

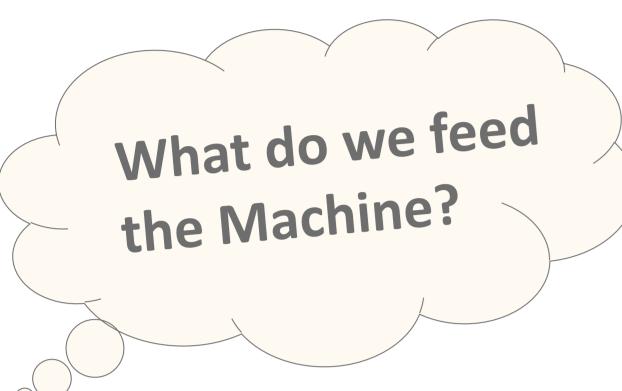
# Can a Machine have a Sense of Senses?

Creating Sense Representations for Danish

Nathalie Carmen Hau Sørensen

nmp@hum.ku.dk

#### Sense Tagging — How to assign a specific sense from a sense inventory (Dictionary) to each content word



Context

Hun skjuler

skjule\_1

target

ansigt\_1

i hænderne og begynder at græde
hånd\_1 begynde\_2 græde\_1

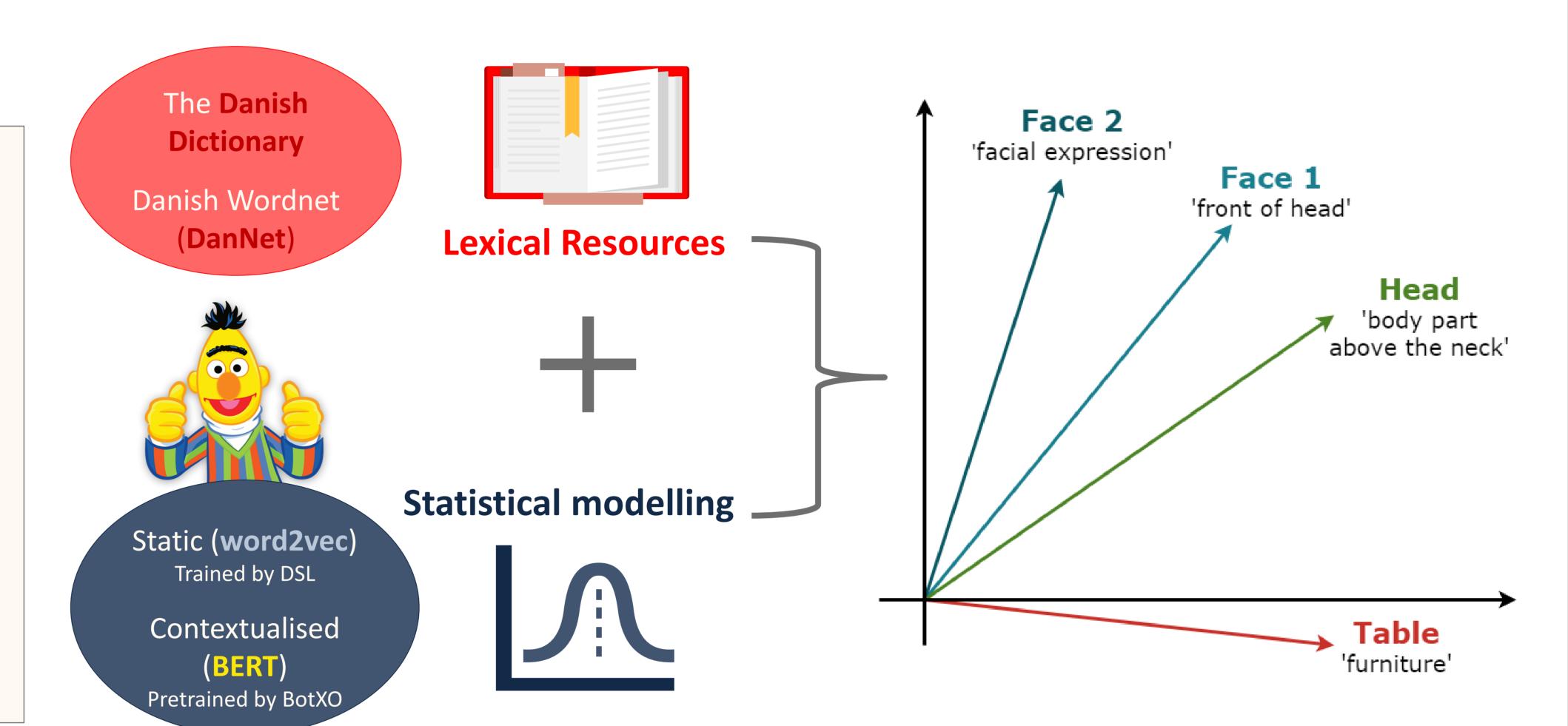
### Senses as vector

**Semantic vector space** where each sense is represented as a **vector** — a list of numbers where each dimension captures some information about the sense.

## **Hybrid approach**

Combining

- 1. The lexical information from human-crafted knowledge bases (Dictionary + WordNet)
- 2. The distributional information from a statistical model (word embedding model)



#### **Evaluation of Sense Vectors**

Intrinsic **Extrinsic DanWiC** SemDaX lexical Sample Word Sense Disambiguation Task Word-in-Context **Automatically** compiled **Human annotated** Collection from **DanNet** sense corpus Broad **Narrow** Coverage **1501** unique lemmas 18 highly polysemous nouns 3 different word classes Instances High Low per lemma Maximum 3 127-536 Explicit ■ Implicit Human Explicit ■ Implicit 0,81 0,72 0,4 0,341 WORD2VEC WORD2VEC BERT HUMAN **BERT** DANWIC SEMDAX

#### **Strong inclusion (explicit)**

Multiple knowledge sources

- DanNet synset connections
- Sense definition
- Quote with target lemma

Relies more on The lexical resources

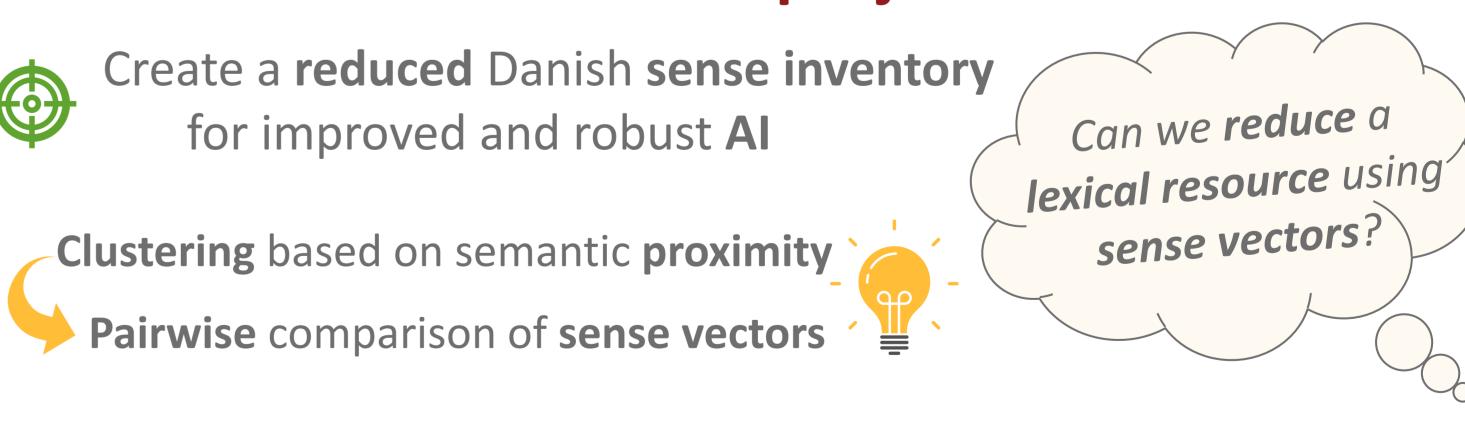
## Weak inclusion (implicit)

A single knowledge source

Just the definition or quote

Relies more on The statistical modelling

### Future Work – The COR-S project



Example of reduction from 4 to 2 senses for dag 'day'

