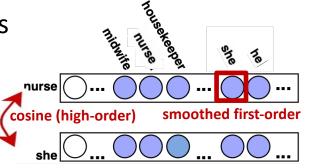# How should we measure biases & societal phenomena from text?
Like in what extent *nurse* is perceived as a *women job* when processing Wikipedia text

- Widely accepted approach to bias measurement is based on high-order relations (like cosine)

- We suggest smoothed first-order measurement
  - ➢ achieved by reconstructing the co-occurrence matrices from word2vec/GloVe word embeddings

- Smoothed first order bias measurement has several **advantages**:
  - ➢ Avoid the inclusion of unwanted factors as in high-order methods
  - ➢ Approaches the sparsity issue of traditional count-based methods

- Our experiments reveal the existence of stronger biases in occupations towards female when processing Wikipedia text

- A Take-Home Message: select a *proper bias measurement method* based on what is aimed to be quantified as *bias construct*