

# Dansk Sprognævn

## Sprog og sprogteknologi i Danmark

Sabine Kirchmeier

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

retskrivning sb., -en, -er, f  
sms. retskrivnings-, fx ret-  
skrivningssystem.  
retslig (el. retlig) adj., -t.  
retslæge sb., -n, -r.  
retslægeråd sb., -et, rets-  
lægeråd, bf. pl. -ene.  
retslærd adj., itk. d.s.  
retsløs adj., -t.

# Oversigt

- Sprog i Danmark
- Hvordan er status for sprogteknologi på dansk?
- Hvordan har man hidtil arbejdet strategisk forskningspolitisk med sprogteknologi?
- Hvad har man gjort i andre lande?

# Sprog i Danmark

- Danmark har et officielt sprog (Dansk).
- Tysk er officielt anerkendt som minoritetssprog i grænseområdet.
- Tegnsprog er ikke officielt anerkendt, men nyder særlig beskyttelse.
- Grønlandsk er sikret gennem selvstyrelovgivningen.
- Færøsk er sikret gennem hjemmestyreløven, som dog samtidigt stiller krav om at man også skal lære dansk.
- I kraft af rigsfællesskabet er der en forventning om at man skal kunne henvende sig til grønlandske og færøske myndigheder på dansk.

# Danmark og Norden

## Den Nordiske Sprogkonvention

- Vedtaget den 17. juni 1981.
- Sikrer nordiske borgeres ret til at bruge deres eget sprog i de andre nordiske lande.
- Omfatter mundtlig og skriftlig kontakt med statslige myndigheder så som domstole, social- og sundhedsmyndigheder, sundhedsvæsen, myndigheder som regulerer arbejdsforhold, skattemyndigheder, politi og skolevæsen.
- Dækker oversættelse og tolkning og permanente tjenester (i områder hvor der er et stort antal borgere som taler et andet sprog).
- Udgifterne betales af landene.

# Den Nordiske Sprogdeklaration

## Parallelsproglighed mellem engelsk og de nordiske sprog

Med parallelsproglighed menes en samtidig anvendelse af flere sprog inden for et eller flere områder. Det ene sprog fortrænger eller erstatter ikke det andet, men sprogene anvendes parallelt.

### Parallelsproglighed mellem engelsk og Nordens sprog

- At såvel Nordens samfundsbærende sprog som engelsk bør kunne anvendes som videnskabelige sprog
- At formidling af videnskabelige resultater på Nordens samfundsbærende sprog bør være meriterende
- At undervisningen i videnskabeligt fagsprog, især det skriftlige, bør tilbydes på både engelsk og på Nordens samfundsbærende sprog
- At universiteter og andre videnskabelige institutioner vil udvikle langsigtede strategier for sprogvalg, parallelsproglighed, sprogundervisning og oversættelsesstøtte inden for deres virkefelter
- At nordiske terminologiorganer kan fortsætte koordineringen af terminologien inden for nye områder
- At virksomheder og arbejdsmarkedsorganisationer bør opfordres til at udvikle strategier for parallelsproglighed

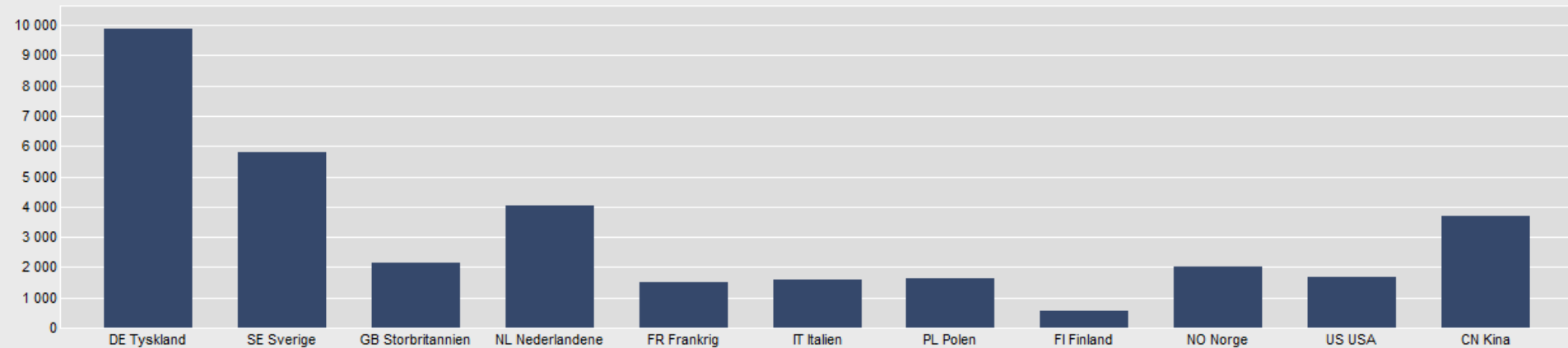
# Kontakter med andre sprog

- Indvandring
- Turisme
- Handel
- Regionalt samarbejde
  - mobilitet og trafik
  - energi og klimaforandringer
  - miljø og naturressurser
- Juridiske forhold på tværs af grænserne
  - ægteskab, fødsel, forældreret, aftaleret mv.

retskrivning sb., -en, -er, f  
sms. retskrivnings-, fx ret-  
skrivningssystem.  
retslig (el. retlig) adj., -t.  
retslæge sb., -n, -r.  
retslægeråd sb., -et, rets-  
lægeråd, bf. pl. -ene.  
retslærd adj., itk. d.s.  
retsløs adj., -t.

# Import. Udenrigshandel med varer.

Den samlede udenrigshandel  
efter land.  
2015M12, SÆSON Sæsonkorrigeret, 11 Import (ekskl. skibe og fly). (Mio. kr.)

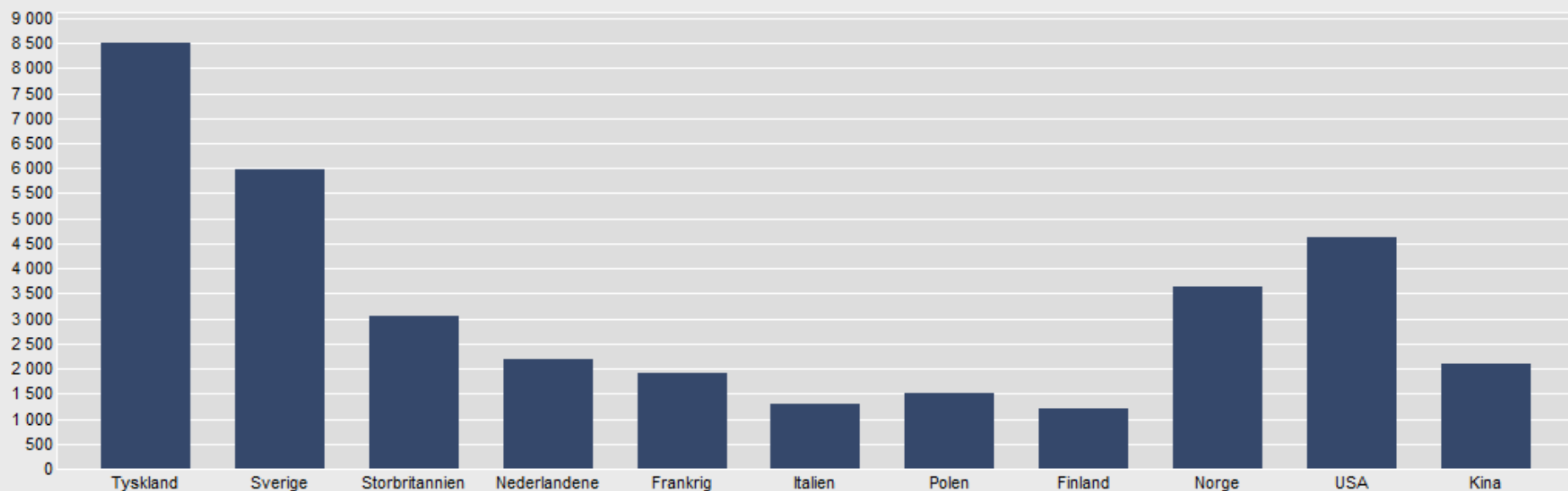


Kilde: Danmarks Statistik

retskrivning sb., -en, -er, f  
sms. retskrivnings-, fx ret-  
skrivningssystem.  
retslig (el. retlig) adj., -t.  
retslæge sb., -n, -r.  
retslægeråd sb., -et, rets-  
lægeråd, bf. pl. -ene.  
retslærd adj., itk. d.s.  
retsløs adj., -t.

# Eksport. Udenrigshandel med varer.

Den samlede udenrigshandel  
efter land.  
2015M12, Sæsonkorrigeret, Eksport (ekskl. skibe og fly). (Mio. kr.)



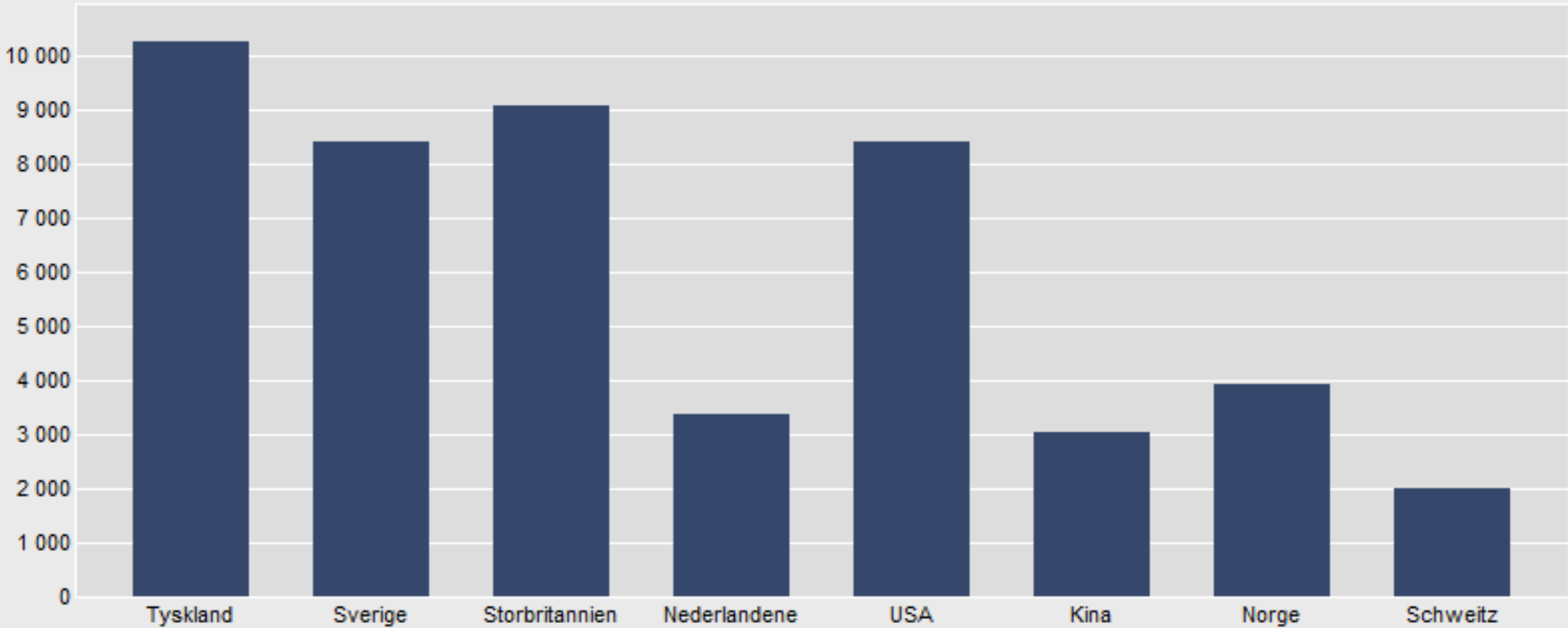
Kilde: Danmarks Statistik



retskrivning sb., -en, -er, f  
sms. retskrivnings-, fx ret-  
skrivningssystem.  
retslig (el. retlig) adj., -t.  
retslæge sb., -n, -r.  
retslægeråd sb., -et, rets-  
lægeråd, bf. pl. -ene.  
retslærd adj., itk. d.s.  
retsløs adj., -t.

# Import. Udenrigshandel med tjenesteydelser

Tjenestehandel  
efter land.  
Sæsonkorrigeret, Import, 2015K4. (Mio. kr.)

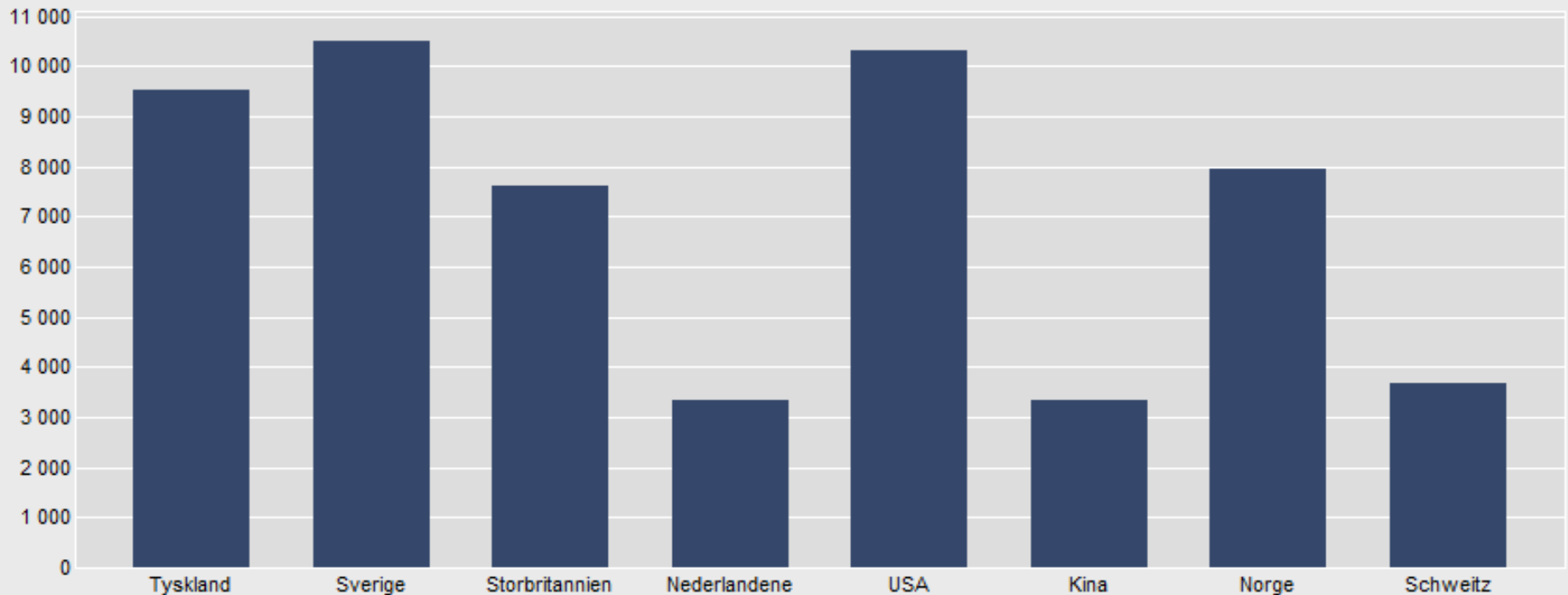


Kilde: Danmarks Statistik

retskrivning sb., -en, -er, f  
sms. retskrivnings-, fx ret-  
skrivningssystem.  
retslig (el. retlig) adj., -t.  
retslæge sb., -n, -r.  
retslægeråd sb., -et, rets-  
lægeråd, bf. pl. -ene.  
retslærd adj., itk. d.s.  
retsløs adj., -t.

# Eksport. Tjenesteydelser

Tjenestehandel  
efter land.  
Sæsonkorrigeret, Eksport, 2015K4. (Mio. kr.)

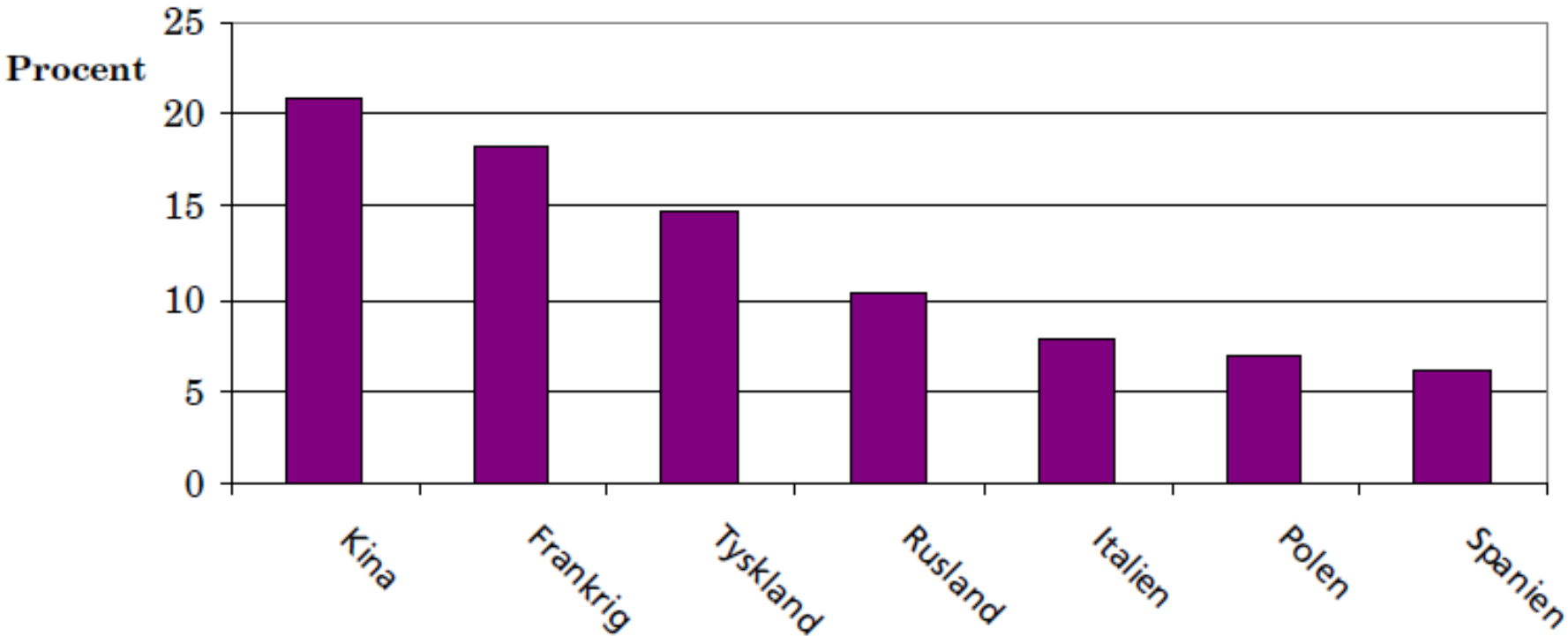


Kilde: Danmarks Statistik

retskrivning sb., -en, -er, l  
sms. retskrivnings-, fx ret-  
skrivningssystem.  
retslig (el. retlig) adj., -t.  
retslæge sb., -n, -r.  
retslægeråd sb., -et, rets-  
lægeråd, bf. pl. -ene.  
retslærd adj., itk. d.s.  
retsløs adj., -t.

# Dansk Industri og Ja til sprog

Lande hvor virksomhederne har oplevet kommunikationsproblemer

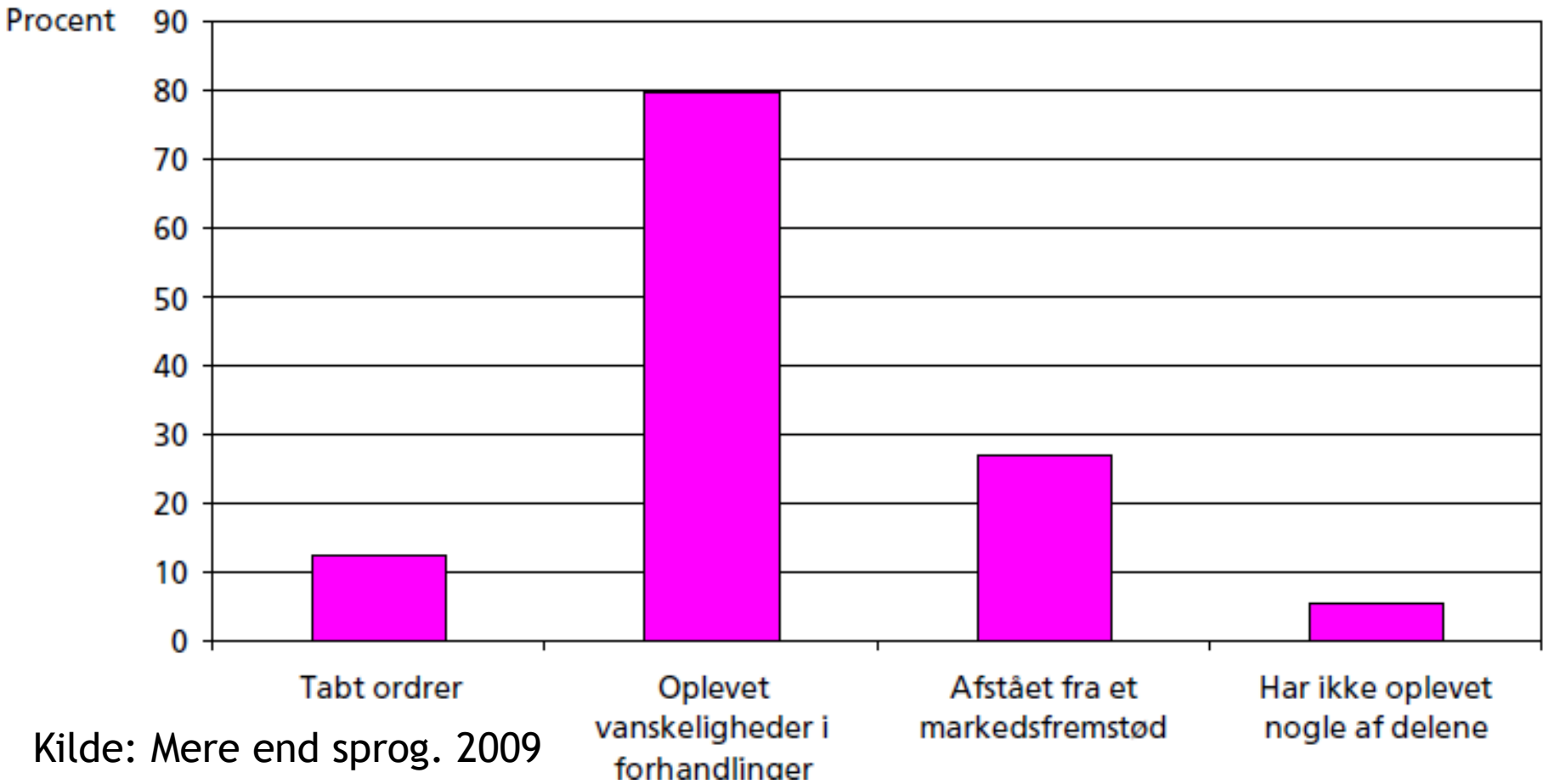


Kilde: Mere end sprog. 2009

retskrivning sb., -en, -er, l  
sms. retskrivnings-, fx ret-  
skrivningssystem.  
retslig (el. retlig) adj., -t.  
retslæge sb., -n, -r.  
retslægeråd sb., -et, rets-  
lægeråd, bf. pl. -ene.  
retslærd adj., itk. d.s.  
retsløs adj., -t.

# Dansk Industri og Ja til sprog

## Problemer der kan tilskrives mangel på sproglige kompetencer



Kilde: Mere end sprog. 2009

# Dansk Sprognævn

Status for sprogteknologi på dansk

# Dansk Sprogteknologi

- Kvaliteten for dansk er ringe
  - Google Translate har hidtil været det bedste bud
  - Danskerne kan som regel først langt senere end andre bruge fx taleinput til deres telefoner, fx begyndte Siri først at virke på dansk i 2014, mens det har virket på engelsk, fransk og tysk siden 2011.
  - Microsofts nye stavekontrol for dansk er utilfredsstillende
  - Kommunernes talegenkendelse bliver udviklet af udenlandske virksomheder som ikke er gode til dansk.
- Der mangler infrastruktur
  - frit tilgængelige datasamlinger og basisprogrammer
- Kun få forsknings- og udviklingsmidler til dansk sprogteknologi i de sidste årtier
- Ekspertisen i dansk sprogteknologi forsvinder



# For dansk er helt grundlæggende problemer stadig ikke løst

Vi finder stadig produkter som ikke kan håndtere de danske bogstaver

æ, ø, å => ae, oe, aa

De kan bruges i webadresser, men ikke i e-mailadresser

Det går ofte galt med alfabetiseringen

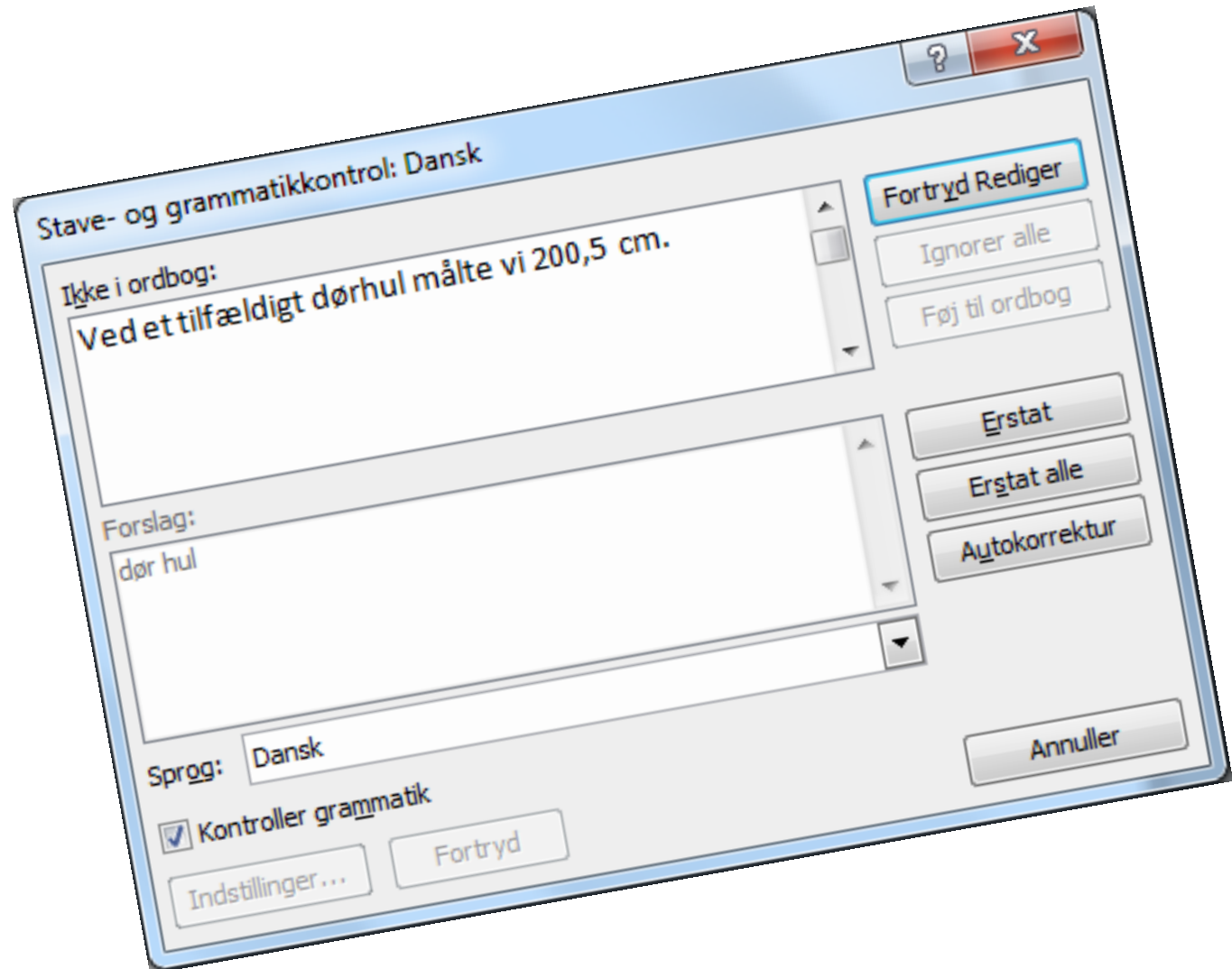
Århus => Aarhus

Børnefonden => Bornefonden

retskrivning sb., -en, -er, l  
sms. retskrivnings-, fx ret-  
skrivningssystem.  
retslig (el. retlig) adj., -t.  
retslæge sb., -n, -r.  
retslægeråd sb., -et, rets-  
lægeråd, bf. pl. -ene.  
retsløst adj., itk. d.s.  
retsløs adj., -t.

retskrivning sb., -en, -er, f  
sms. retskrivnings-, fx ret-  
skrivningssystem.  
retslig (el. retlig) adj., -t.  
retslæge sb., -n, -r.  
retslægeråd sb., -et, rets-  
lægeråd, bf. pl. -ene.  
retslærd adj., itk. d.s.  
retsløs adj., -t.

# Dansk stavekontrol i Word 2010 understøtter ikke korrekt dansk





retskrivning sb., -en, -er, i  
sms. retskrivnings-, fx ret-  
skrivningssystem.  
retslig (el. retlig) adj., -t.  
retslæge sb., -n, -r.  
retslægeråd sb., -et, rets-  
lægeråd, bf. pl. -ene.  
retslærd adj., itk. d.s.  
retsløs adj., -t.

# Sprogteknologiske anbefalinger

- Strategisk satsning på sprogteknologi (2004)
- Sprog til tiden (2008)
- Dansk Sprogs status (2012)
- Det danske sprog i den digitale tidsalder  
META-NET (2012)



retskrivning sb., -en, -er, f.  
 sms. retskrivnings-, fx retskrivningssystem.  
 retslig (el. retlig) adj., -t.  
 retslægeråd sb., -n, -r.  
 lægeråd, bf. pl. -ene.  
 retslærd adj., itk. d.s.  
 retsløs adj., -t.

# Dansk sprogteknologi ifølge META-net-rapporten (2012)

Excellent support	Good support	Moderate support	Fragmentary support	Weak/no support
	English	Czech Dutch Finnish French German Italian Portuguese Spanish	Basque Bulgarian Catalan <b>Danish</b> ←	Croatian Icelandic Latvian Lithuanian Maltese Romanian

Taleteknologi

8: Speech processing: state of language technology support for 30 European languages

Excellent support	Good support	Moderate support	Fragmentary support	Weak/no support
	English	Czech Dutch French German Hungarian Italian Polish Spanish Swedish	Basque Bulgarian Catalan Croatian <b>Danish</b> ←	Icelandic Irish Latvian Lithuanian Maltese

Sprogresurser til tekst og tale

11: Speech and text resources: State of support for 30 European languages

Excellent support	Good support	Moderate support	Fragmentary support	Weak/no support
	English	Dutch French German Italian Spanish	Basque Bulgarian Catalan Czech <b>Danish</b> ←	Croatian Icelandic Irish Latvian Lithuanian Maltese Serbian

Tekstanalyse

10: Text analysis: state of language technology support for 30 European languages

Excellent support	Good support	Moderate support	Fragmentary support	Weak/no support
	English	French Spanish	Catalan Dutch German Hungarian Italian Polish Romanian	Basque Bulgarian Croatian Czech <b>Danish</b> ←

Maskinoversættelse

9: Machine translation: state of language technology support for 30 European languages

# Dansk Sprognævn

Hvordan har man hidtil arbejdet strategisk forskningspolitisk med sprogteknologi?

# Sprogteknologiske aktører i Danmark

- Forskningsmiljøer: KU, CBS, SDU (AAU, DTU, ITU)
- Virksomheder: Ankiro, PDC, MIRSK, Mikroværkstedet, Microsoft, Goggle, IBM
- Forskningsråd- og forskningsprogrammer
  - Danske
  - Nordiske
  - EU

# Sprogpolitiske aktører

- Politikere
- Kommuner
- Staten
  - Digitalisering i det offentlige - PSI-direktivet
  - Forskningsinfrastruktur
  - Digitaliseringen af kulturarven
- Faglige Interessegrupper
  - OS2Talk. Offentligt digitaliseringsfællesskab. Fælleskommunalt samarbejde om talegenkendelse (OS2 -KOMBIT -CBS)
  - Terminologigruppen
  - FORVIR - Forum for Vidensmodellering i Offentligt Regi
  - Nordiske Sprognævn: Arbejdsgruppen for sprogteknologi i Norden (ASTIN)
  - Sprognævnets fagråd for sprogteknologi

# Sprogteknologi og terminologi i Folketinget

## Dansk Sprognævn

retskrivning sb., -en, -er, i  
sms. retskrivning, for ret-  
skrivning, sb.  
retslig (el. retlig) adj., -t.  
retslæger sb., -en, -t.  
retslægeråd sb., -et, rets-  
lægeråd, bf. pl. -ene.  
retslærd adj., itk. d.s.  
retsløs adj., -t.

# Sprogteknologi og terminologi i Folketinget

retskrivning sb., -en, -er, f  
sms. retskrivnings-, fx ret-  
skrivningssystem.  
retslig (el. retlig) adj., -t.  
retslæge sb., -n, -r.  
retslægeråd sb., -et, rets-  
lægeråd, bf. pl. -ene.  
retslærd adj., itk. d.s.  
retsløs adj., -t.

2008 Sprog til tiden

2009 Beretning i Folketingets Kulturudvalg (S, DF, SF og EL)

*Flertallet i udvalget vil have et terminologicenter oprettet under Dansk Sprognævn, som skal drive en offentligt tilgængelig flersproglig termbank.*

*Terminologicenteret skal understøtte arbejdet med fagsproget i de faglige miljøer og sikre videndeling indbyrdes mellem miljøerne og det omgivende samfund.*

Nedstemt. Der blev iværksat en sprogkampagne (Gang i sproget 2009-2011).

# Sprogteknologi i Folketinget

2012 Dansk Sprogs status

2012 Beslutningsforslag nr. **B 13** Folketinget 2012-13. Dansk Folkeparti

## *”VII. Sprogteknologiske tiltag*

*Der sattes fuldt og helt på udvikling af dansk sprogteknologi i form af erhvervsstøtteordninger og øremærkede midler til sprogteknologisk forskning og infrastruktur.*

## *VIII. Sprogtermbank*

*Offentligheden skal have adgang til en national flersproget termbank, hvilket var en af de vigtige anbefalinger i »Sprog til tiden« fra 2008. Det vil gøre det lettere at undervise og formidle på dansk, så det helt undgås at udbyde uddannelser på engelsk for danske studerende.”*

Nedstemt. Det blev besluttet at der skal udarbejdes jævnlige statusrapporter for brugen af engelsk på universiteterne.



# Tiltag til udvikling af sprogteknologi i Danmark

- Eurotra (1984-1982)
- SHF/FKK
  - Constraint grammar dansk, engelsk, portugisisk (VISL 1995-2000 (?))
  - UDOG (1993-1997) korpusbaseret udforskning af ordforråd og grammatik
  - CMOL (Computational modelling of language 2003-2007) statistiske metoder
  - DANNET (2005-2008) (bygger på EU-projekter (PAROLE og SIMPLE)
  - Semantic processing across domains (CST/DSL 2013-2016)
- STVF (1998-2002) fokus på udvikling af komponenter til taleteknologi
- Tværvideenskabelige IT-forskningsprogrammer under forskningsrådene
  - Spoken language dialogue systems (1991-1995)
  - Ontoquery (1999-2004)
- Private fonde
  - Carlsberg
  - Velux: DANTERMbank (2010-2015)

Kilde: Strategisk satsning på sprogteknologi (FKK) 2004

retskrivning sb., -en, -er, f  
sms. retskrivnings-, fx ret-  
skrivningssystem.  
retslig (el. retlig) adj., -t.  
retslæge sb., -n, -r.  
retslægeråd sb., -et, rets-  
lægeråd, bf. pl. -ene.  
retslærd adj., itk. d.s.  
retsløs adj., -t.

# Regeringsinitiativer og kommunale initiativer

- Regeringsinitiativer
  - Udvikling af talesyntese 1998-2001
  - Forskning i multimedia (1997-2002) Staging: dialog med tale, håndbevægelser mm.
  - It for alle (2003) talesyntese til blinde, svagtsynede og læsesvage
  - IT-infrastruktur
    - Clarin-dk (2008-2011) indsamling og tilgængeliggørelse af tekst, tale og billeder til humanistisk forskning
    - DigHumLab (2012- ?)
    - Deic (Dansk e-infrastruktur 2012-2016) skal sikre den bedst mulige nationale resurseudnyttelse på e-infrastrukturuområdet
- Kommunale initiativer
  - Taleteknologi OSTALK, KOMBIT i samarbejde med DanCast (CBS)
- Problem: Kun få midler. Ingen sammenhæng mellem initiativer.

# Dansk Sprognævn

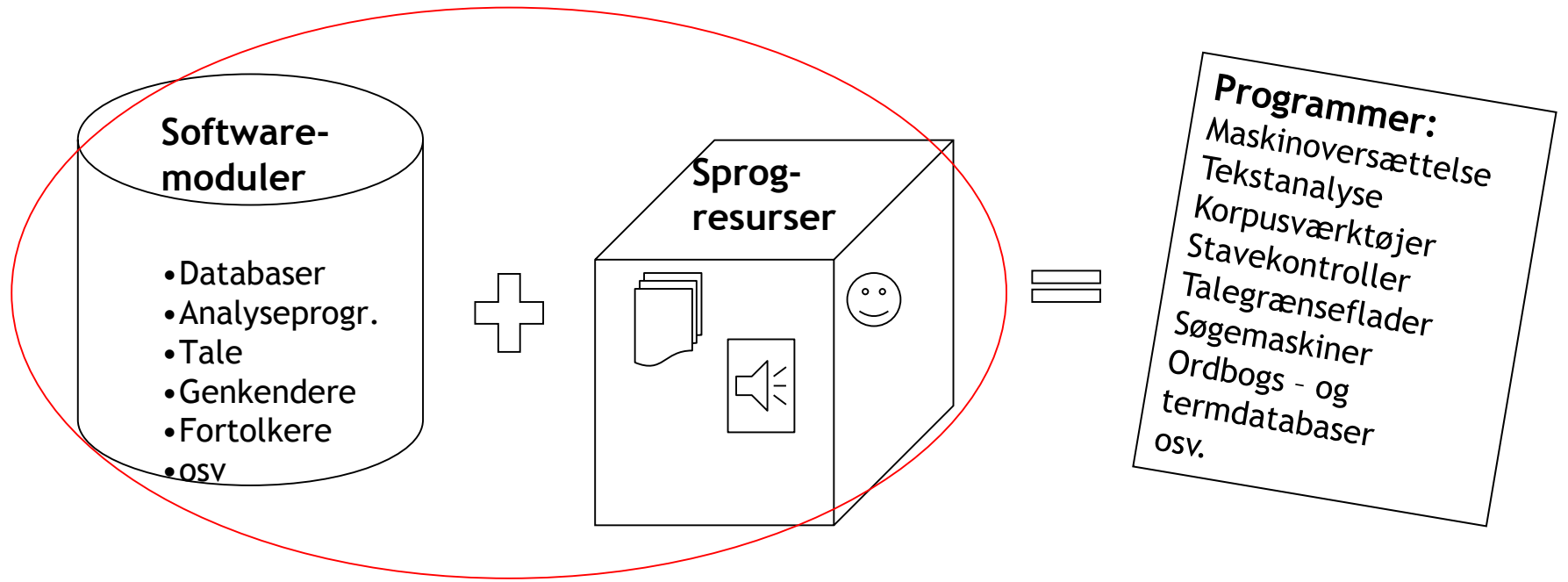
Hvad har man gjort i andre lande?

# Hvad har man gjort i andre lande?

- Nederlandsk (Holland og Belgien)
  - Eurotra (1984-1992)
  - Stevin (2004-2009)
  - => **BLARK (Basic language kit)**
  - CLARIN 2009-2015
- Letland (og andre nyere EU-medlemslande)
  - Statsstøttede projekter fra 2005-2009 - systematisk opbygning af sprogteknologikomponenter
- Sverige, Norge, Finland
  - Termbanker (nationale)
  - Store infrastrukturprojekter
  - Flere statslige initiativer (udarbejdelse af **BLARK** for norsk og svensk)

retskrivning sb., -en, -er, f  
sms. retskrivnings-, fx ret-  
skrivningssystem.  
retslig (el. retlig) adj., -t.  
retslæge sb., -n, -r.  
retslægeråd sb., -et, rets-  
lægeråd, bf. pl. -ene.  
retslærd adj., itk. d.s.  
retsløs adj., -t.

# BLARK - Sprogteknologiens basisbygggesten



## Software-moduler

- Databaser
- Analyseprogr.
- Tale
- Genkendere
- Fortolkere
- OSV

## Sprog-resurser

## Programmer:

- Maskinoversættelse
- Tekstanalyse
- Korpusværktøjer
- Stavekontroller
- Talegrænseflader
- Søgemaskiner
- Ordbogs - og termdatabaser
- osv.

retskrivning sb., -en, -er, f  
sms. retskrivnings-, fx ret-  
skrivningssystem.  
retslig (el. retlig) adj., -t.  
retslæge sb., -n, -r.  
retslægeråd sb., -et, rets-  
lægeråd, bf. pl. -ene.  
retslærd adj., itk. d.s.  
retsløs adj., -t.

# BLARK prioriteringer for nederlandsk

## For language technology:

### Modules:

- Robust modular text pre-processing (tokenisation and named entity recognition)
- Morphological analysis and morphological disambiguation
- Syntactic analysis
- Semantic analysis

### Data:

- Monolingual lexicon
- Annotated corpus of text (a treebank with syntactic, morphological, and semantic structures)
- Benchmarks for evaluation

## For speech technology:

### Modules:

- Automatic speech recognition (including tools for robust speech recognition, recognition of speaker, and prosody)
- Text-to-speech (including tools for unit selection)
- Confidence measures
- Speaker identification (speaker identification as well as dialect identification)
- Automatic annotation of speech
- Tools for specific applications, such as transcription assistance, etc.
- Multi-modal speech corpora
- Multi-media speech corpora
- Multi-lingual speech corpora
- Benchmarks for evaluation

Only by dedicated programmes such as the STEVIN programme was it possible to create language resources and basic tools to be able to carry out research on language technology for the Dutch language, and to make it more attractive to companies to develop and offer products and services in the Dutch language.  
Meta-net, NL.

# Konklusion

- Sprogteknologi er infrastruktur.
- Der skal arbejdes efter en klar strategi.
- Der skal løbende indsamles data for at holde systemerne ajour.
- Offentlige institutioner kan spille en afgørende rolle.