CA4012 SMT Assignment 1
Eoin Murphy 11487358

MT Outputs:

1. Salmons swim in river .
2. Fish swim in the river .
3. The salmon swam in the river .

Reference:

- Salmons swim in the river .

N-grams

Translation 1: "Salmons swim in river ."

1-grams: Salmons, swim, in, river, .
2-grams: Salmons swim, swim in, in river, river .
3-grams: Salmons swim in, swim in river, in river .
4-grams: Salmons swim in river, swim in river .

Translation 2: "Fish swim in the river ."

1-grams: Fish, swim, in, the, river, .
2-grams: Fish swim, swim in, in the, the river, river .
3-grams: Fish swim in, swim in the, in the river, the river .
4-grams: Fish swim in the, swim in the river, in the river .

Translation 3: "The salmon swam in the river ."

1-grams: The, salmon, swam, in, the, river, .
2-grams: The salmon, salmon swam, swam in, in the, the river, river .
3-grams: The salmon swam, salmon swam in, swam in the, in the river, the river .
4-grams: The salmon swam in, salmon swam in the, swam in the river, in the river .

Reference: "Salmons swim in the river ."

1-grams: Salmons, swim, in, the, river, .
2-grams: Salmons swim, swim in, in the, the river, river .
3-grams: Salmons swim in, swim in the, in the river, the river .
4-grams: Salmons swim in the, swim in the river, in the river .

BLEU

$$BLEU = min\left(1, \frac{output\ length}{reference\ length}\right)\left(\prod_{i=1}^{4} percision_i\right)^{\frac{1}{4}}$$

$$Precision_n = \frac{number\ of\ clipped\ ngram\ in\ output}{total\ number\ of\ ngram\ in\ output}$$

Precision

Precision is a metric used in BLEU to determine how accurate the translation is against a reference source. It is measured from 1-n grams typically a range of 1-4 is used. Precision does have its issues as if a n-gram has a precision value of zero then the entire BLEU score will be zero as it is a multiplicative function.

Translation 1: "Salmons swim in river ."

$Precision_1 = 5/5 = 1.0$
$Precision_2 = 3/4 = 0.75$
$Precision_3 = 1/3 = 0.33$
$Precision_4 = 0/2 = 0$

Translation 2: "Fish swim in the river ."
$Precision_1 = 5/6 = 0.83$
$Precision_2 = 4/5 = 0.8$
$Precision_3 = 3/4 = 0.75$
$Precision_4 = 2/3 = 0.66$

Translation 3: "The salmon swam in the river ."
$Precision_1 = 4/7 = 0.57$
$Precision_2 = 3/6 = 0.5$
$Precision_3 = 2/5 = 0.4$
$Precision_4 = 1/4 = 0.25$

Brevity Penalty

$$Brevity = min\left(1, \frac{output\ length}{reference\ length}\right)$$

The closest reference length is used for the Brevity penalty. If the closest reference is greater then 1 we just set it to 1. In this example we only have one reference so our brevity penalty will always be the MT output we are currently checking's length over the one reference length. If we had multiple references we would need to check the output against each reference and chose the reference which has the closest match as out reference length.

Translation 1: "Salmons swim in river ."

$Brevity = 5/6$

Translation 2: "Fish swim in the river ."

Brevity $= 6/6$

Translation 3: "The salmon swam in the river ."

Brevity $= 7/6 := 1$

BLEU Scores

Translation 1:

$$BLEU = \frac{5}{6} \times (1.0 \times 0.75 \times 0.33 \times 0)^{\frac{1}{4}} = 0$$

* will always equal zero if a Precision value of zero is present

Translation 2:

$$BLEU = 1 \times (0.83 \times 0.8 \times 0.75 \times 0.66)^{\frac{1}{4}}$$
$$= 1 \times (0.32868)^{\frac{1}{4}}$$
$$= 0.7571$$

Translation 3:

$$BLEU = 1 \times (0.57 \times 0.5 \times 0.4 \times 0.25)^{\frac{1}{4}}$$
$$= 1 \times (0.0285)^{\frac{1}{4}}$$
$$= 0.4108$$