

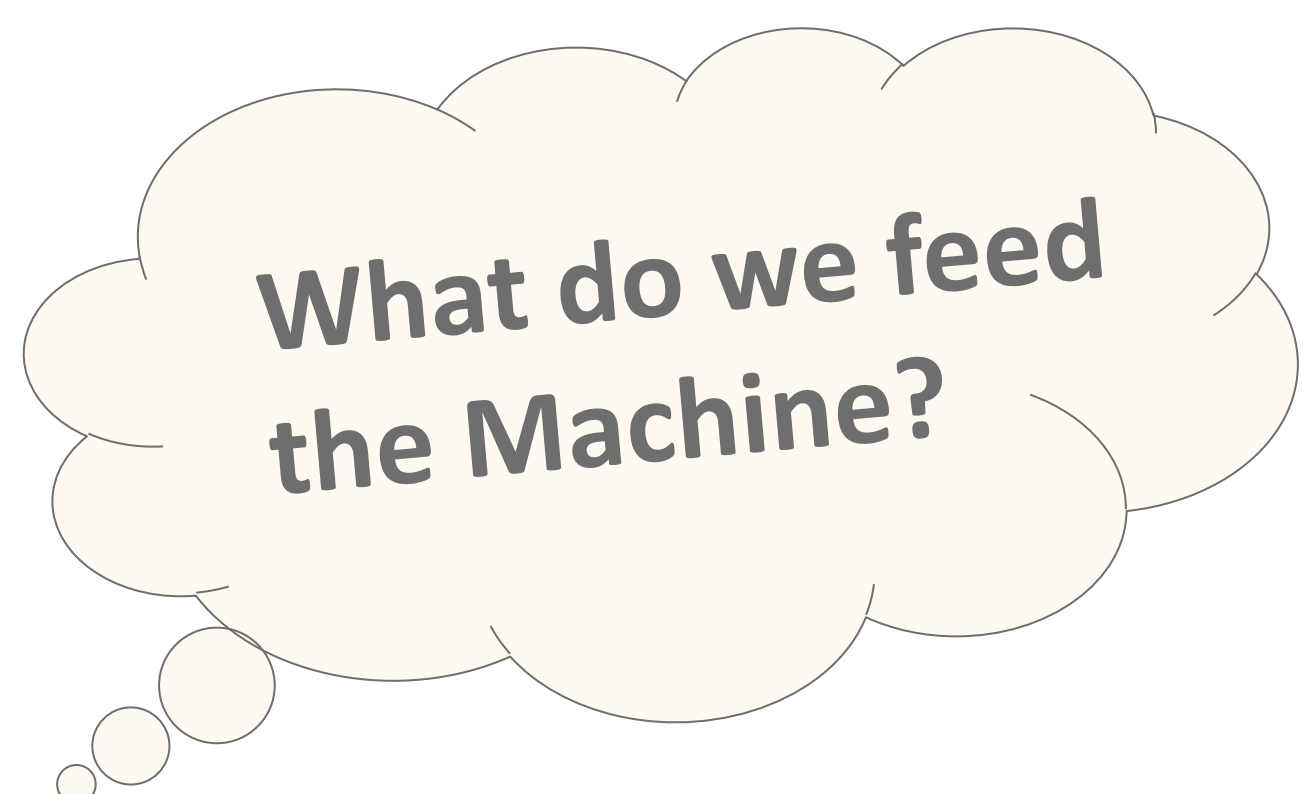
# Can a Machine have a Sense of Senses?

## 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**



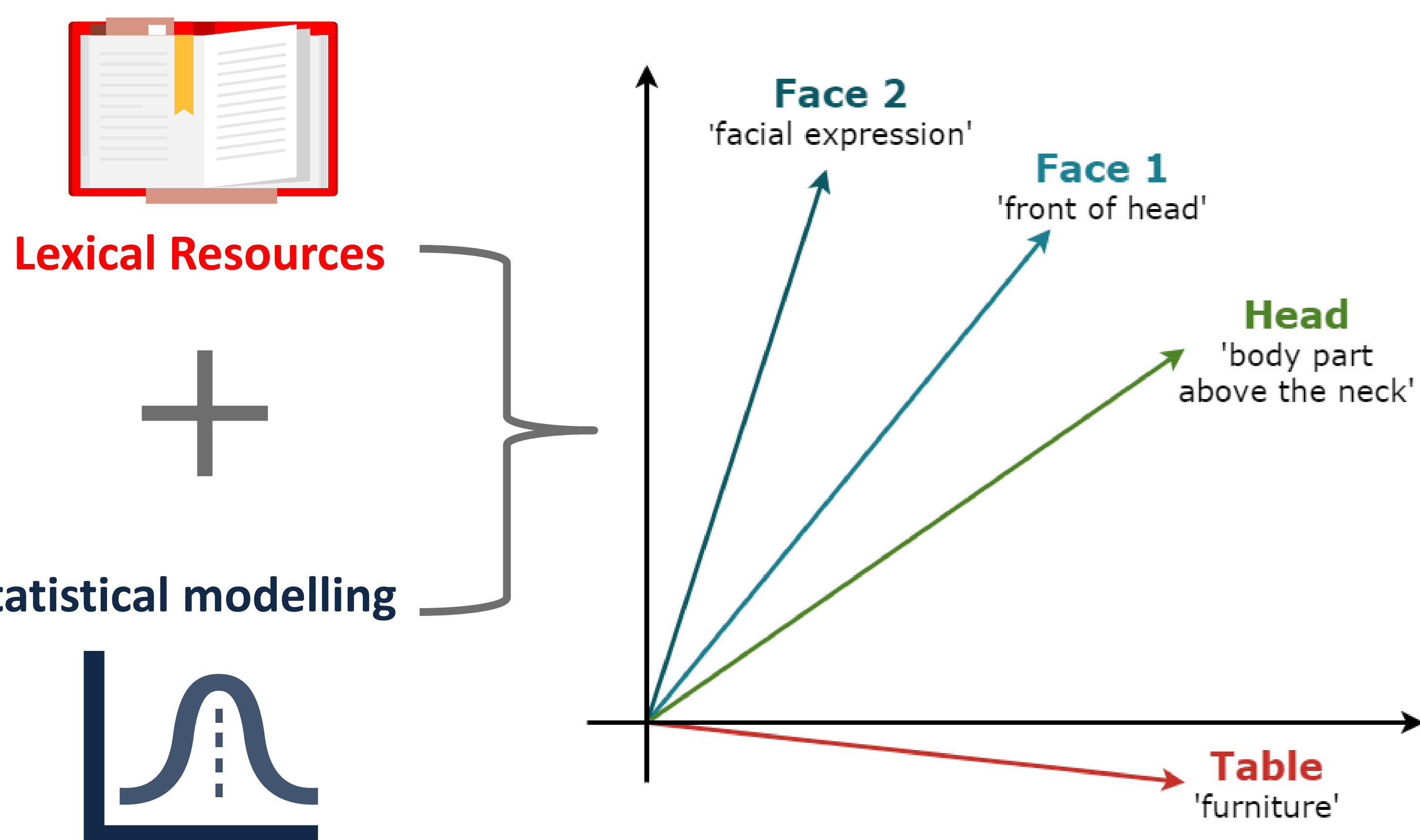
### Senses as vector

**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

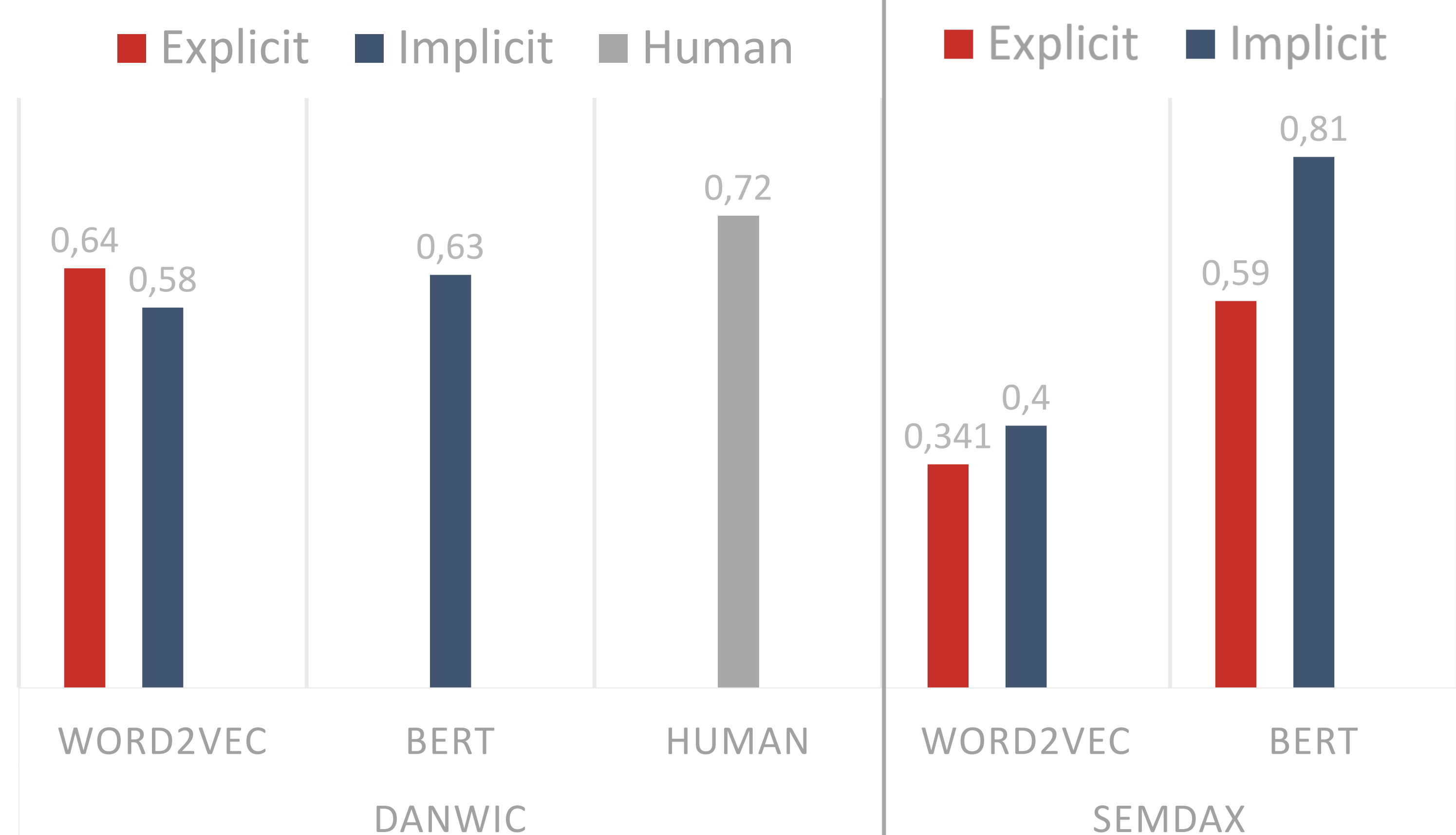
Combining

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

Task	Intrinsic <b>DanWiC</b>	Extrinsic <b>SemDaX lexical Sample</b>
	Word-in-Context	Word Sense Disambiguation
Collection	Automatically compiled from DanNet	Human annotated sense corpus
Coverage	<b>Broad</b> 1501 unique lemmas 3 different word classes	<b>Narrow</b> 18 highly polysemous nouns
Instances per lemma	<b>Low</b> Maximum 3	<b>High</b> 127-536



### Strong inclusion (explicit)

- Multiple knowledge sources
- DanNet synset connections
- Sense definition
- Quote with target lemma

Relies more on **The lexical resources**

### Weak inclusion (implicit)

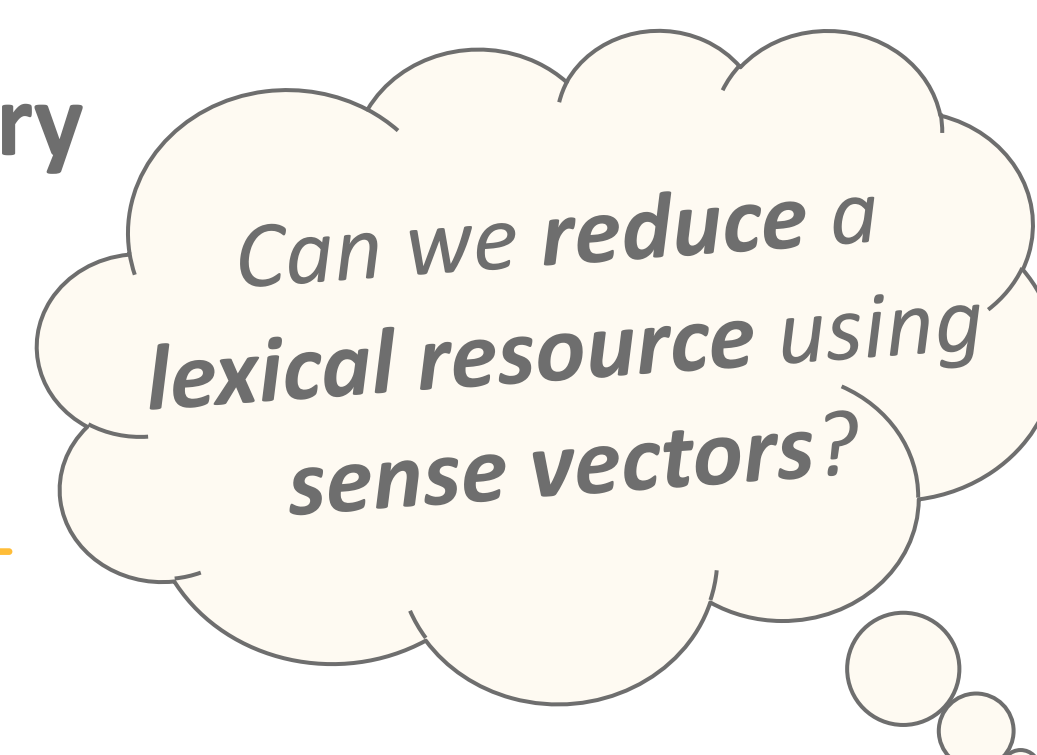
- A single knowledge source
- Just the definition or quote

Relies more on **The statistical modelling**

### Future Work – The COR-S project

- Create a **reduced Danish sense inventory** for improved and robust AI

- Clustering based on semantic proximity
- Pairwise comparison of sense vectors



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

