



Statusrapport for dansk sprogteknologi 2022

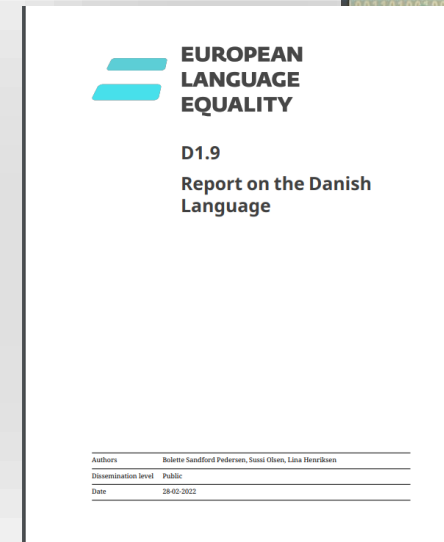
Bolette S. Pedersen, Center for Sprogteknologi, KU (inkl. slides fra Maria Giagkou)

Tre rapporter i nyere tid om status for dansk sprogteknologi:

2012 Det danske sprog i den digitale tidsalder

2018 Dansk sprogteknologi i verdensklasse

2022 Report on the Danish Language

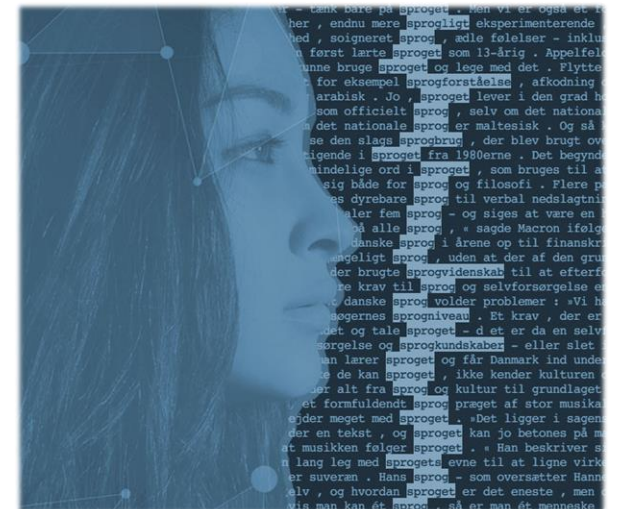


Highlights fra 2022-rapporten

Drejning mod sprogcentreret AI *øger fokus* på sprogteknologi i Danmark

Drejning mod brug af ‘pre-trained’ sprogmodeller som fine-tunes til specifikke opgaver

Større fokus på *deling af sprogressourcer* både i virksomhederne og på tværs af offentlige institutioner



Highlights fra 2022-rapporten

Nationale initiativer er nu i fuld gang under Digitaliseringsstyrelsen:

Sprogteknologi.dk



Det Centrale OrdRegister



Highlights fra 2022-rapporten

General opgradering af **sprogressourcer**, både *tekstkorpora* og *sprogmodeller*

Stor stigning i antallet af danske virksomheder og offentlige institutioner der arbejder med NLP – fra niche til ‘alle-mandseje’



Highlights fra 2022-rapporten

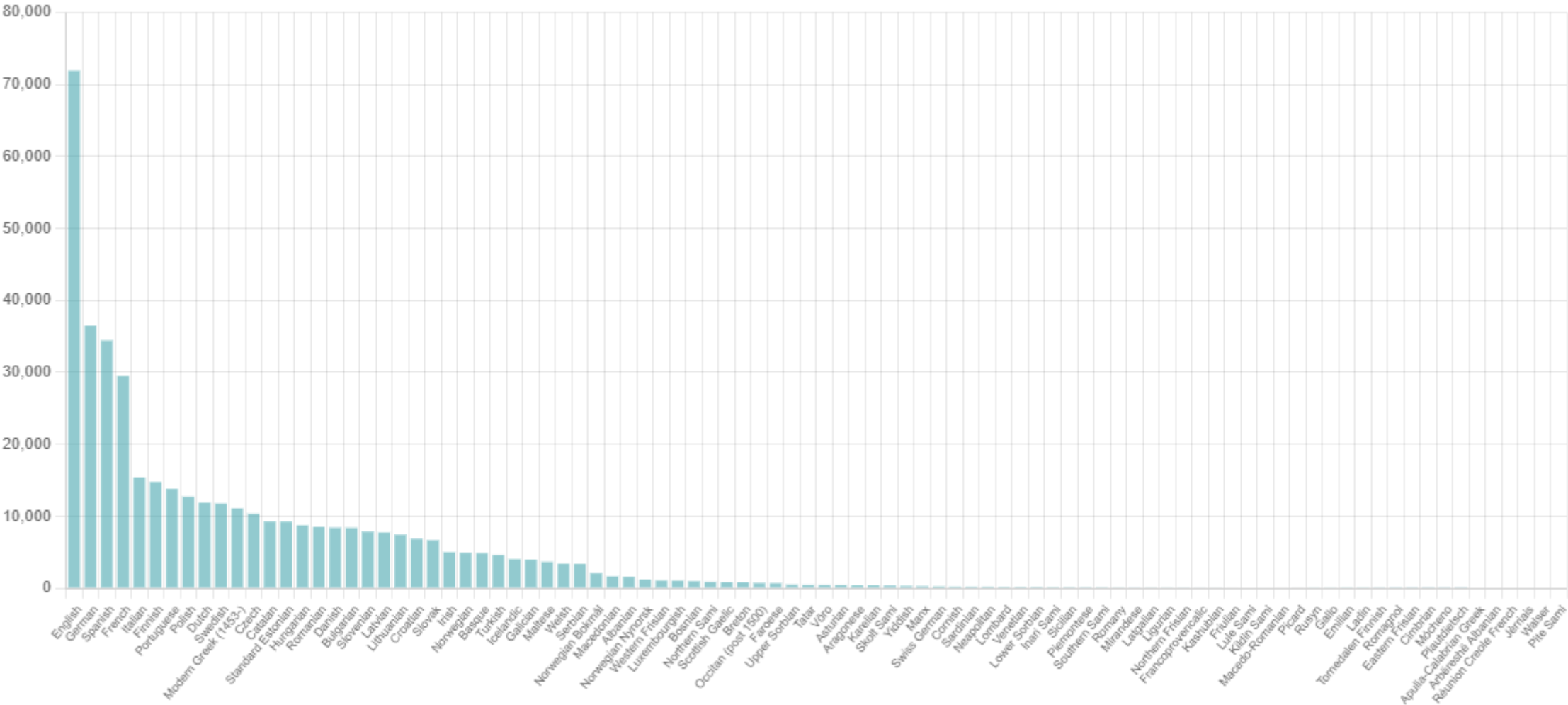
Store privatstøttede AI-initiativer i gang som også har NLP/sprogteknologi med, særligt:

AI Pioneer Centre

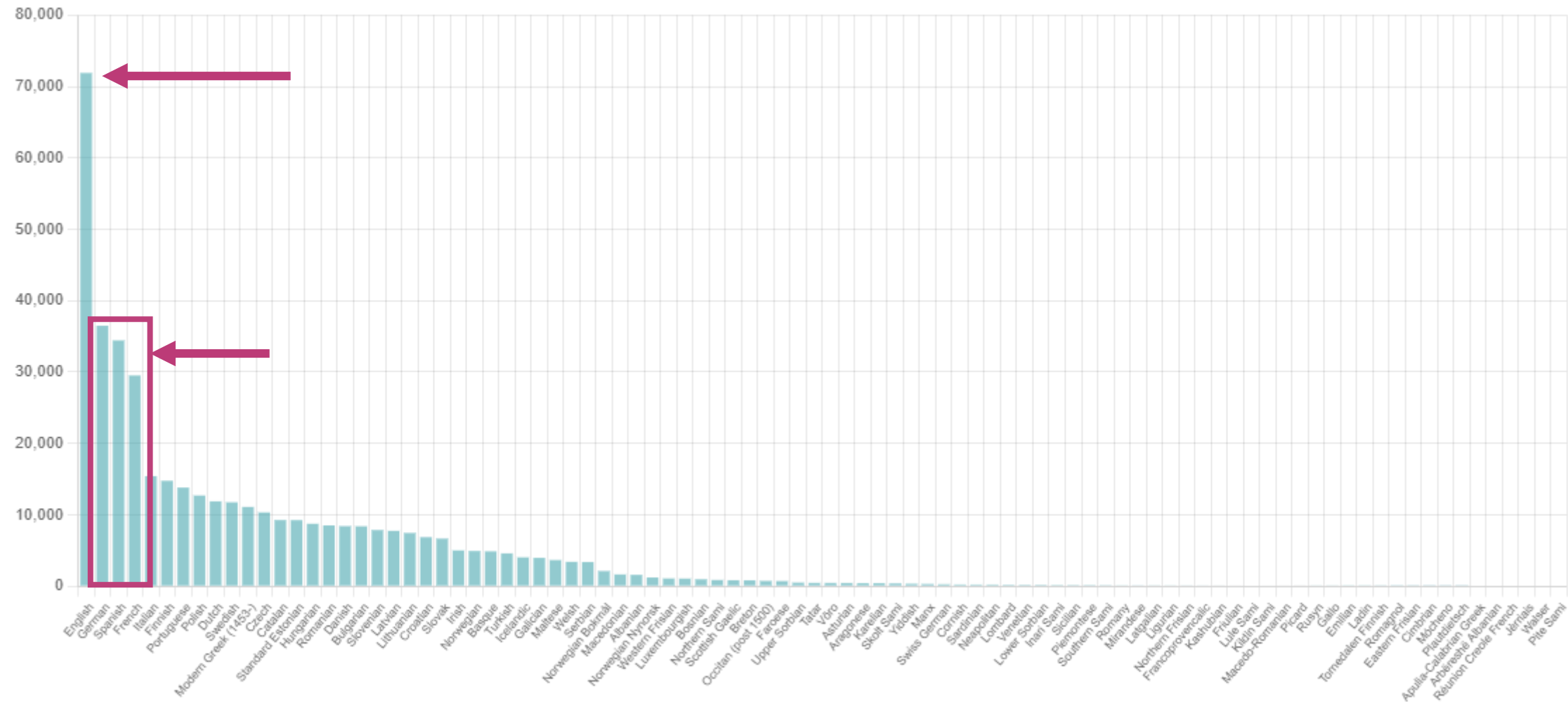
Carlsbergfondet, Novo Nordisk Fonden, Velux Fonden, Grundforskningsfonden (330 mio. kr.)



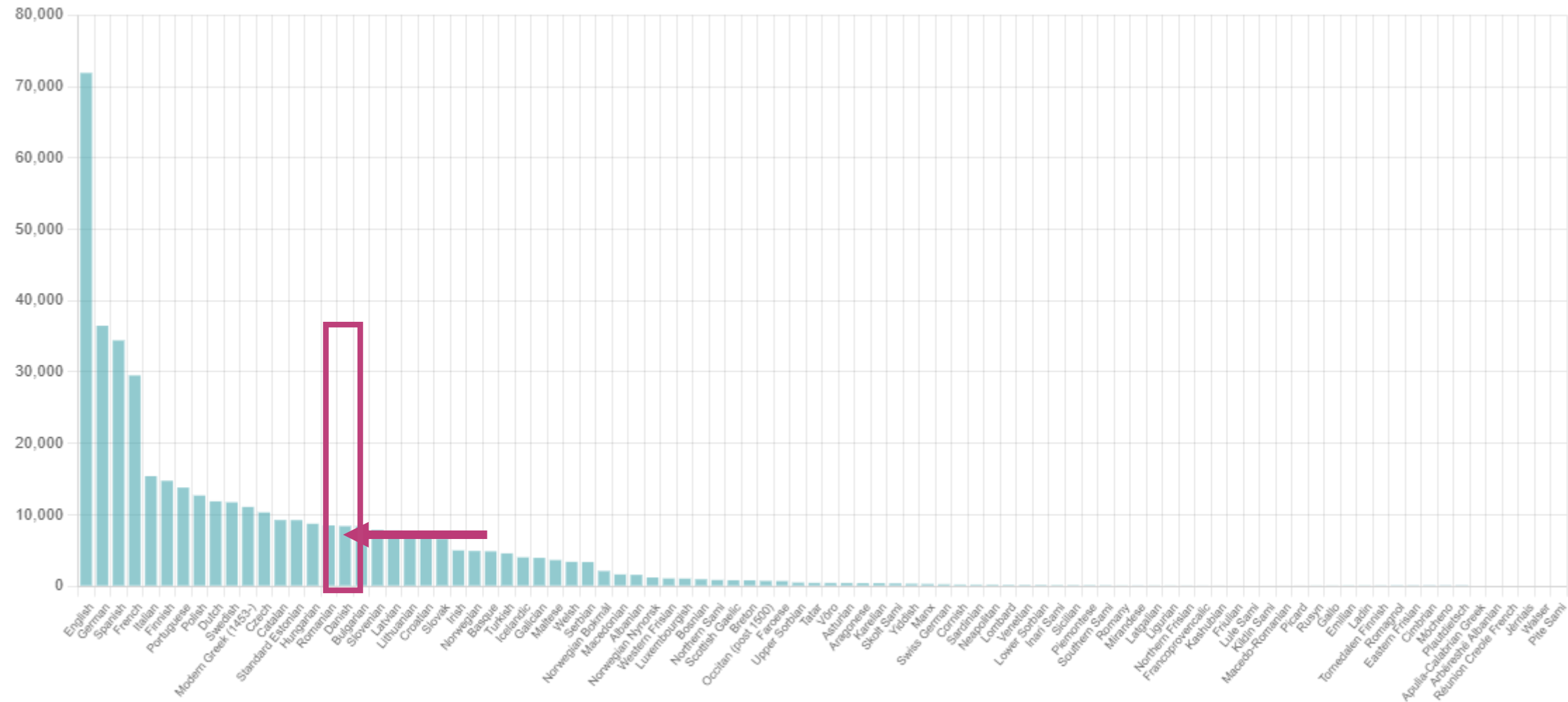
Stor ulighed i sprogteknologi



Stor ulighed i sprogteknologi



Stor ulighed i sprogteknologi



Hvordan måles uligheden?

Teknologiske faktorer

- Værktøjer og tjenester
- Korpora
- Sprogmodeller og sprogbeskrivelser
- Leksikalske og semantiske ressourcer

- Projekter
- Organisationer

Andre faktorer i sprogsamfundet

- Industri
- Online virksomhed

- Økonomi
- Uddannelse
- Bevillingsforhold
- Jura
- Medier
- Politik
- Offentlig administration
- Samfund

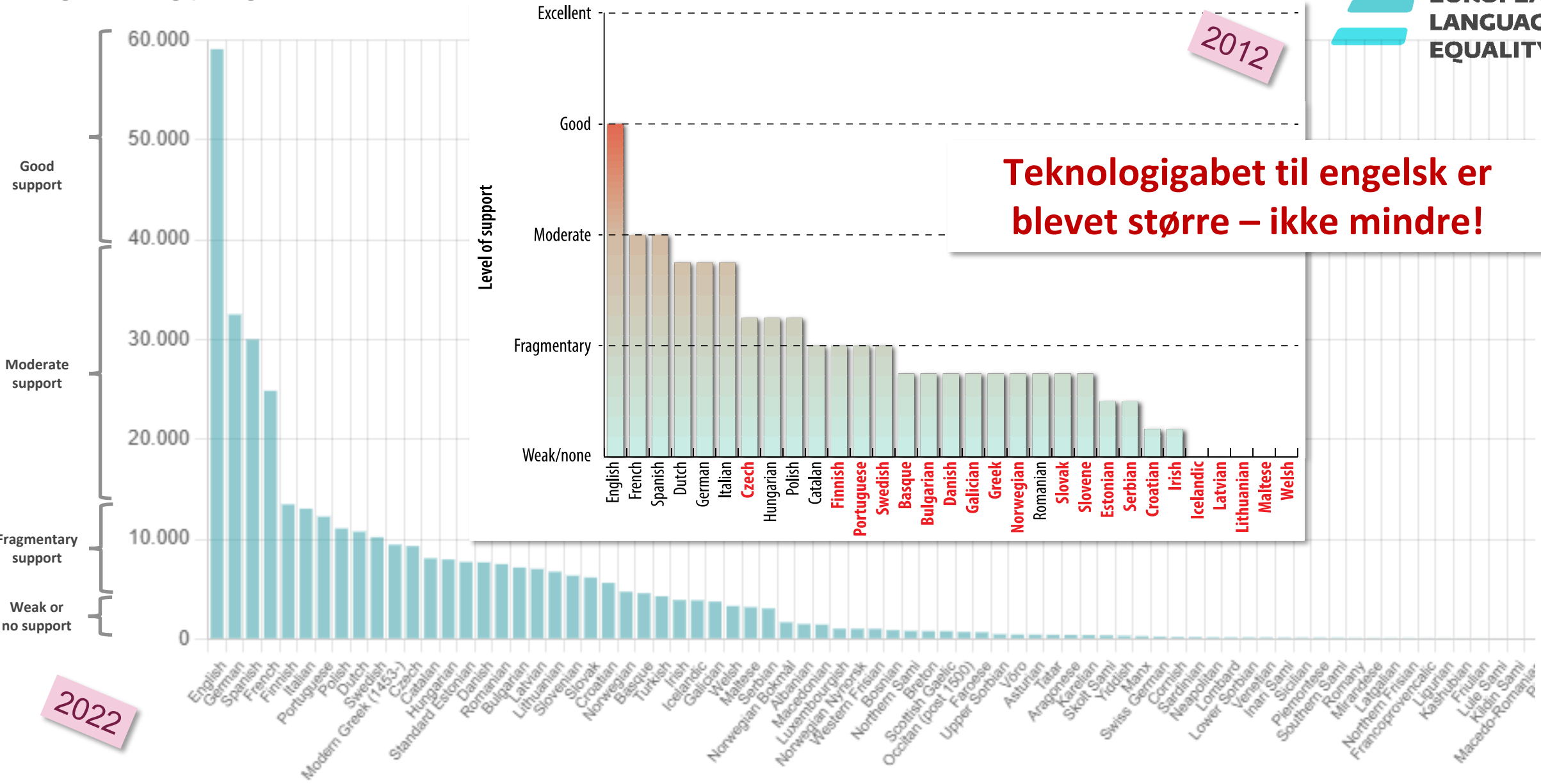
Dansk i sammenligning med andre sprog

Fire kategorier:

- God støtte (grøn)
 - Moderat støtte (lysegrøn)
 - Fragmentarisk støtte (gul)
 - Svag eller ingen støtte (lysegul)
-
- Kun i relation til maskinoversættelse ligger dansk i lysegrønt

		Tools and Services							Language Resources					
		Text Processing	Speech Processing	Image/Video Processing	Information Extraction and IR	Human-Computer Interaction	Translation Technologies	Natural Language Generation	Text Corpora	Multimodal Corpora	Parallel Corpora	Models	Lexical Resources	Overall
EU official languages	Bulgarian													
	Croatian													
	Czech													
	Danish													
	Dutch													
	English													
	Estonian													
	Finnish													
	French													
	German													
	Greek													
	Hungarian													
	Irish													
	Italian													
	Latvian													
	Lithuanian													
	Maltese													
	Polish													
	Portuguese													
	Romanian													
Slovak														
Slovenian														
Spanish														
Swedish														
(Co-)official languages	National level	Albanian												
		Bosnian												
		Icelandic												
		Luxembourgish												
		Macedonian												
		Norwegian												
		Serbian												
	Regional level	Basque												
		Catalan												
		Faroese												
		Frisian (Western)												
		Galician												
		Jerriais												
		Low German												
		Manx												
		Mirandese												
		Occitan												
		Sorbian (Upper)												
		Welsh												
All other languages														

2022 vs. 2012



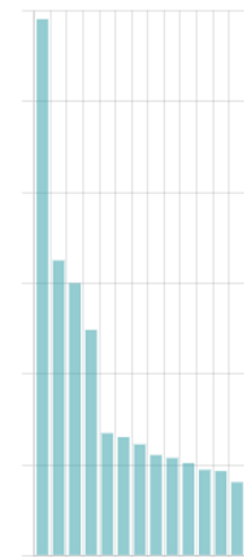
Konklusion

Større viden om og opmærksom på sprogteknologi i DK – og flere gode løsninger!

MEN: udviklingen går stærkt, engelsk og andre store sprog øger deres forspring

Fortsat fokus på udvikling af dansk sprogteknologi er vigtig!

God fornøjelse på Sprogteknologisk Konference 2022!!



Dagens program

09:30-
09:40 Velkomst v. Institutleder Anne Jensen, Nordiske Studier og Sprogvidenskab (NorS), KU

09:40 – Introduktion v. Jane Sofia Gyberg Rasmussen, kontorchef i
10:05 Kontor for teknologi og data i Digitaliseringsstyrelsen og
Bolette Sandford Pedersen, professor, Center for
Sprogteknologi, NorS, KU, herunder pointer fra den [nye
sprogteknologirapport for dansk](#)

10:05 – Hvordan kan NLP hjælpe i hjemmeplejen? v. Julie
10:25 Sohn, Systematic (Se [abstract](#))

10:25 – Kaffe
11:00

11:00 – Talegenkendelse for dansk v. Kasper Bay Noer og Anders Jess
11:30 Pedersen, Alexandra Instituttet, (Se [abstract](#)).

11:30 – Finding Meaning in Data across Languages v. Daniel
12:00 Hershovich, Datalogisk Institut, KU (Se [abstract](#))

12:00 – Poster pitch
12:20

12:20 – **Frokost, netværk og posters**
13:40

13:40 – Opening up dictionaries for natural language understanding
14:10 v. Simon Krek, Institut Jozef Stefan, Slovenien (se [abstract](#))

14:10 – Improving critical 911 patient communication with machine
14:30 learning v. Lars Maaløe, Corti (se [abstract](#))

14:30 – Tryg Digital By v. Sebastian Jørgensen, Nordic Safe Cities, og
14:50 Niels Ørbæk, Analyse og Tal (Se [abstract](#))

14:50 – Afrunding
15:00