

# Performance of STE and BPE

- German→Czech T2T experiments (Macháček et al., 2018).
- The underscore trick:
  - Append “\_” to tokens before learning splits.

split	underscore	BLEU
STE	after every token	18.58 $\pm$ 0.06
BPE	after non-final tokens	18.24 $\pm$ 0.08
BPE	after every token	13.88 $\pm$ 0.18
BPE	-	13.69 $\pm$ 0.66

- +5(!) BLEU points from the underscore trick.
  - If not attached at the end of the sentence.