

Other evaluation metrics

- **BLEU:** Bilingual Evaluation Understudy
- **ROUGE:** (Recall-Oriented Understudy for Gisting Evaluation)
- **METEOR:** (Metric for Evaluation of Translation with Explicit ORdering)

Other MT evaluation metrics

- **BLEU (Bilingual Evaluation Understudy)**
 - A score for evaluating the performance of a MT.
 - BLEU score compares automatically translated sentences (hypothesis) with one or more reference translations.
- **ROUGE (Recall-Oriented Understudy for Gisting Evaluation)**
 - Usually used for evaluating the quality of text summaries.
 - ROUGE compares word sequences, word pairs, and n-grams with a set of reference summaries created by humans.
 - Many ROUGE variants such as
 - **ROUGE-N** (Overlap of N-grams between the system and reference summaries)
 - **ROUGE-L** (Longest Common Subsequence (LCS) based statistics)
 - **ROUGE-W** (Weighted LCS-based statistics that favors consecutive LCSes)
 - **ROUGE-S** (Skip-bigram based co-occurrence statistics)
 - **ROUGE-SU** (Skip-bigram plus unigram-based co-occurrence statistics).
- **METEOR (Metric for Evaluation of Translation with Explicit ORdering)**
 - METEOR or is another metric used to evaluate the performance of a MT model.
 - METEOR compares Standard word segments with the reference texts. Stems and synonyms are also considered for matching.