BLEU (Bilingual Evaluation Understudy)

Candidate (1) shares more words and word-N-grams with the reference translations than candidate (2)!
→ Compute **precision scores** (proportion of correct N-grams)

**Modified N-gram precision (for each N-gram):**

\[
\text{count}_{\text{clip}}(w) = \min(\text{count}_{\text{candidate}}(w), \max\text{count}_{\text{reference}}(w))
\]

*What is \(\text{count}_{\text{clip}}(w = \text{"the"})\) for candidate (1)?*  \(\min(3, 4) = 3\)

Do the same thing for larger units:
- **unigram precision** = proportion of correct words among all candidate words
- \(p_{\text{unigram}} = \frac{\text{count}_{\text{clip}}(w_1) + \text{count}_{\text{clip}}(w_2) + \ldots + \text{count}_{\text{clip}}(w_M)}{\text{count}(w_1) + \text{count}(w_2) + \ldots + \text{count}(w_M)}\)
- \(p_{\text{unigram}} = \frac{\sum_{i=1}^{M} \text{count}_{\text{clip}}(w_i)}{\sum_{i=1}^{M} \text{count}(w_i)}\)
- compute \(\text{count}_{\text{clip}}(w)\) for all translated sentences
- compute \(p_1 = p_{\text{unigram}}\) for all translated sentences together

---

**Note:**
- **candidate translations**
- **reference translations**

1. It is a guide to action which ensures that the military always obeys the commands of the party.
2. It is to insure the troops forever hearing the activity guidebook that party direct.
3. It is the practical guide for the army always to heed the directions of the party.

1. It is a guide to action that ensures that the military will forever heed Party commands.
2. It is the guiding principle which guarantees the military forces always being under the command of the Party.
3. It is the practical guide for the army always to heed the directions of the party.
BLEU (Bilingual Evaluation Understudy)

Put all precision scores together: Use geometric mean!

\[ P = \sqrt[n]{p_1 \cdot p_2 \cdot \ldots \cdot p_n} \]

\[ = (p_1 \cdot p_2 \cdot \ldots \cdot p_n)^{\frac{1}{n}} = \left( \prod_{i=1}^{n} p_i \right)^{\frac{1}{n}} \]

This can be transformed into: \( P = \exp \left( \frac{1}{n} \sum_{i=1}^{n} \log_e(p_i) \right) \)

\[ \exp(\log_e(x)) = x \]
\[ \log_e(x^y) = y \cdot \log_e(x) \]
\[ \log_e(x \cdot y) = \log_e(x) + \log_e(y) \]

Brevity penalty (BP) for short candidates (c):

\[ BP \begin{cases} 1 & \text{if candidate } c > \text{reference } r \\ \exp(1 - r/c) & \text{if candidate } c \leq \text{reference } r \end{cases} \]

Putting it all together:

\[ \text{BLEU} = BP \cdot \exp \left( \frac{1}{n} \sum_{n=1}^{N} \log_e(p_n) \right) \]

Common setting: \( N = 4 \), lowercased n-grams