

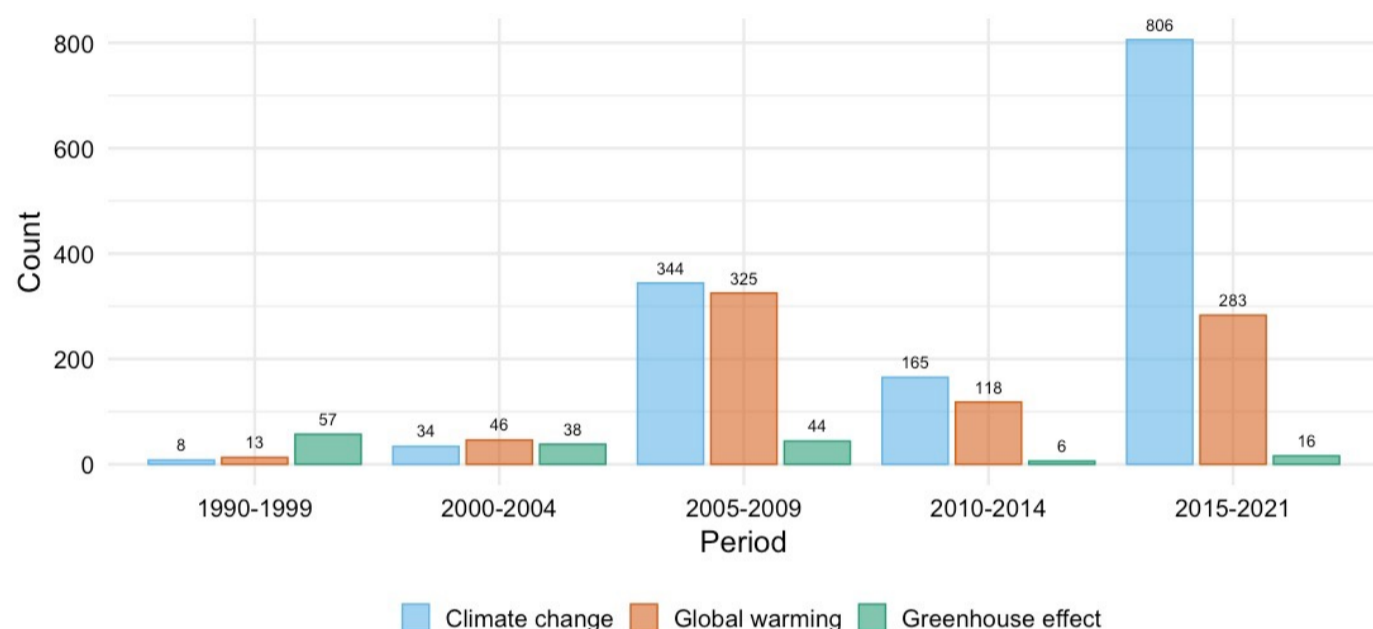
# Using BERT to Study Semantic Variations of Climate Change Keywords in Danish News Articles

Florian Meier

## 1. Background

- **Greenhouse effect (GE)**, **global warming (GW)** and **climate change (CC)** are often used interchangeably
- However, these terms have their own nuanced meanings, historical patterns, and communicative implications
- Their interpretation and use in CC communication have yet to be thoroughly examined
- Previous work has mostly relied on qualitative methods or costly to collect survey data
- We propose the use of masked language modelling (DanishBERT) for this task

## 2. Data and approach



### Masked language modelling

#### Original sentence:

Der er således stærke beviser på, at isbjørnene er påvirket af klimaforandringerne.

#### CC keyword masked:

Der er således stærke beviser på, at isbjørnene er påvirket af [MASK].

#### Predictions:

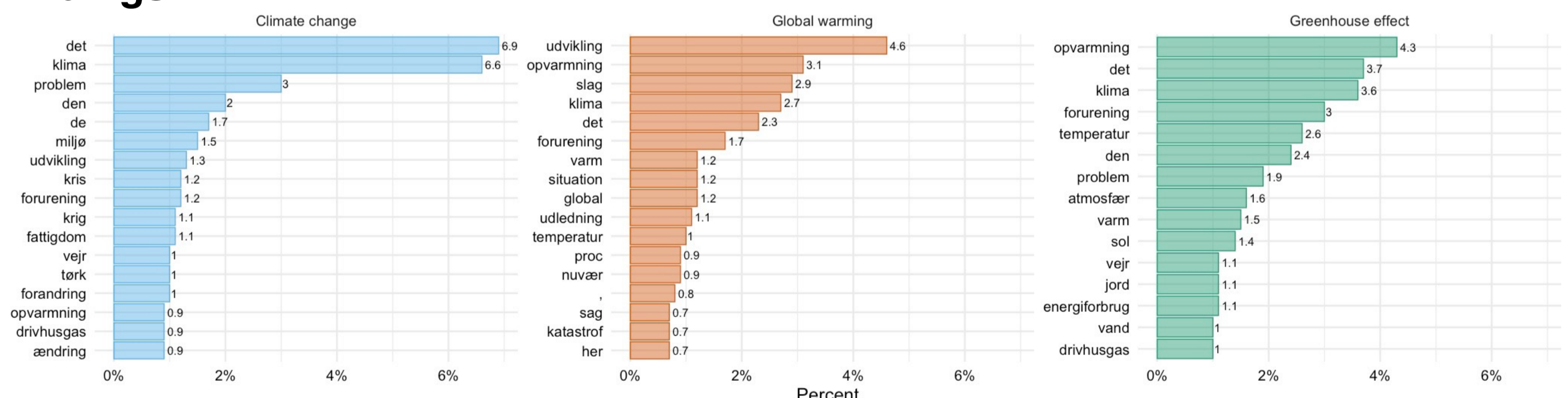
aber, natur, mennesker, is, naturen



- 32,214 articles containing one of the three keywords at least twice were sourced from **Infomedia** (9 different outlets, 1990 – 2021)
- Separation in 5 periods and random sampling of 5% of sentences per keyword per period
- GE 161 | GW 785 | CC 1357 = 2303 sentences
- High keyword diversity in 2005-2009

- We study the semantic variation of CC keywords by performing masked word prediction (fill-mask) using DanishBERT via **the Huggingface Inference API**
- These tasks are useful when we want to get a statistical understanding of the language and usage of words in which the model is trained in

## 3. Findings



- **Bootstrapped TTR<sub>100</sub> (n=1000):** CC 0.78, GW 0.81, GE 0.80
- Climate (klima), warming (opvarmning) and pollution (forurening) are always among the top 15 words
- **GE:** mainly linked to natural science terms: atmosphere (atmosfær) or greenhouse gas (drivhusgas)
- **CC:** words that emphasize the continuous change (forandring, ændring), the visible consequences (tørk, miljø and vejr); framed like other social issues, e.g. war (krig) or poverty (fattigdom)