



# Indsamling af sprogressourcer i Danmark til CEF – vanskeligheder og muligheder

**Bolette Sandford Pedersen &  
Lina Henriksen**  
Center for Sprogteknologi,  
Nordiske Studier og Sprogvidenskab



KØBENHAVNS UNIVERSITET

EU-workshop 8. oktober 2018



# Hvem er vi i DK

ELRC har to såkaldte **National Anchor Points** i hvert land.

I Danmark er

- Digitaliseringsstyrelsen og
- Dansk Sprognævn tilsammen '**Public Anchor Point**'
- Københavns Universitet, Center for Sprogteknologi, er '**Technical Anchor Point**'

Center for Sprogteknologi er ansvarlig for *at støtte omkring det tekniske når det drejer sig om at få indsamlet og uploadet parallelle tekster og terminologiske ordlister*



# Hvem er vi i DK

## ELRC-staben ved **Center for Sprogteknologi:**

- Bolette Sandford Pedersen [bspedersen@hum.ku.dk](mailto:bspedersen@hum.ku.dk)
- Lina Henriksen [linah@hum.ku.dk](mailto:linah@hum.ku.dk)
- Claus Povlsen [cpovlsen@hum.ku.dk](mailto:cpovlsen@hum.ku.dk)
- Sussi Olsen [saolsen@hum.ku.dk](mailto:saolsen@hum.ku.dk)





# Hvorfor indsamler vi sproressourcer?

Et godt oversættelsessystem kræver meget store mængder sproglige data:

- for hvert eneste sprog der skal oversættes til/fra
- for hvert fagområde der skal oversættes til/fra

Den eneste måde at forbedre systemets oversættelser til og fra dansk er derfor ved at vi i Danmark:

- leverer gode danske sprogdata inden for relevante fagområder i form af **tekster** og **terminologier**

# Hvilke slags sproressourcer

Der er brug for parallelle tekster og tosprogede termlister, der

- har relevant indhold i forhold de aktuelle institutioner
- findes på *både dansk og engelsk*

Der er brug for *mange*, gerne lange tekster, fx publikationer.

Termlister/terminologisamlinger må selvfølgelig også gerne være lange, men selv små lister/samlinger har interesse.



# Parallelle tekster

## Før graviditeten

- ▶ Du kan forberede en sund graviditet, allerede inden du bliver gravid. Begynd at tage et kosttilskud af folsyre. Og hold dig fra alkohol og rygning.

### Folsyre – et kosttilskud

Tag 400 mikrogram folsyre dagligt, fra du overvejer at blive gravid. Folsyre tilhører gruppen af B-vitaminer. Folsyre nedsætter risikoen for, at barnet bliver født med misdannelse af rygmarven (neuralrørsdefekt).

### Ingen alkohol

Alkohol kan skade barnet, fra graviditeten indtræder og frem. Da det kan være svært at kende det nøjagtige tidspunkt for undfangelsen, bør du lade være med at drikke alkohol, hvis du planlægger graviditet.

Hvis graviditeten er uplanlagt, og du har drukket alkohol, er der dog kun sjældent grund til bekymring. Tal evt. med lægen/jordemoderen om det ved første graviditetsbesøg.



op med at ryge, allerede i tiden planlægges.

### Tjek medicinen

Er du i behandling med medicin i den periode, hvor du plan-

## Before getting pregnant

- ▶ You can prepare for a healthy pregnancy even before you become pregnant. Start taking folic acid. And stay clear of alcohol and smoking.

### Folic acid – a dietary supplement

You should begin taking 400 mcg folic acid daily as soon as you start planning your pregnancy. Folic acid is part of the vitamin B group. Folic acid reduces the risk of your baby being born with neural tube defect.

### No alcohol

Alcohol can harm your baby from the start of pregnancy and onwards. As it is often difficult to know the exact time of conception, you should not drink alcohol if you are planning on having a baby.

However, if you find that you are pregnant without having planned it, and you have been drinking alcohol, it is rarely a cause for concern. You may like to talk to your doctor/midwife about this when going for your first consultation.



### Check medicines

If you are being treated with any medications when you begin planning your pregnancy, you should discuss



# Parallele termlister

Dansk-Engelsk Uddannelsesterminologi	
Udtræk pr. 2-2-2018	
Gældende version og bemærkninger findes her: <a href="http://ufm.dk/termbase">http://ufm.dk/termbase</a>	
Term	Oversætte Bemærkning
13-skala	00 to 13 grading scale
7-trins-skala	7-point grade se også syvtrinsskala
absolut karaktergivning	absolute grading
AC	se Akademikernes Centralorganisation
ad hoc-udvalg	ad hoc committee
adfærdsforstyrrelse	behavioural disorder
adfærdsvanskelig	ebd (emotional and behavioural difficulties)
adgang (til beskæftigelse)	access
adgang (til uddannelse)	entry
adgang (til uddannelse)	access Jf. Lissabon-konventionen. I Lissabon-konve
adgang (til uddannelse)	admission
adgangsbegrænsning	restricted admission
adgangsbegrænsning	numerus c primært i tysk- og fransktalende lande
adgangsberettiget	eligible for admission
adgangsberettiget	entitled to admission
adgangsbetingelse	entry requirement
adgangsbetingelse	entrance requirement
adgangsbetingelse	admission requirement



# Sprogressourcer i ELRC

http://lr-coordination.eu/resources



Home Discover Resources Services Events Anchor Points News Helpdesk Re:










# Sprogressourcer i ELRC


Filter by:

- Language
- Resource Type
- Media Type
- Licence
- Conditions of Use
- Linguality Type
- Multilinguality Type
- Data Format
- Domain
- Appropriateness For DSI
- Publication Status


Resource Type:

- Corpus: 
- Lexical/Conceptual: *ab*
- Language Description: 
- Tool/Service: 

## 11 Language Resources

 **DA-EN Danish Health Authority**  
Danish | English

 **DA-EN Danish Ministry of Economic Affairs and the Interior**  
Danish | English

 **DA-EN Danish Ministry of Finance**  
Danish | English

 **DA-EN Danish Ministry of Finance 2**  
Danish | English

 **DA-EN Danish Ministry of Higher Education and Science**  
Danish | English

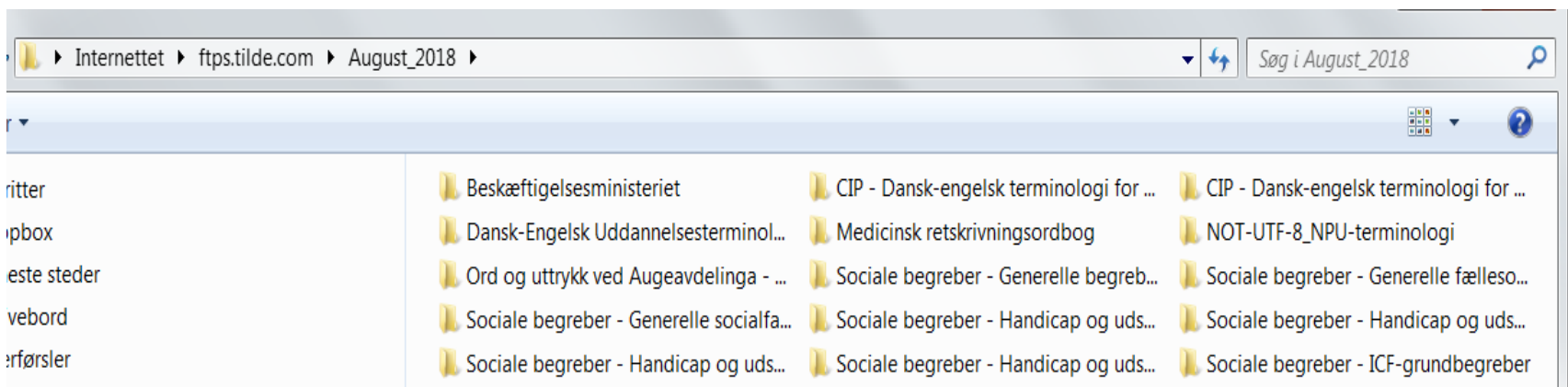
 **DA-EN Danish Ministry of Higher Education and Science 2**  
Danish | English



# Terminologiresourcer i ELRC (e-Translation Termbank)

Eks. på danske institutioner der allerede har bidraget med terminologi:

- Socialstyrelsen
- Beskæftigelsesministeriet
- Københavns Universitet
- Forsknings- og Uddannelsesministeriet
- Sundhedsdatastyrelsen
- Landsorganisationen i Danmark



# Udfordringer

Der er i DK for lidt opmærksomhed omkring:

- Hvor værdifulde tekster og termlister er for forskning og udvikling i dansk sprogteknologi
- At vi ikke har en klar fælles sprogpolitik og fælles fokus på sproglige problemstillinger i det offentlige (som i andre nordiske lande)
- At nogle offentlige institutioner hverken anvender terminologiske ressourcer el.lign. som kan sikre ensartet tekstkvalitet
- At vi ikke deler sproglige data inden for det offentlige (som i andre nordiske lande)
- At denne skepsis over for deling af sproglige data mellem offentlige institutioner har konsekvenser for Danmark



# Udfordringer

## Mulige årsager

- I DK har vi ikke tradition for den helt store fokus på det sproglige område
- Mange steder føler man stort ansvar og derfor stor nervøsitet ved at give data fra sig – selvom det er til formål inden for det offentlige
  - Kan data blive misbrugt?
  - Hvad nu hvis vi opdaterer vores data og denne opdatering ikke sker alle steder?
  - kan vores data ende med at gøre mere skade end gavn?



*Resultatet er alt for få store parallelle ressourcer*



# Mulige løsninger

Behov for

- større politisk viden og bevågenhed på området
- en mere proaktiv tilgang til feltet fx vha

*en fælles organiseret infrastruktur således at det bliver lettere og mere anerkendt at samle og dele data*



# Hvordan leverer man data

Som situationen er i dag

- Identificér de relevante tekster
- Angiv emne
- Angiv url for deres placering (eller send teksterne med)
- Tag stilling til hvad teksterne må bruges til:
  - må udelukkende bruges i CEF.AT-systemet
  - må derudover også bruges til forskning i sprog og sprogteknologi
  - må gøres tilgængelige for alle, herunder virksomheder der udvikler sprogteknologi for dansk



# Hvordan leverer man data

Send url eller tekster, emneangivelse og brugstilladelse til Københavns Universitet:

- Bolette Sandford Pedersen [bspedersen@hum.ku.dk](mailto:bspedersen@hum.ku.dk)
- Lina Henriksen [linah@hum.ku.dk](mailto:linah@hum.ku.dk)
- Claus Povlsen [cpovlsen@hum.ku.dk](mailto:cpovlsen@hum.ku.dk)
- Sussi Olsen [saolsen@hum.ku.dk](mailto:saolsen@hum.ku.dk)

Kontakt os for hjælp til at uploade tekster og ordlister eller upload dem selv under elrcs hjemmeside, kontakt evt. [elrc-share@ilsp.gr](mailto:elrc-share@ilsp.gr) for yderligere teknisk hjælp