# Multi-way MT

Firat et al. 2016 (two papers)

- As in Bahdanu et al. (2014), attention mechanism is a feedforward function of both decoder hidden state and encoder context vector.

- **Shared** between all encoders and decoders.

$$p(f_i|f_{i-1}, ..., f_1, \mathbf{e}) = g(\boxed{f_{i-1}, s_i, c_i})$$

Everything we need is right here!

$$\alpha_{ij} = \frac{\exp(a_{ij})}{\sum_{k=1}^{|\mathbf{e}|} \exp(a_{ik})}$$

$$c_i = \sum_{j=1}^{|\mathbf{e}|} \alpha_{ij} h_j$$

$$a_{ij} = a(s_{i-1}, h_j)$$