

What Went Wrong?

$$\hat{e}_1^I = \operatorname{argmax}_{I, e_1^I} p(f_1^J | e_1^I) p(e_1^I) = \operatorname{argmax}_{I, e_1^I} \prod_{(\hat{f}, \hat{e}) \in \text{phrase pairs of } f_1^J, e_1^I} p(\hat{f} | \hat{e}) p(e_1^I) \quad (12)$$

- Too strong phrase-independence assumption.
 - Phrases do depend on each other.
Here “nemám” and “žádného” jointly express one negation.
 - Word alignments ignored that dependence.
But adding it would increase data sparseness.
- Language model is a separate unit.
 - $p(e_1^I)$ models the target sentence independently of f_1^J .