

# Evaluation

- How to compare two arbitrary translations?
- Low agreement rate even between reviewers
- BLEU score – a popular automatic technique

*Reference:*

*E-mail was sent on Tuesday.*

*System output:*

*The letter was sent on Tuesday.*

*1-grams: 4 / 6*

*2-grams: 3 / 5*

*3-grams: 2 / 4*

*4-grams: 1 / 3*