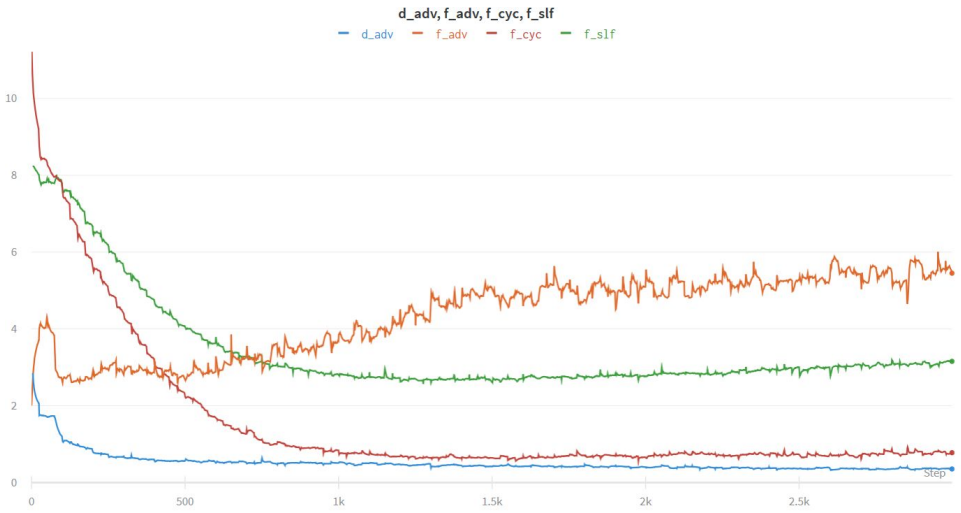
# HW5-1 Report

Student ID: R08921040, R08942087
github ID: wubinary

- Configuration

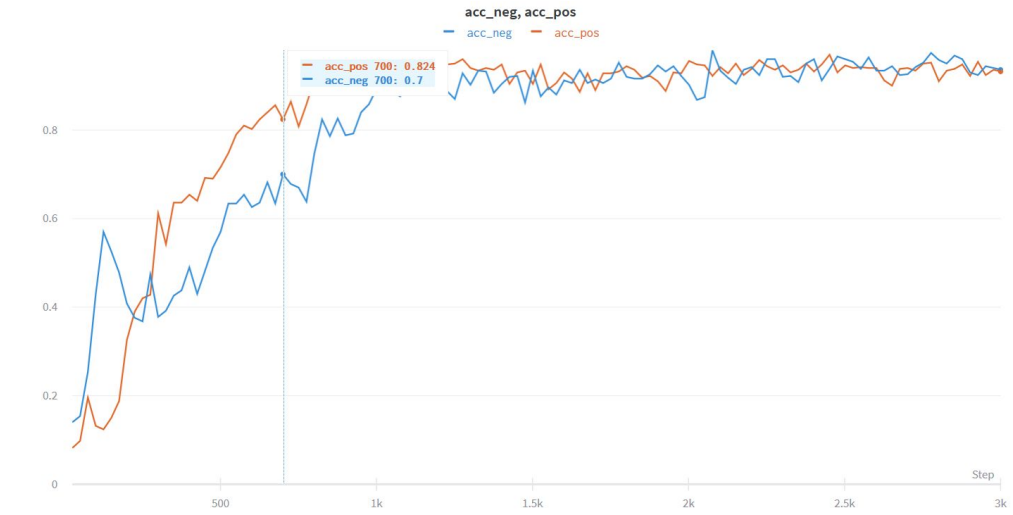
  We change batch size to 128 and set slf_factor as 0.5. Other arguments follow the default settings.

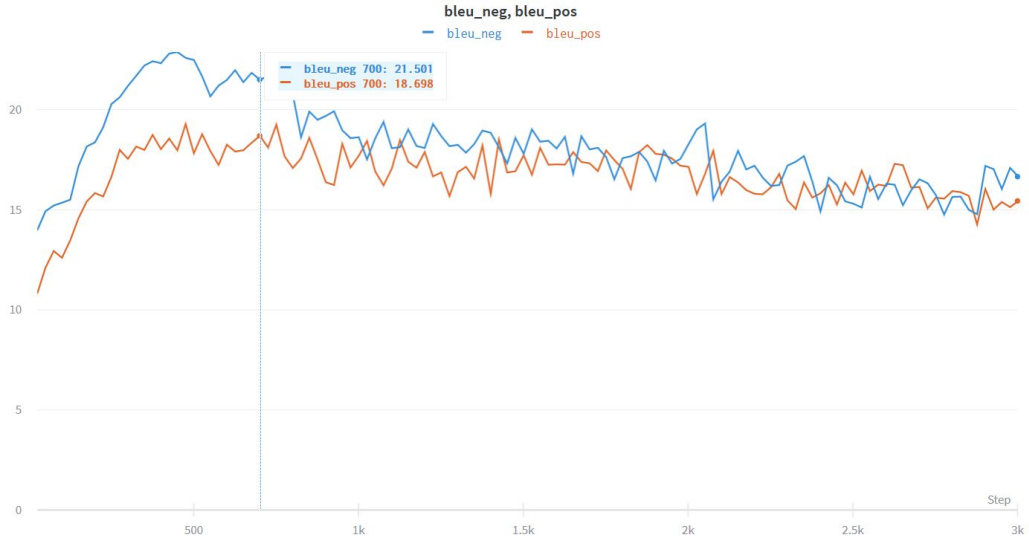
- Loss

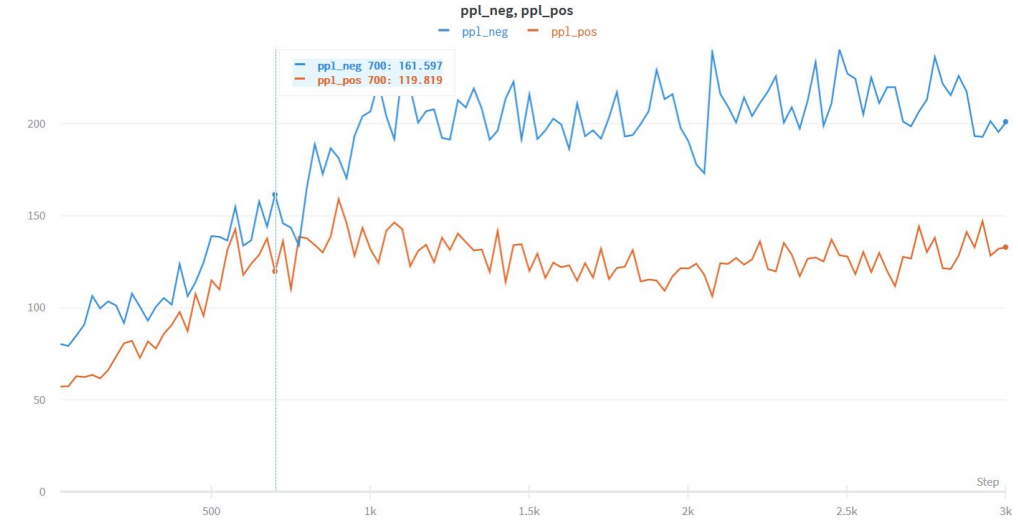



- Metrics

| | pos | neg | avg |
|---|---|---|---|
| accuracy | 0.7 | 0.824 | 0.762 |
| ref-bleu | 18.698 | 21.501 | 20.0995 |
| perplexity | 119.819 | 161.597 | 140.708 |


acc_neg, acc_pos
acc_neg
acc_pos
acc_pos 700: 0.824
acc_neg 700: 0.7
0.8
0.6
0.4
0.2
0
Step
500
1k
1.5k
2k
2.5k
3k


bleu_neg, bleu_pos
bleu_neg
bleu_pos
bleu_neg 700: 21.501
bleu_pos 700: 18.698
20
15
10
5
0
Step
500
1k
1.5k
2k
2.5k
3k


ppl_neg, ppl_pos
ppl_neg
ppl_pos
ppl_neg 700: 161.597
ppl_pos 700: 119.819
200
150
100
50
0
Step
500
1k
1.5k
2k
2.5k
3k

- Examples
  - Good Examples

```
****************** neg sample ******************
[gold] i ca n't believe how inconsiderate this pharmacy is .
[raw ] i ca n't believe how inconsiderate this pharmacy is .
[rev ] i ca always believe how inconsiderate this pharmacy is .
```

```
****************** neg sample ******************
[gold] she was not happy being there .
[raw ] she was not happy being there .
[rev ] she was definitely happy being there .
```

```
****************** neg sample ******************
[gold] the sales people here are terrible .
[raw ] the sales people here are terrible .
[rev ] the sales people here are great .
```

```
****************** pos sample ******************
[gold] great service by the staff .
[raw ] great service by the staff .
[rev ] terrible service by the staff .
```

```
****************** pos sample ******************
[gold] ambiance and setting were superb .
[raw ] ambiance and setting were superb .
[rev ] ambiance and setting were depressing .
```

From the examples above, we find that our model changes negative words to positive ones, and vice versa. For example, “not” becomes “always” and “great” turns into “terrible”.

  - Problematic Examples

```
****************** neg sample ******************
[gold] had to wait a month to get in .
[raw ] had to wait a month to get in .
[rev ] had to personable a month to get in .
```

```
****************** neg sample ******************
[gold] the salads are ok , nothing special .
[raw ] the salads are ok , nothing special .
[rev ] the salads are disappoint , enormous special .
```

```
******************** pos sample ********************
[gold] good food , great prices on wings on tuesdays .
[raw ] good food , great prices on wings on tuesdays .
[rev ] rude food , terrible prices on wings on tuesdays .
```

```
****************** pos sample ******************
[gold] nice menu , good prices , great service - for both dinner and breakfast !
[raw ] nice menu , good prices , great service - for both dinner and breakfast !
[rev ] disappointing menu , rude prices , terrible service - for then dinner and breakfast !
```

To transfer the style of a sentence, replacing a word with some specific word may not work. For example, replacing “wait” with “personable” in the first case dosen’t work. Actually, we can simply add “Don’t” at the beginning of the sentence. Aside from this, if the replacement word is an adjective, then the noun should be taken into consideration. In the third case, it is unsuitable to describe “the food” as “rude”.

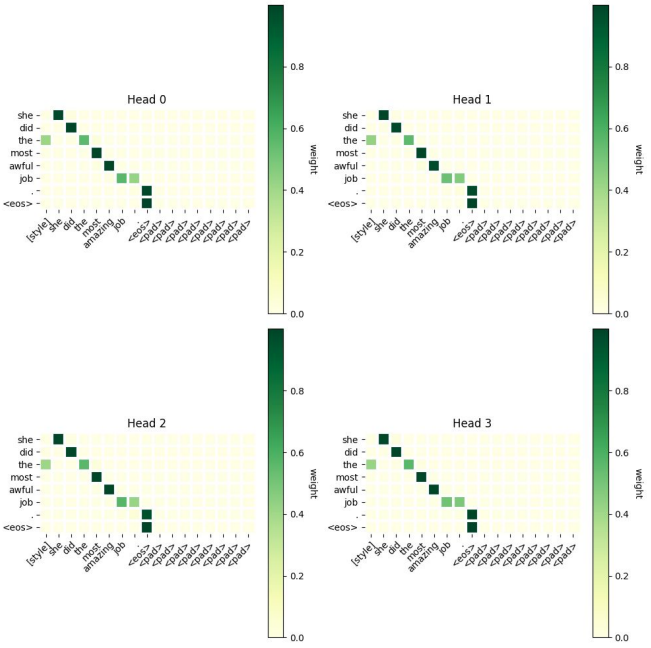
# HW5-2 Report

Student ID: R08921040, R08942087
github ID: wubinary

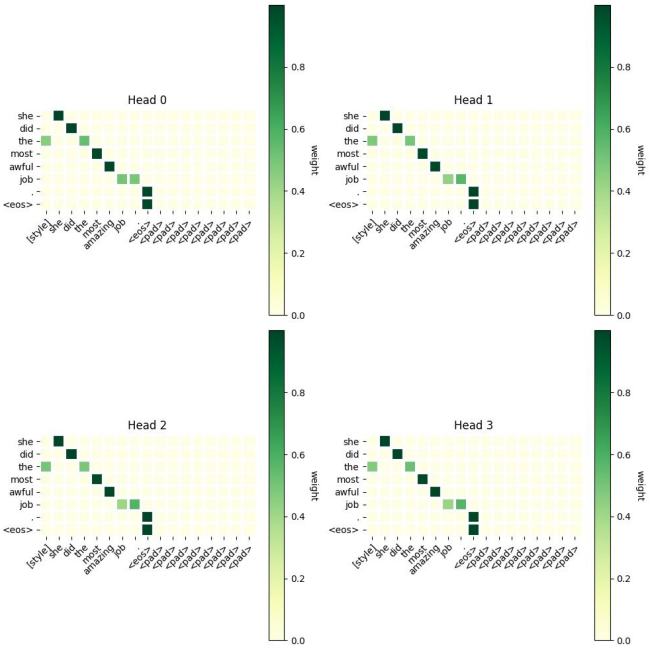
1. Problem 1
   a. sample one sentence and apply style tranfer
   b. show the attention maps and your explanation

```
****************** pos sample ******************
[gold] she did the most amazing job .
[rev ] she did the most awful job .
```

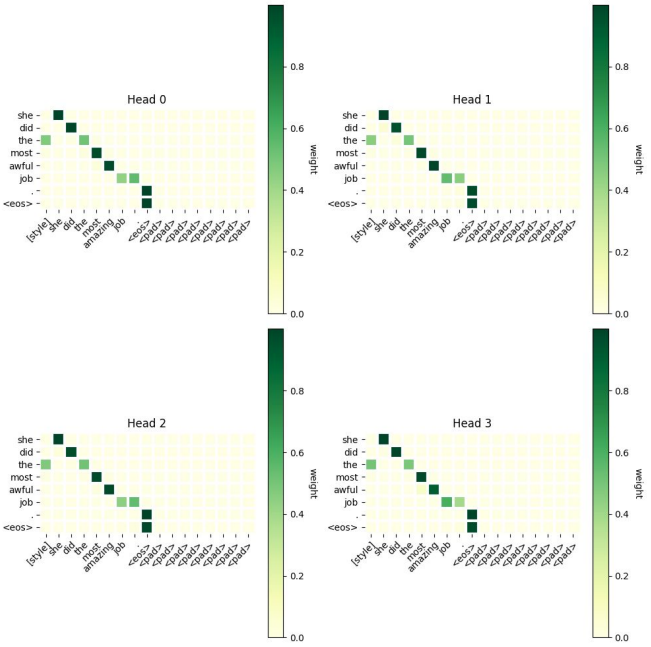
- Layer 0



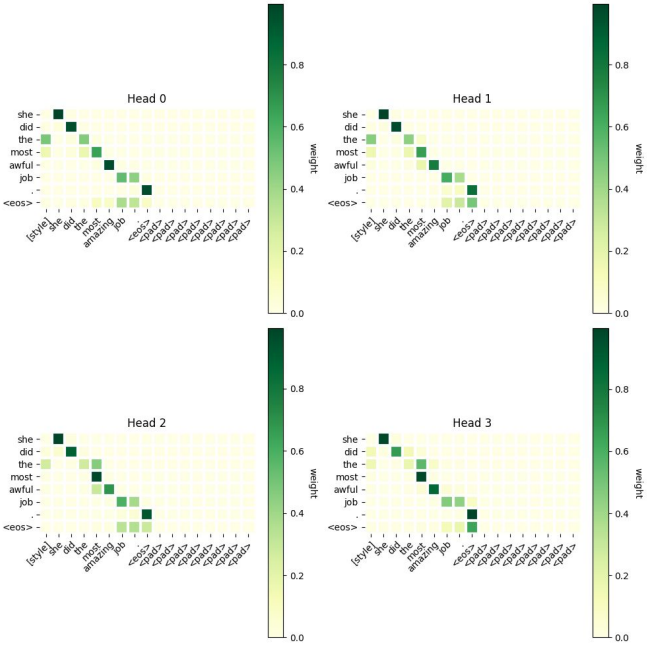

- Layer 1



- Layer 2



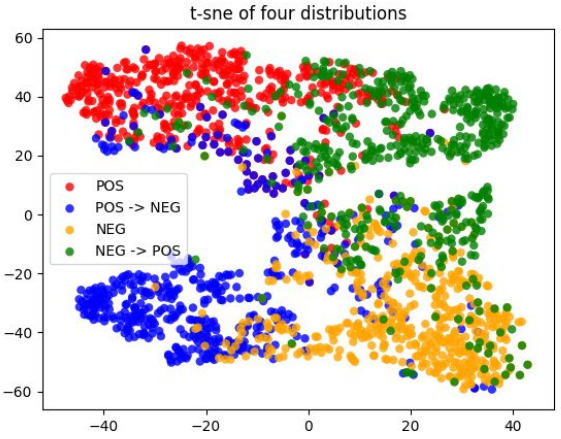

- Layer 3




c. The style token seems not being attended to while decoding, but the style is actually being transferred. Do you think it's reasonable? Why or why not?( Hint: how does the information of "style" flow? )

Since the discriminator learns to distinguish the style of sentences depending on the content of  the sentence, it's reasonable that the style token is not being attended to while decoding.

2. Problem 2
   a. apply style transfer on test data and collect the sentence embedding
   b. visualize the distribution of embedding by T-sne




   c. Does the result look reasonable?

      We observe that NEG → POS is closer to POS and POS → NEG is near to NEG. The result is reasonable since the transferred data should approach its target style.

3. Problem 3
   a. sample one sentence, mask one position at a time and do style transfer

```
***** 2-3: mask input *****
-
[ORG] the food here is delicious .
[REV] the food here is dry .
-
[ORG] <unk> food here is delicious .
[REV] the food here is dry .
-
[ORG] the <unk> here is delicious .
[REV] the food here is dry .
-
[ORG] the food <unk> is delicious .
[REV] the food was is dry .
-
[ORG] the food here <unk> delicious .
[REV] the food here was dry .
-
[ORG] the food here is <unk> .
[REV] the food here is not .
***** 2-3 end *****
```

b. show the results and your explanation
(Hint: Does masking N or ADJ make any difference? what kind of token affect the mostly on style, why?)

We observe that adjectives have the most influence on style. In this sentence, “delicious” is replaced with “dry”. If “delicious” is masked, the model may not be able to find a suitable word. In contrast, if a noun is masked, our model can fill in the blank with a noun. In fact, the results meet our expectation, where adjectives are usually key to the style of a sentence.

# HW5-3 Report

Student ID: R08921040, R08942087
github ID: wubinary

Option 2

- Configuration
  - Data: the provided gender data (Pos: male, Neg: female)
  - Settings:
    We use the sinusoidal positional embedding and change batch size to 128. As for model architecture, we set the number of head to 8. Other arguments follow the default settings.

```
5 train_gender:
6     python3 main.py --do_train -train_iter 600 -eval_steps 50 --train_gender -data_path ./data
  /gender_data/ -learned_pos_embed True -head 8 -save_path ./save/gender/ --use_wandb
7
```

- Loss


- Metrics

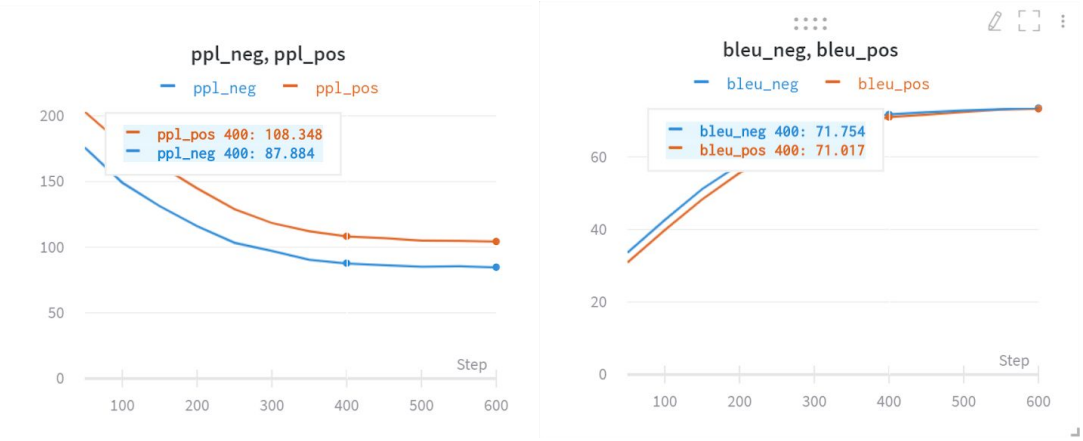
  (epoch 400)

| | male | female | avg |
|---|---|---|---|
| accuracy | 0.3487 | 0.1608 | 0.2547 |

| | | | |
|---|---|---|---|
| ref-bleu | 71.017 | 71.754 | 71.381 |
| perplexity | 108.348 | 87.884 | 98.116 |

(epoch 600)

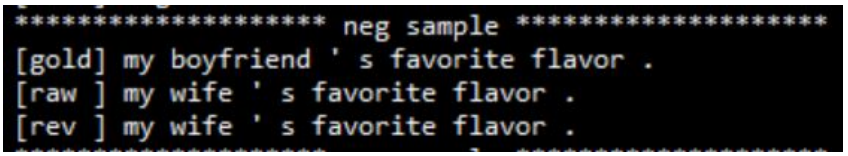
| | male | female | avg |
|---|---|---|---|
| accuracy | 0.355 | 0.179 | 0.267 |
| ref-bleu | 73.305 | 73.555 | 73.582 |
| perplexity | 104.397 | 84.801 | 94.599 |

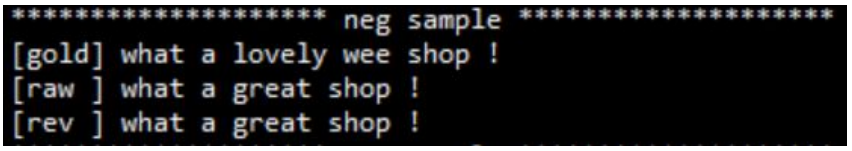


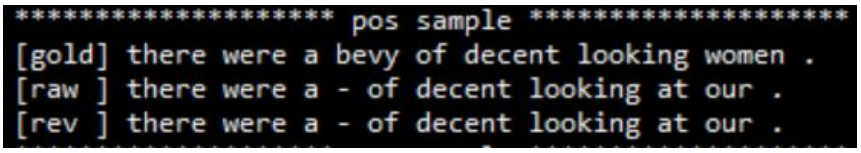

```
wandb: Waiting for W&B process to finish, PID 7640
wandb: Program ended successfully.
wandb: Run summary:
wandb:    _timestamp 1592849748.4206545
wandb:         f_cyc 1.3515025987625122
wandb:         f_slf 3.795697923660278
wandb:      _runtime 19828.757932424545
wandb:         f_adv 1.7499544792175292
wandb:         _step 600
wandb:         d_adv 2.736828130245209
wandb:       acc_pos 0.355469071586274
wandb:       ppl_pos 104.39770351279864
wandb:       ppl_neg 84.8017884425159
wandb:      bleu_pos 73.30505197372676
wandb:      bleu_neg 73.55533708576623
wandb:    total_bleu 146.860389059493
wandb:       acc_neg 0.17965048834337463
wandb: Syncing 5 W&B file(s), 0 media file(s), 0 art
wandb:
wandb: Synced desert-mountain-18: https://app.wandb.
q
aa@aa (~/Desktop/hw/DLHLP2020-SPRING/hw5/HW5-TextSty
```

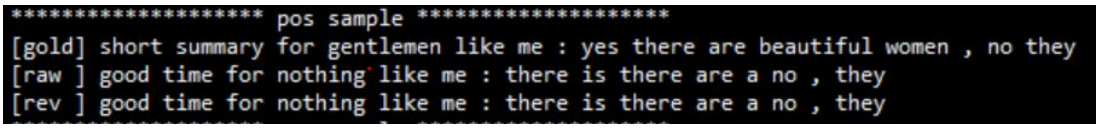
- Examples
  - Good Examples

```
********** neg sample **********
[gold] i had prosecco and my boyfriend ordered a beer .
[raw ] i had times and my wife ordered a beer .
[rev ] i had times and my wife ordered a beer .
```

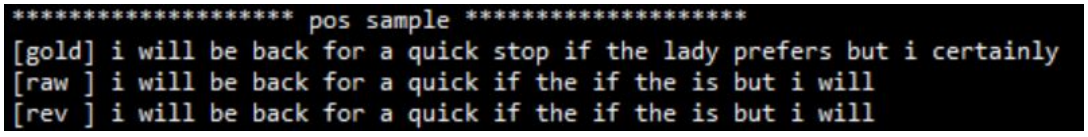


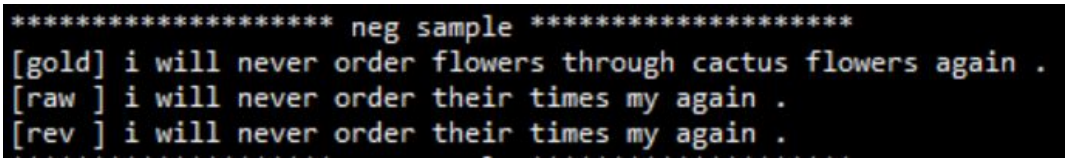



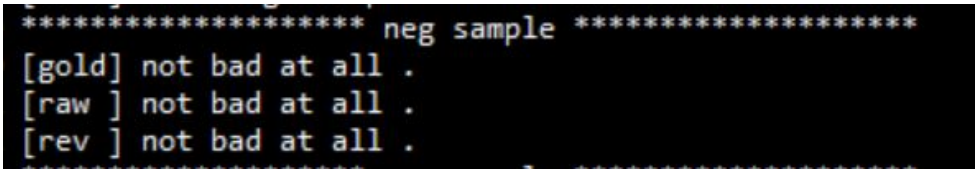



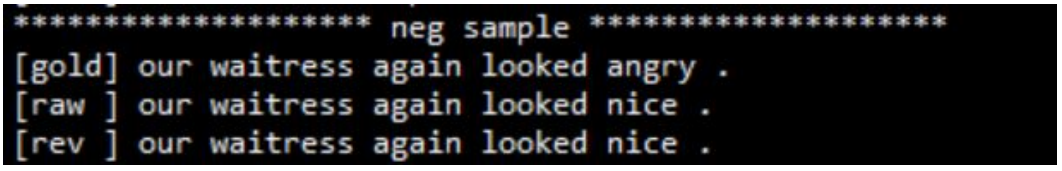



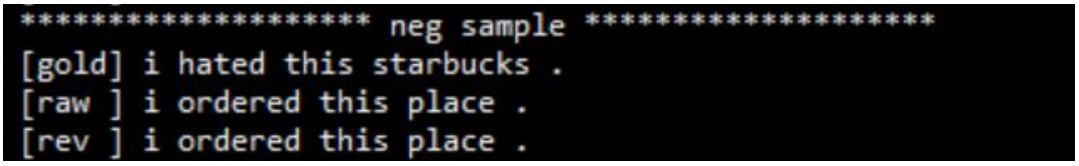

- Problematic Examples




Unfortunately, some examples don't change after transferring style. Also, the transferred sentences have grammar error in most cases. However, we still can observe that our model alters the word from masculine style to feminine style, and vice versa. For example, some words like "boyfriend", which usually comes

from women,  turn into “wife” and the word “women” is replaced when we set the style as feminine style.