## Full Results – correlations

BLEU vs. other metrics:  $-0.57 \pm 0.31$  (en  $\rightarrow$  cs)

**-0.36 ± 0.29** (en → de)

Pairwise average (except BLEU):  $0.78 \pm 0.32$  (en  $\rightarrow$  cs)  $0.57 \pm 0.23$  (en  $\rightarrow$  de)

