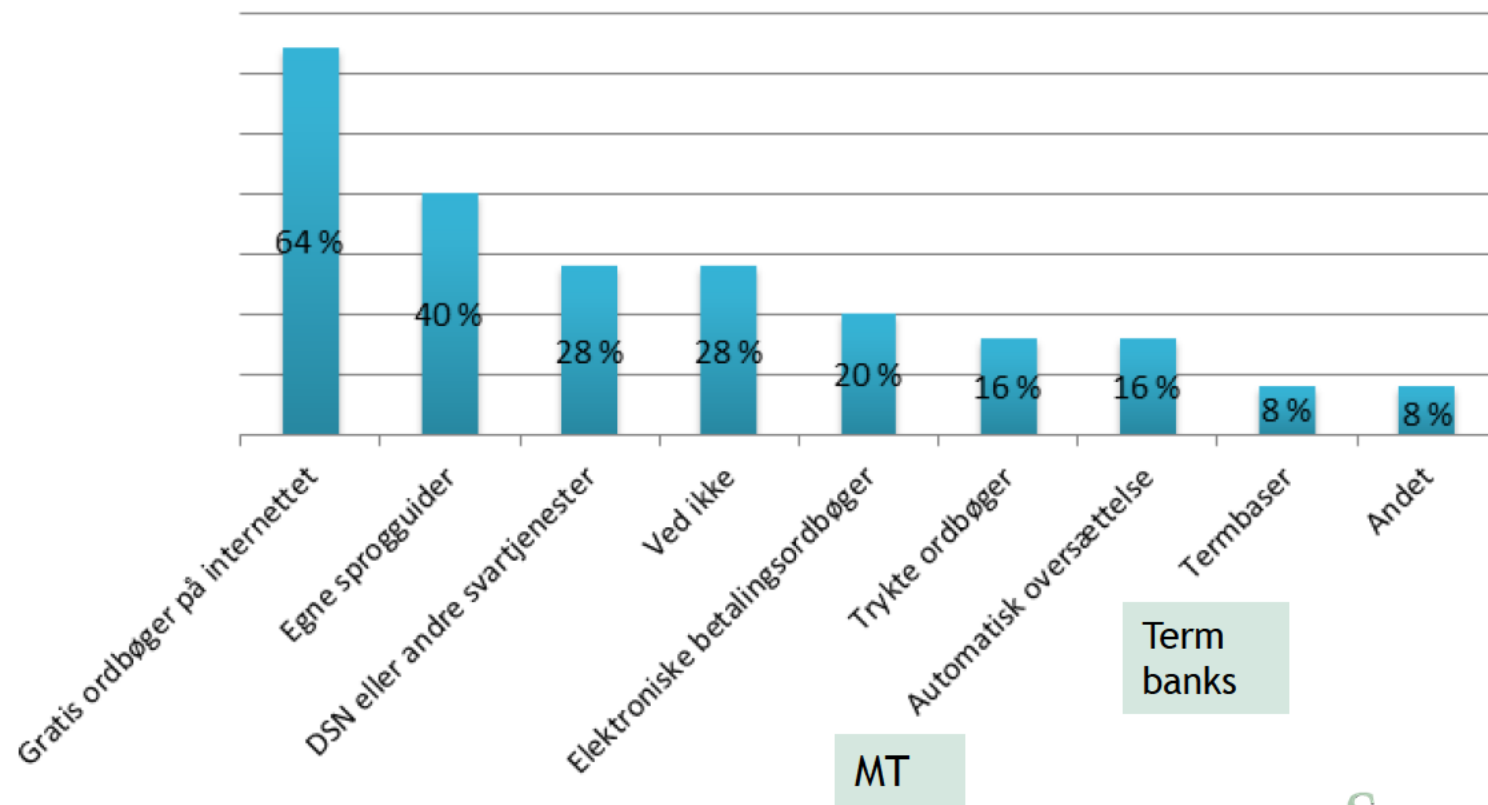
# Translation services in Danish Municipalities

If you use other languages than Danish, how do you deal with translation tasks?



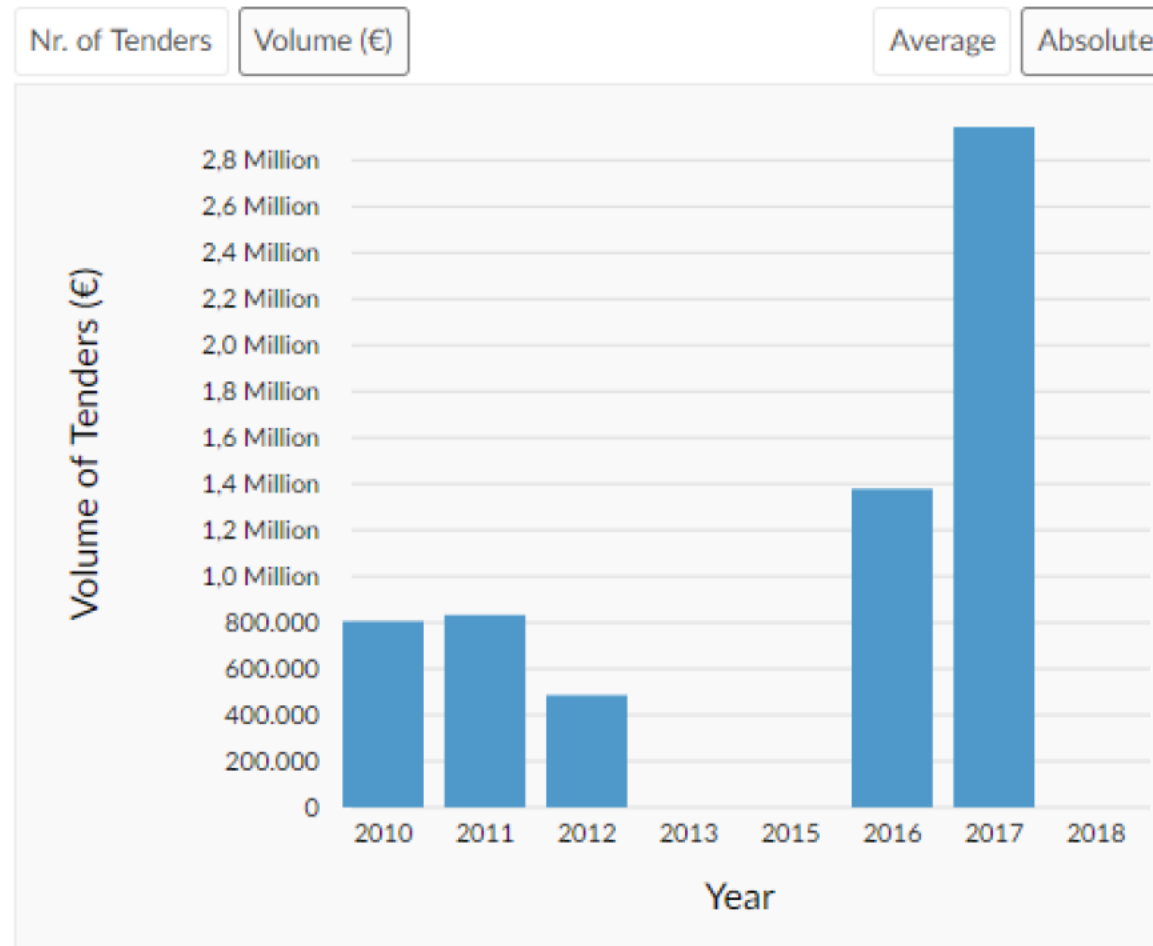


# Use of language tools in Danish Municipalities



# Translation volume

Sector over Time

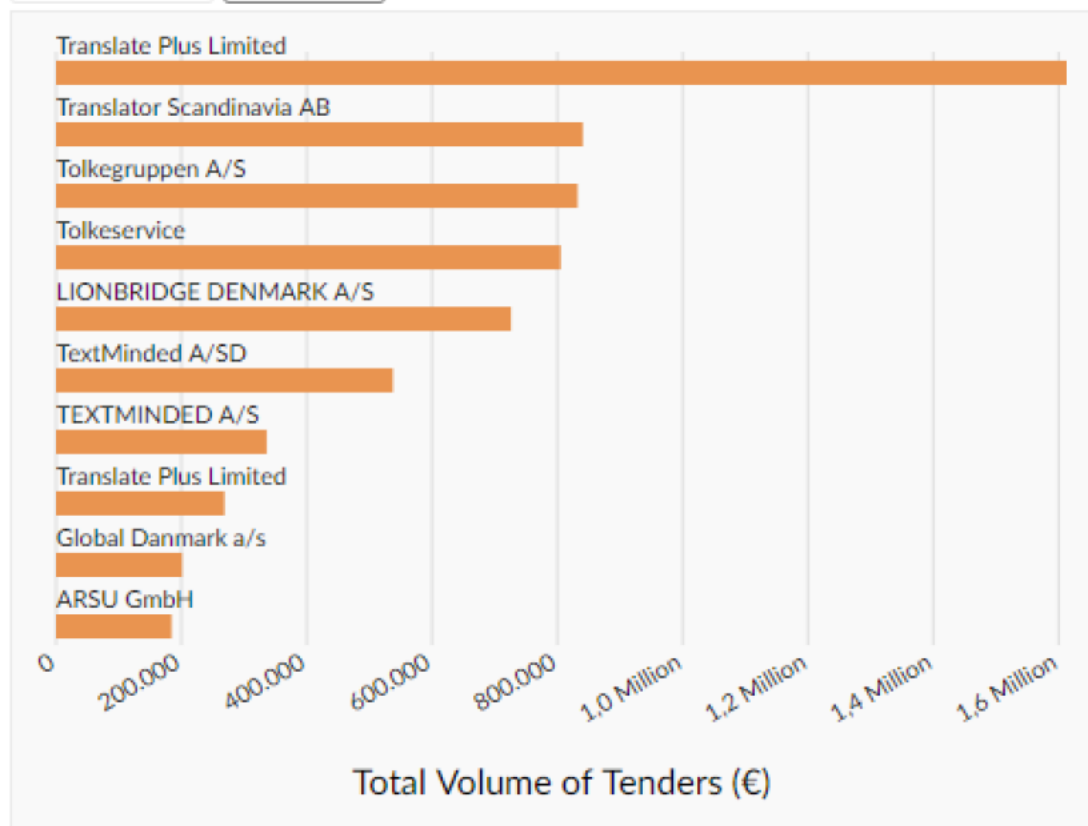


# Biggest translation providers

## Main Suppliers

Nr. of Tenders

Volume (€)

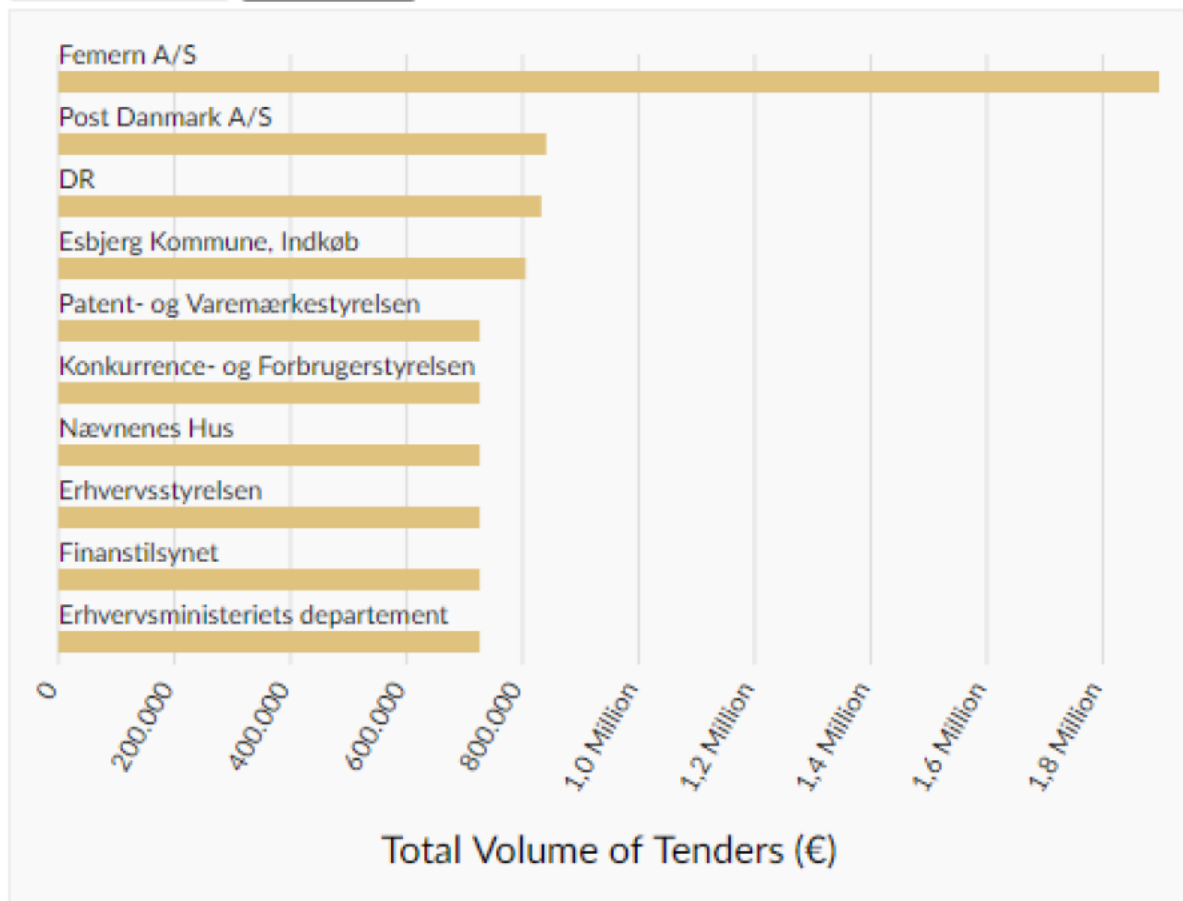


# Biggest costumers

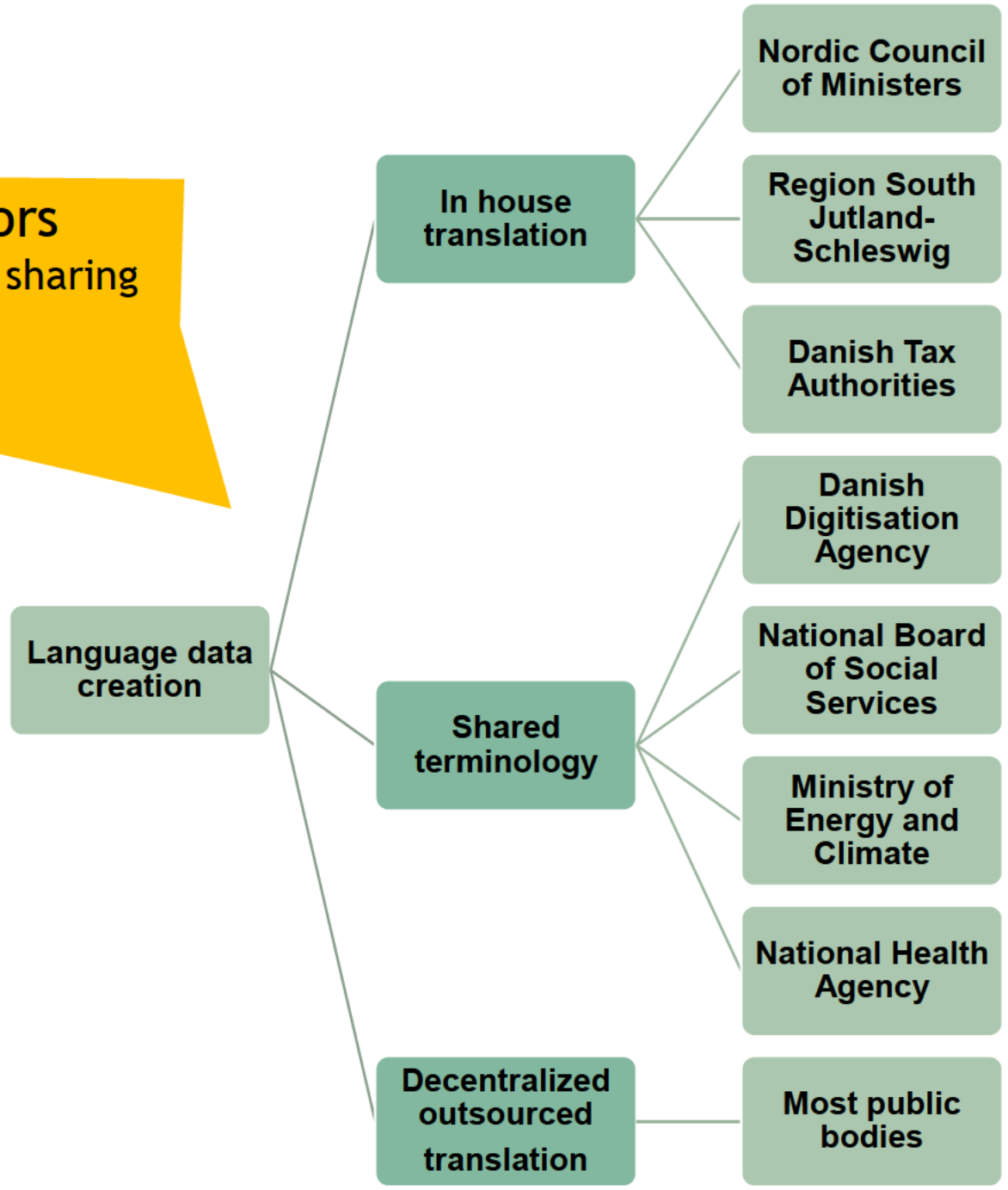
## Main Buyers

Nr. of Tenders

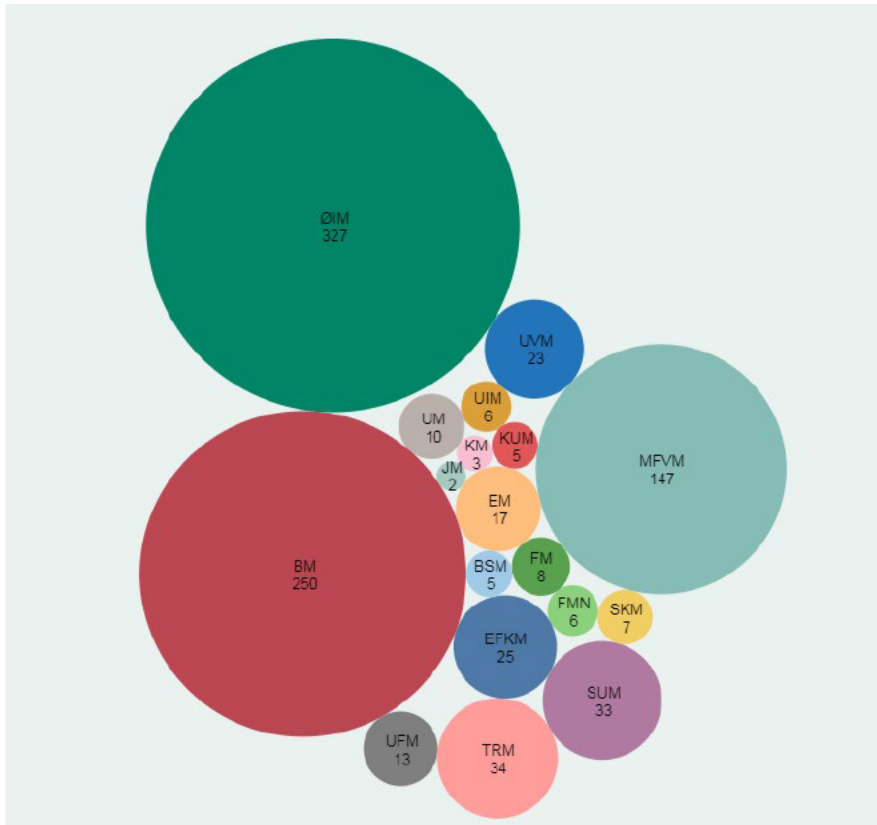
Volume (€)



**State Auditors**  
Control of data sharing  
procedures

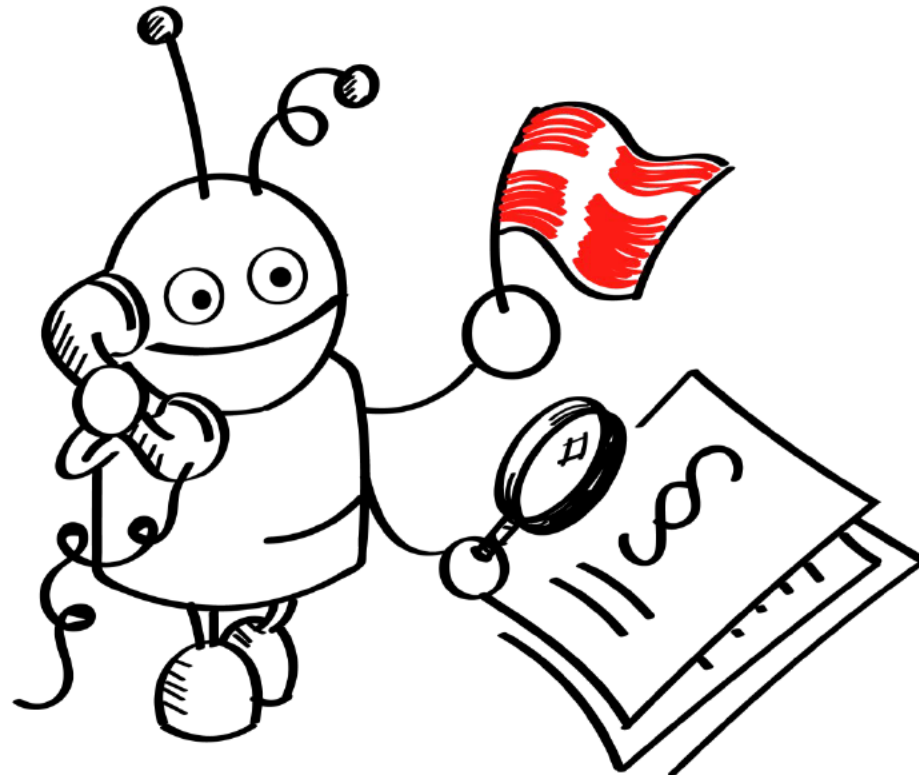


# State auditors' report on PSI (March 2019)



- Few large data providers
- 1000 data sets spread in 82 places
- Mainly numbers, very little text

# Recommendations by the language technology committee (sprogtek2018.dk)



# Problems for language technology for Danish

- Size of the language community. Small market for language technology for Danish
- Special features of the Danish language (complex phonology, compounding, movement phenomena)
- Lack of basic resources: Quality data for speech technology, annotated text, terminology and semantics (WordNet, Frame Net etc.)
- Lack of coordination
- Lack of strategic research and low priority of further education and training in computational linguistics and LT.



# Not sufficiently many high quality resources for Danish

## Oversigt over danske sprogresurser

Indhold/type	Parallel tekst	Tale	Tekst	Tekst og tale	Terminologi	Ord-net	I alt
Formelle grammatikker			1				1
Sentimentopmærkning			1				1
Stavning/fejlopmærkning			2				2
Lister over stednavne og geodata			2				2
Træbanker			4				4
Semantisk ordbog eller opmærket tekst			2			3	5
Tekstkorpusser (Ældre litteratur)			6				6
Ordbøger	3	2	10		3		18
Værktøj til danske sprogdata	4	4	20		1		29
Tekstsamlinger og korpusser	11	8	11	1	1		32
<b>I alt</b>	<b>18</b>	<b>14</b>	<b>59</b>	<b>1</b>	<b>5</b>	<b>3</b>	<b>100</b>

# Political initiatives

Dansk  
Sprognævn

# Political initiatives

- Few initiatives from 2000-2015
- MetaNet report 2012
- 2015-2016: Series on Language Technology in the communications of the Danish Language Council
- 2016: Proposal in parliament to establish LT committee
- 2017: Hearing in parliament
- 2018: Establishment of LT committee under the Danish Language Council by the minister of culture

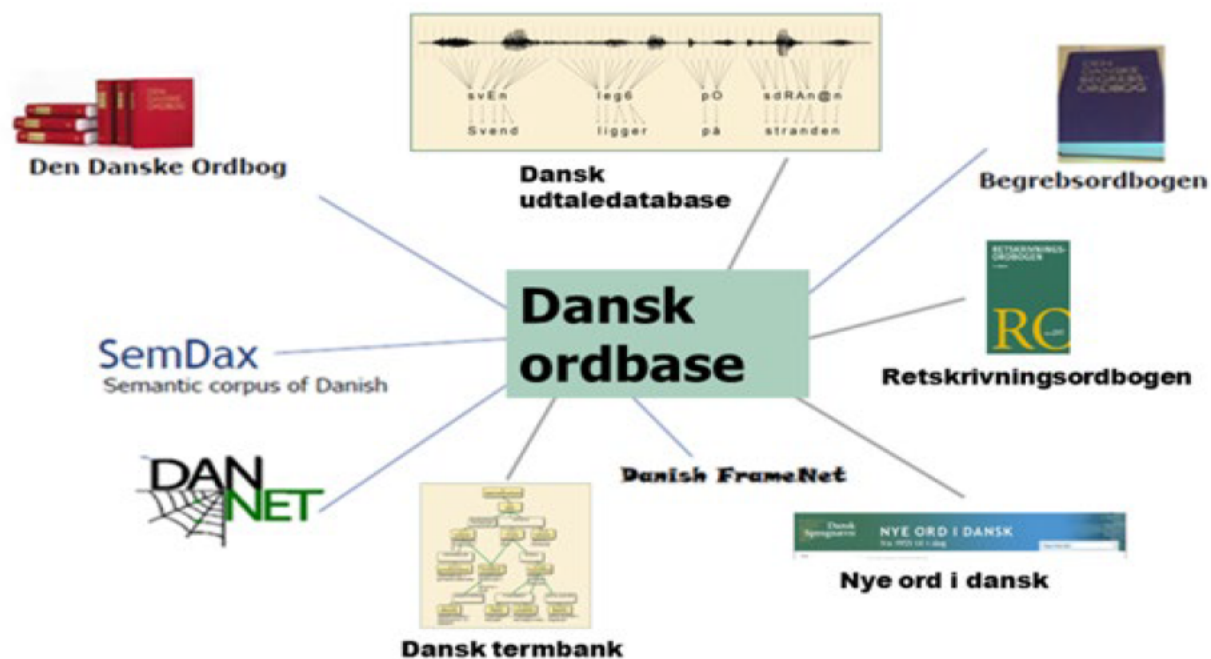
# Political initiatives

- October 2018: Ministry of Finance – World class digital service
  - „World class language technology“
- March 2019: Ministry of Finance – National AI-strategy
  - 4 mio. EUR for language resources
- March 2019: First PSI-report by State Auditors

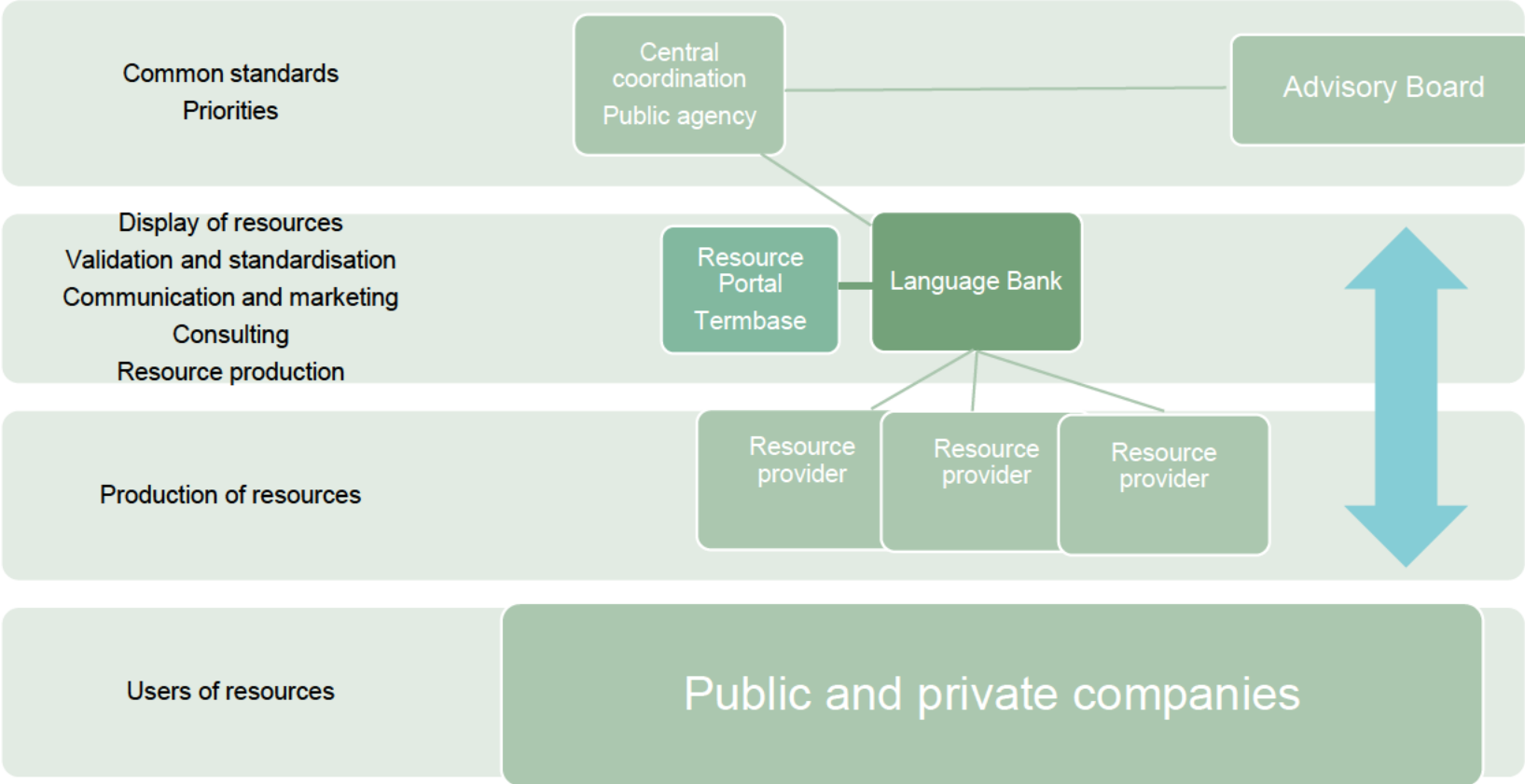
# Recommendations by the LT committee

1. Establishment of a central organisation to plan and manage a National Danish language resource bank
2. The language bank should contain high quality resources of the following kind:
  - 2.1. A time annotated Danish speech technology corpus
  - 2.2. Danish text corpora and annotated gold standards for machine learning
  - 2.3. A comprehensive lexical database
  - 2.4. A Danish terminology bank
  - 2.5. A language technology tool-kit
  - 2.6. A language portal for the distribution of resources
3. More education of training of experts in language technology for Danish
4. More research on language technology for Danish

# A comprehensive lexical database



# Organisation



# Supplementary recommendations

- Change in the law of copyright deposit in order to make deposited material available for language technology
- Guidelines for collecting texts in public institutions and in publicly financed projects
- Mapping of translation tasks in public institutions and revision of translation contracts for public institutions wrt. to sharing of TM



# Conclusion

- Language technology is an important infrastructure
- We need a strategic approach
- Datasets must be collected and kept up to date
- Market forces are not sufficient, but a data sharing strategy will stimulate the market
- The state and public institutions play a crucial role.



Thank you!

Sabine Kirchmeier

[sabine@dsn.dk](mailto:sabine@dsn.dk)

Dansk  
Sprognævn

---