Multi-head inner attention

- Liu et al. (2016), Li et al. (2016), Lin et al. (2017)
- α_{ij}: weight of the jth encoder state for the ith column of $M^\top$
- concatenate columns of $M^\top$
  - $\rightarrow$ sentence embedding
- linear projection of columns to control embedding size