

# CSC401 Homework Assignment #2

## Analysis

Daniyal Shahzad

Student number: 1010137629

UTORid: shahz166

d.shahzad@mail.utoronto.ca

## 1 Training Results

### 1.1 Training Loop Printout

#### Model with Pre-layer Normalization

```
[Device:cuda] Epoch 1 Training ====
[Device:cuda] Epoch 1 Validation ====
Epoch 1: loss=6.35202525816223, BLEU: skipped until epoch 4, time=00:01:41
[Device:cuda] Epoch 2 Training ====
[Device:cuda] Epoch 2 Validation ====
Epoch 2: loss=2.8117863269842807, BLEU: skipped until epoch 4, time=00:03:22
[Device:cuda] Epoch 3 Training ====
[Device:cuda] Epoch 3 Validation ====
Epoch 3: loss=1.9384291363662973, BLEU: skipped until epoch 4, time=00:05:04
[Device:cuda] Epoch 4 Training ====
[Device:cuda] Epoch 4 Validation ====
Epoch 4: loss=1.5827857448120153, BLEU-4: 35.2011 BLEU-3: 42.1095, time=00:07:21
[Device:cuda] Epoch 5 Training ====
[Device:cuda] Epoch 5 Validation ====
Epoch 5: loss=1.343592219586638, BLEU-4: 36.1027 BLEU-3: 42.9181, time=00:09:37
```

#### Model with Post-layer Normalization

```
[Device:cuda] Epoch 1 Training ====
[Device:cuda] Epoch 1 Validation ====
Epoch 1: loss=6.358696772611839, BLEU: skipped until epoch 4, time=00:02:01
[Device:cuda] Epoch 2 Training ====
[Device:cuda] Epoch 2 Validation ====
Epoch 2: loss=2.799971314137677, BLEU: skipped until epoch 4, time=00:04:03
[Device:cuda] Epoch 3 Training ====
[Device:cuda] Epoch 3 Validation ====
Epoch 3: loss=1.9602262958410641, BLEU: skipped until epoch 4, time=00:06:06
[Device:cuda] Epoch 4 Training ====
[Device:cuda] Epoch 4 Validation ====
Epoch 4: loss=1.6276981408340159, BLEU-4: 35.2809 BLEU-3: 42.2410, time=00:08:42
[Device:cuda] Epoch 5 Training ====
[Device:cuda] Epoch 5 Validation ====
Epoch 5: loss=1.4009998585429955, BLEU-4: 35.8643 BLEU-3: 42.6337, time=00:11:17
```

### 1.2 Test Set BLEU Score

This section lists the test set BLEU score reported on the test set for each model in table 1.

Model	BLEU-4	BLEU-3
Model Pre-layer Normalization	41.8109	49.0464
Model Post-layer Normalization	41.3010	48.3197

Table 1: The BLEU score reported on the test set for each model.

## 2 Translation Analysis

### 2.1 Translations

- Voila des mesures qui favorisent la famille canadienne.  
My Model: '<s> these are things that the canadian family will benefit </s>'  
Bart MT: These are measures that favour the Canadian family.  
Google Translate: These are measures that favor the Canadian family.
- Je voudrais aussi signaler aux deutes qu'ils peuvent maintenant s'avancer pour voter.  
My Model: '<s> i would also like to remind members of the house to vote now </s>'  
Bart MT: I also want to tell members that they can now vote.  
Google Translate: I would also like to point out to members that they can now come forward to vote.
- Trudeau embauche un cabinet de consultants pour examiner la dependance excessive du gouverne-  
ment a l'egard des cabinets de consultants.  
My Model: '<s> trudeau s executive officer of consultants time to examine excessive dependency of  
funds </s>'  
Bart MT: Trudeau is creating a consulting cabinet to look at the excessive dependence of the  
government on.  
Google Translate: Trudeau hires consulting firm to examine government's overreliance on consulting  
firms.
- Pierre demande s'il vous plait s'il peut s'attribuer le merite d'avoir supprime 600 emplois a la SRC.  
My Model: '<s> pierre s request was abolished by the cbc </s>'  
Bart MT: Pierre asks if he can assist you in eliminating 600 jobs at the RCMP.  
Google Translate: Pierre please asks if he can take credit for cutting 600 jobs at the SRC.
- La France a remporte la coupe du monde 2018.  
My Model: '<s> france has been cut by the world </s>'  
Bart MT: France won the world in 2018.  
Google Translate: France won the 2018 World Cup.
- J'avais a peine de l'eau a boire pour huit jours.  
My Model: '<s> <unk> barely <unk> for eight days </s>'  
Bart MT: I had water to drink for eight days.  
Google Translate: I barely had water to drink for eight days.
- Toronto est une ville du Canada.  
My Model: '<s> toronto is a city of canada </s>'  
Bart MT: Toronto is a city of Canada.  
Google Translate: Toronto is a city in Canada.
- Les etudiants de l'Universite de Toronto sont excellents.  
My Model: '<s> the students of toronto are very good news </s>'  
Bart MT: The students at the University of Toronto are excellent.  
Google Translate: The students at the University of Toronto are excellent.

## 2.2 Discussion

*In this section, write a brief discussion on your findings. Describe the quality of those sentences. How's your model compared with Google Translate or ChatGPT?*

After getting the translations from all the models and comparing them. I find Google Translate consistently delivers translations with very good accuracy, fluency, and contextual understanding, making it the best among the three. Bart MT follows closely behind, giving translations that generally capture nuances and context accurately, although with occasional hiccups. In contrast, my model demonstrates potential but falls short in maintaining coherence and accuracy across translations, often producing less polished outputs or failing to account for the context. So the ranking is Google translate, Bart MT, and then my model based on accuracy of translations, considering the context, and producing intelligible and fluent outputs.

*What attributes and factors of the models do you think play a role in determining the quality?*

The first factor to consider would be the size of training data. Having a larger training data most of the time guarantees that our model would have more observations to learn from, and hence be more accurate in translations. Architecture also plays a role in determining the quality of models. For example, Transformer based models generally perform better than traditional RNN based models because they can capture context better through attention mechanism. Fine-tuning pretrained models can also affect the quality of predictions since the model will be able learn the domain specific information, or would be able to generate predictions that are more relevant for the task. Lastly, Post processing techniques, like greedy search or beam-search can affect how the translations are generated and their quality.

*Now, let's observe the quality of your model's translations for individual sentences. Which sentences does your model translate better, and which does it translate worse? Can you identify a pattern? Describe the pattern of quality across different types of sentences.*

"Voila des mesures qui favorisent la famille canadienne." My model translates this sentence relatively well, although it misses the time aspect of the model since it uses "will benefit", but captures the essence of the sentence.

"Toronto est une ville du Canada." My model translates this sentence accurately and fluently. It captures the simple statement about Toronto being a city in Canada effectively.

"Les etudiants de l'Universite de Toronto sont excellents." My model translates this sentence reasonably well, although it adds "news" at the end, which doesn't fit contextually since we are talking about university of Toronto. Still, it conveys the main message about the excellence of students.

"Pierre demande s'il vous plait s'il peut s'attribuer le merite d'avoir supprime 600 emplois a la SRC." My model's translation of this sentence is quite far off from the original meaning. It simply states "pierre s request was abolished by the cbc," which doesn't make much sense in context.

"Trudeau embauche un cabinet de consultants pour examiner la dependance excessive du gouvernement a l'egard des cabinets de consultants." My model's translation of this sentence lacks coherence and accuracy. It doesn't capture the essence of Trudeau hiring a consulting firm to examine the government's dependency on consulting firms.

"J'avais a peine de l'eau a boire pour huit jours." My model's translation of this sentence is incomplete and lacks coherence. It fails to convey the meaning of the original sentence about having barely enough water to drink for eight days and it gives out "<unk>" tokens.

My model seems to perform better with simple sentences that convey simple statements. It struggles more with sentences containing complex structures, and nuanced meanings such as those involving political context or nuanced requests like Pierre's statement. There is a pattern in terms of sentence length and complexity as well, where shorter and simpler sentences tend to be translated more accurately compared to longer and more complex sentences.

134 *What about the fine-tuned pre-trained model and Google Translate/ChatGPT's quality for individual*  
135 *sentences? Does the previous pattern still persist? Why?*  
136

137 Both Bart MT and Google Translate provide accurate and fluent translations across the provided  
138 sentences, maintaining consistency in quality. But when the sentence contains nuanced meanings and  
139 complex structures, it also tends to perform worse, but not as frequently as my model. This can be seen  
140 by the translation of "La France a remporte la coupe du monde 2018.", where it translates "France won  
141 the world in 2018." Also, the translation for "Pierre demande s'il vous plait s'il peut s'attribuer le merite  
142 d'avoir supprime 600 emplois a la SRC." is also not correct since it produced "Pierre asks if he can assist  
143 you in eliminating 600 jobs at the RCMP." The pattern of performing worse in complex structure and  
144 nuanced meanings holds here aswell. Lastly, pattern of high performance on shorter and simpler setence  
145 also holds since we can see Bart MT performing accurately on examples like, "Toronto est une ville du  
146 Canada".