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.***