

## between word embeddings

## Optimal Transport

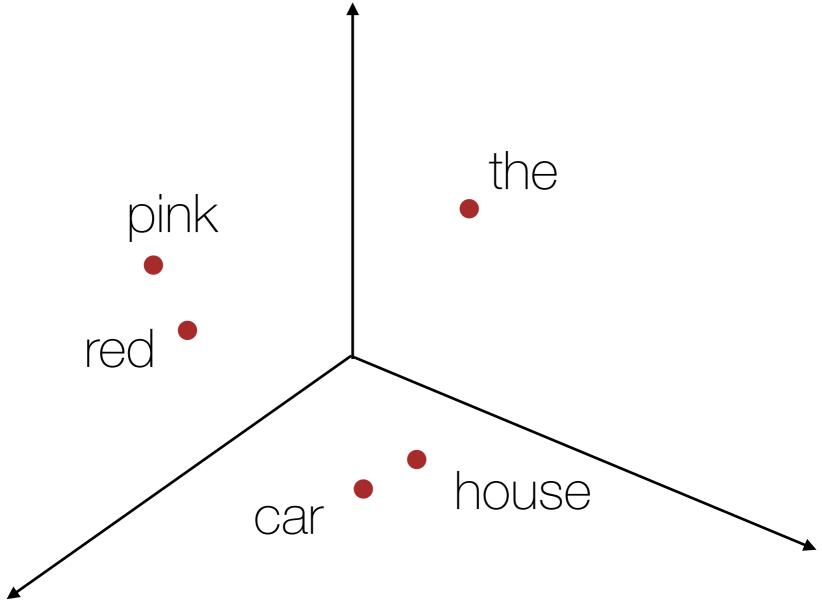
Treat embeddings as support points of discrete distribution

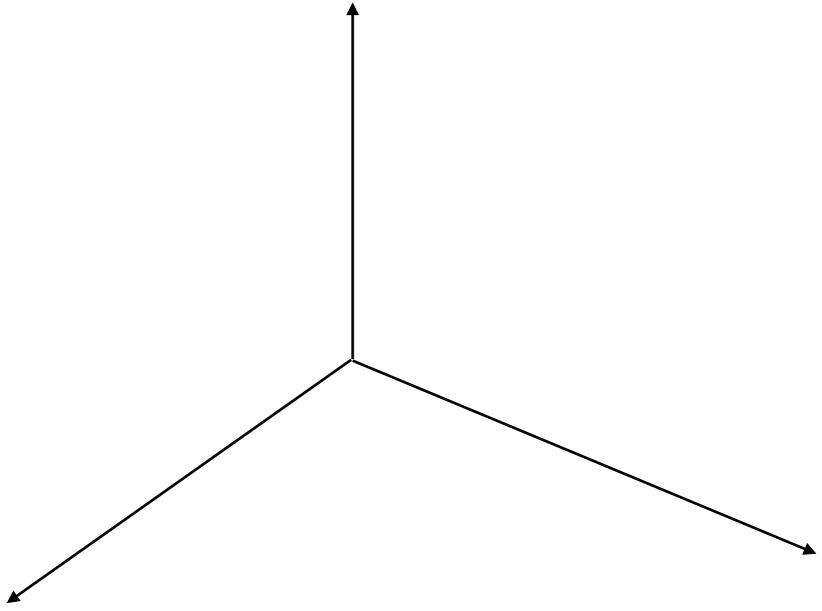




But this assumes the two spaces are registered (~axes are in correspondence)

Not true in general for word embeddings in different languages!











## [Kusner et al. 2015]

e.g. WMD [Kusner et al., 2015]

$$c(w_i^{EN}, w_j^{ES}) = d(\mathbf{v}^{EN}(w_i), \mathbf{v}^{ES}(w_j))$$



(identical under word

## embedding objective)



