

Multi-way MT

Firat et al. 2016 (two papers)

- Assume only many bilingual parallel corpora.
- For N languages: learn N encoders and N decoders.
- But what about attention?

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

$$c_i = \sum_{j=1}^{|\mathbf{e}|} \alpha_{ij} h_j$$
$$\alpha_{ij} = \frac{\exp(a_{ij})}{\sum_{k=1}^{|\mathbf{e}|} \exp(a_{ik})}$$
$$a_{ij} = a(s_{i-1}, h_j)$$