

# Multi-way multi-source MT

Firat et al. 2016 (two papers)

- Still assumes only many bilingual parallel corpora.
- What to do if there are multiple input sentences?
- Early averaging (average context vectors).  $\mathbf{c}_t = \frac{\mathbf{c}_t^1 + \mathbf{c}_t^2}{2}$ .
- Late averaging (aka linear interpolation).

$$P(w_i|\mathbf{c}) = \sum_{k=1}^K \lambda_k(\mathbf{c}) P_k(w_i|\mathbf{c})$$

Early and late averaging are orthogonal, can be combined.