

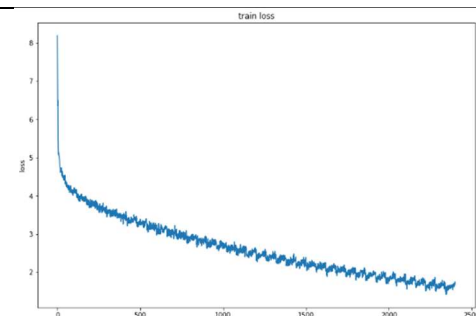
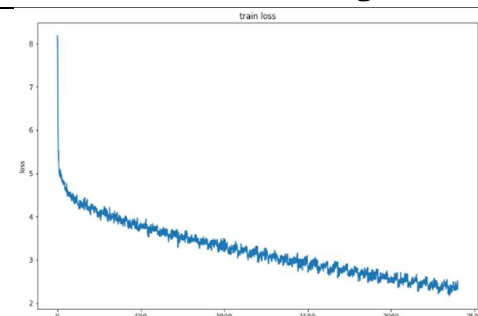
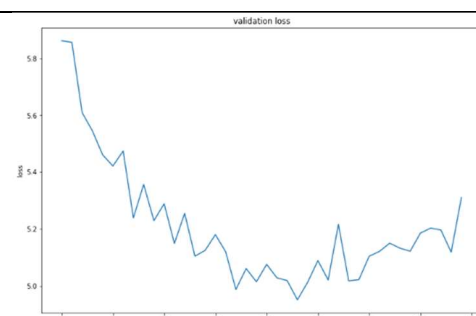
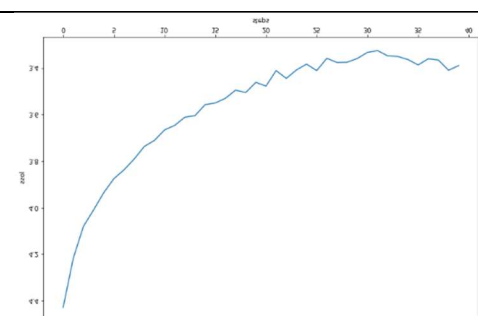
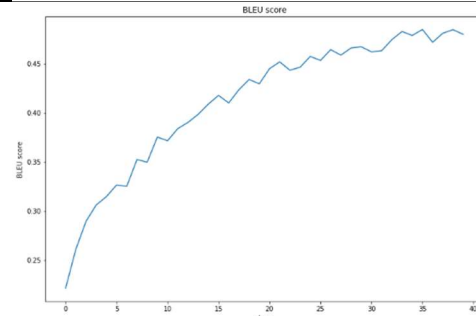
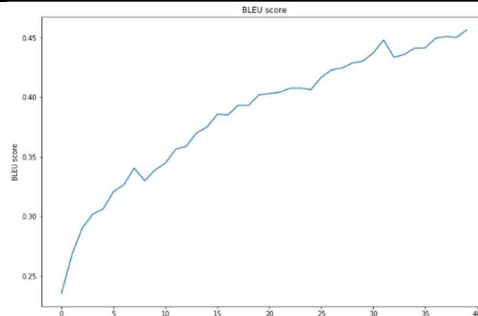
學號：R08922167 系級：資工碩一 姓名：曾民君

1. (20%) Teacher Forcing:

a. 請嘗試移除 Teacher Forcing，並分析結果。

Ans:

移除 Teacher Forcing 後，在 training loss 部分收斂比較慢且值比較大，valid loss 部分在去掉Teacher Forcing 後持續呈現不減反增的現象，在 valid 的 blue score 卻是收斂到差不多的值。在 Test 時會更爛，另外倆者在 blue 的表現都很糟。

	Baseline	No Teacher Forcing
Train Loss		
Valid Loss		
Valid BLUE Score		
Test Loss:	8.247097495474113	8.637846298444311
Test BLUE Score	0.0015560877977023727	0.0010741075476523379

2. (30%) Attention Mechanism:

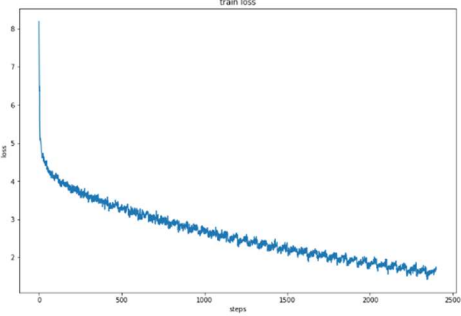
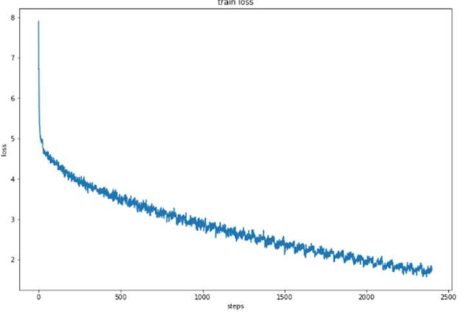
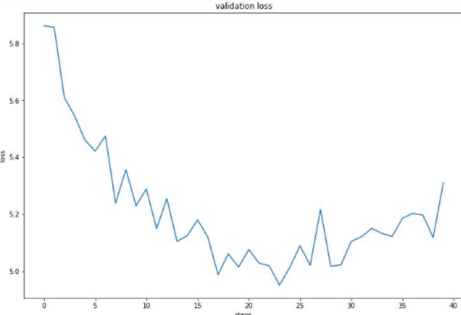
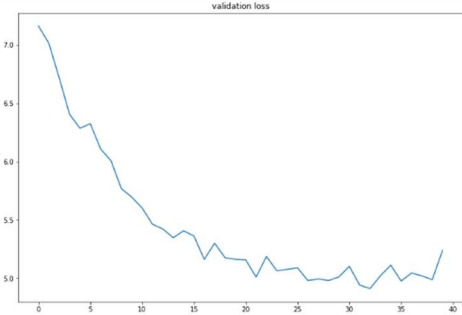
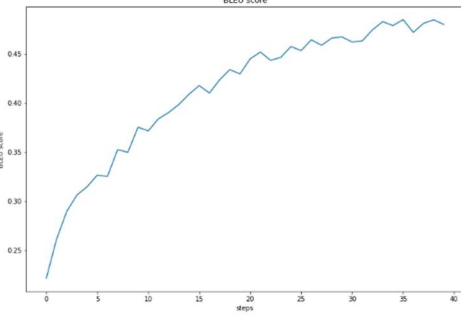
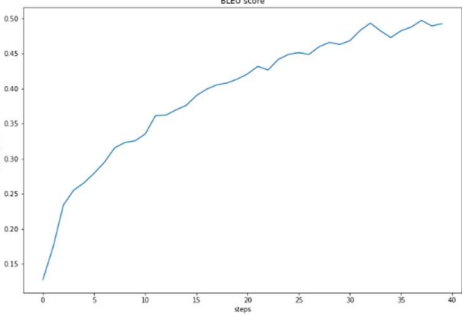
a. 請詳細說明實做 attention mechanism 的計算方式，並分析結果。

Ans:

實作步驟:

- 將 encoder 的 output 乘上 decoder 的 hidden_states
- 將上述乘積經過 softmax 運算得到一個 vector
- 將上方得到的 vector 與 encoder 的 output 做乘積
- 之後將上述依照對應的 decoder 的 hidden_state n 所計算出的結果，與 decoder 每一層的 input 相加再餵進 decoder layer 中

實際表現似乎不佳，在 test 的 blue score 所得到的分數反而比 base line 還要差，可能是因為直接相加會破壞掉一些原有的資訊。

	Baseline	With Attention
Train Loss		
Valid Loss		
Valid BLUE Score		
Test Loss:	8.247097495474113	8.233547497591589
Test BLUE Score	0.0015560877977023727	0.0012231167171067507

3. (30%) Beam Search:

a. 請詳細說明實做 beam search 的方法及參數設定，並分析結果。

Ans:

實作步驟:

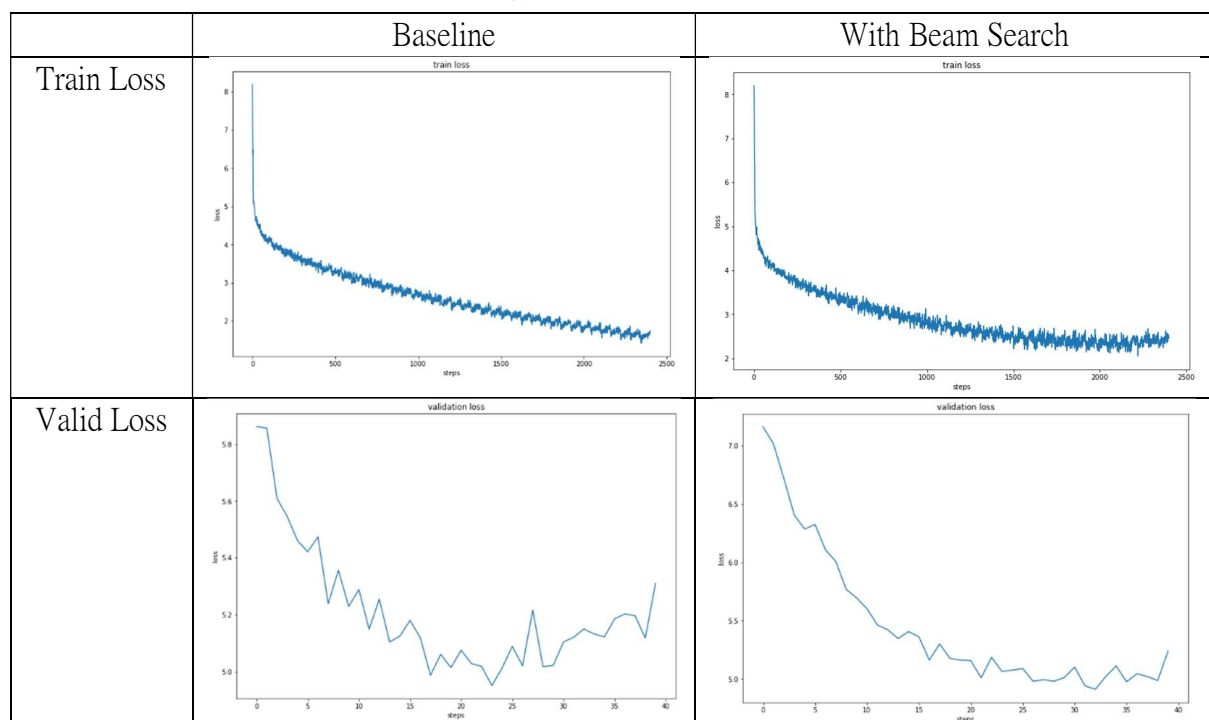
- 用一個candidates list 紀錄可能路徑的資訊，內容有 input/ hidden/ seq/ outputs/ score，其中 input與 hidden 為上一個output 結果，seq與 outputs為紀錄歷史 output 結果，score 為記錄某路徑的分數(越大越好)，以 $\log(\text{softmax}(\text{output}))$
- 在每次探索新的可能時，指挑出所有 candidates 前 n_beams (hyper parameters)，則會挑出 $n_beams * n_beams$ 個可能路徑，在依照 score 排序，只保留前 n_beams 個可能路徑於 candidates，已用於為來探索依據。
- 最後只回傳 score 最好的路徑

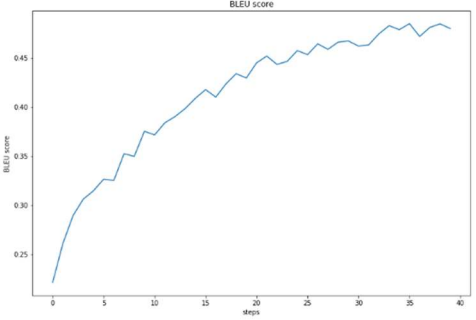
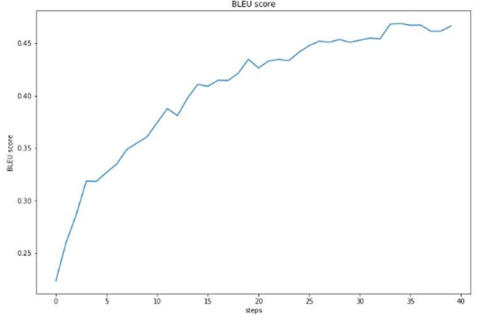
實際測驗結果所生成的文字有容易連續生成同一字元的傾向如:

```
['<BOS>', 'did', 'he', 'go', 'to', 'see', 'mary', '?']
```

[‘他’, ‘他’, ‘看見’, ‘去’, ‘嗎’, ‘嗎’, ‘嗎’, ‘嗎’, ‘嗎’, ‘?’, ‘?’, ‘?’, ‘?’, ‘?’, ‘?’, ‘?’, ‘?’,
‘。’, ‘對’, ‘的’, ‘?’, ‘。’, ‘?’, ‘。’, ‘?’, ‘。’, ‘。’, ‘?’, ‘。’, ‘。’, ‘。’, ‘。’, ‘。’,
‘。’, ‘。’, ‘。’, ‘。’, ‘。’, ‘。’, ‘。’, ‘。’, ‘。’, ‘。’, ‘。’, ‘。’],

這一部分還不確定問題出在哪裡。



Valid BLEU Score		
Test Loss:	8.247097495474113	8.31197484357742
Test BLEU Score	0.0015560877977023727	0.0012534374107020675

4. (20%) Schedule Sampling:

a. 請至少實做 3 種 schedule sampling 的函數，並分析結果。

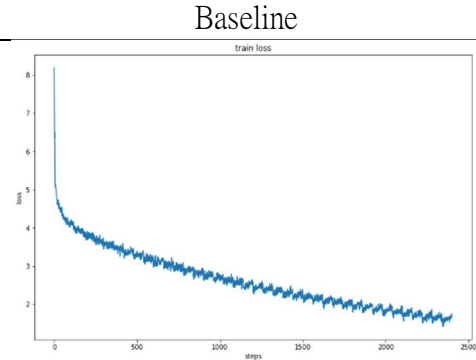
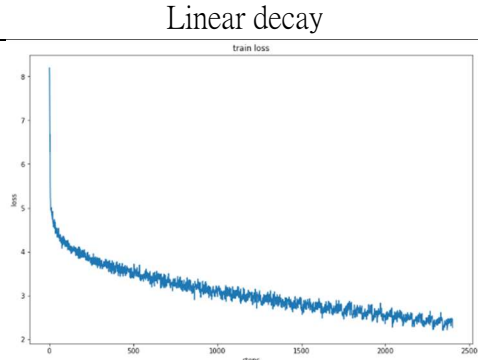
Ans:

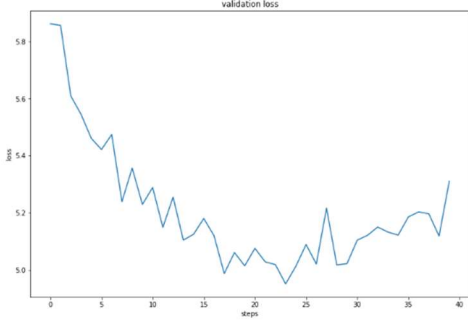
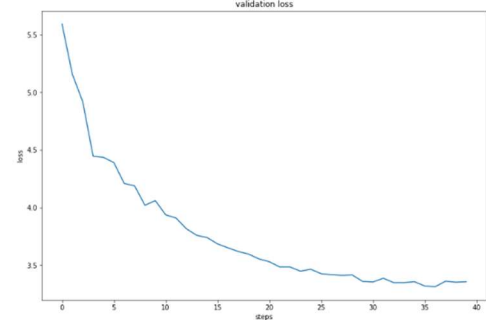
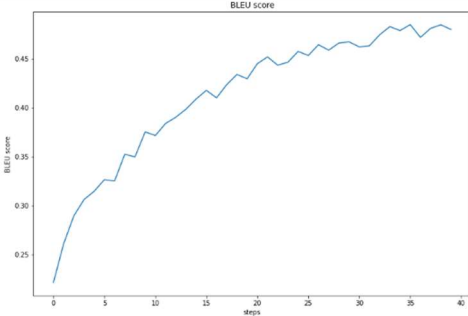
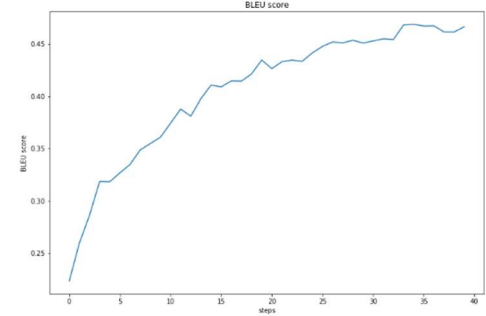
實做的3個方法的公式如下:

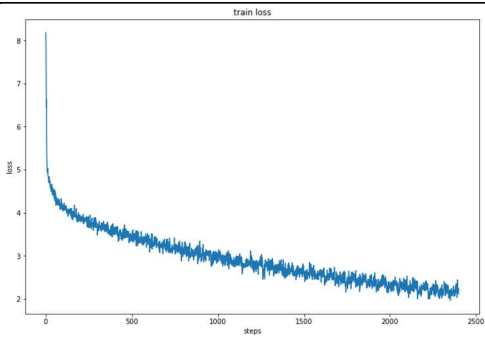
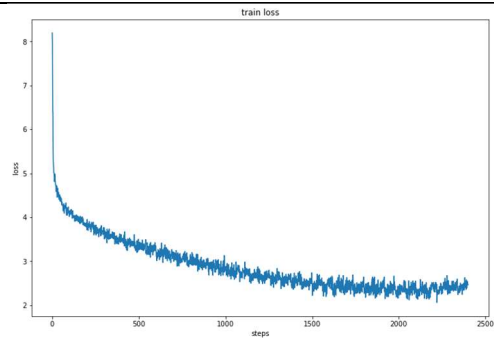
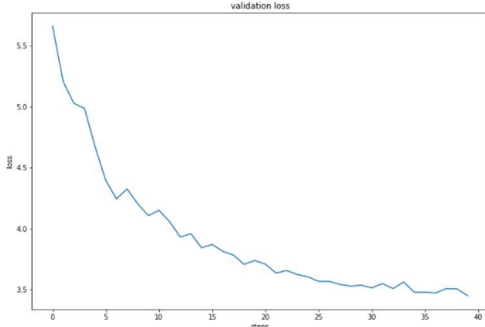
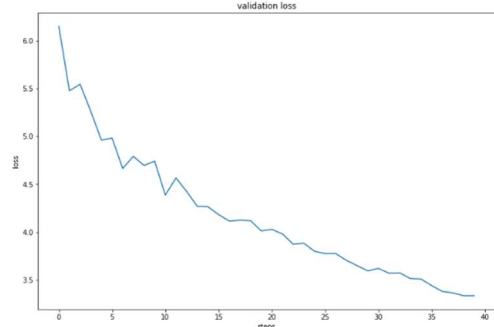
$$\text{令 } x = \frac{\text{Steps}_{\text{current}}}{\text{Steps}_{\text{total}}}$$

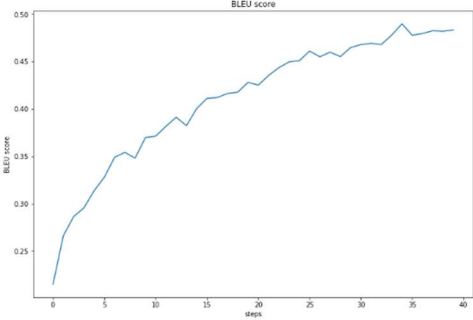
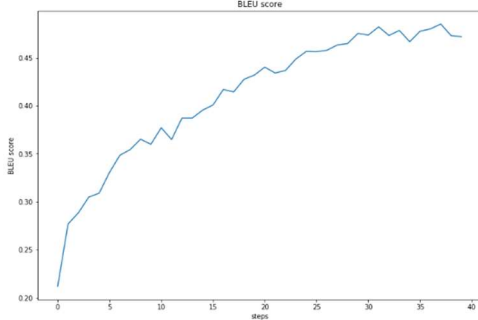
- ◆ Linear decay: $1 - x$
- ◆ Exponential decay: $0.5^{(10*x)}$
- ◆ Inverse Sigmoid decay: $1 - 0.2 * \frac{x}{0.2 - x + 1}$

實作結果為有使用 Schedule Sampling 的方式在 Testing 時的 loss 與 baseline 差不多，但是在 test blue score 上都有大幅度改進，其中表現最好的是 Linear decay 方法。

	Baseline	Linear decay
Train Loss		

Valid Loss		
Valid BLUE Score		
Test Loss:	8.247097495474113	8.24243473753401
Test BLUE Score	0.0015560877977023727	0.008235573582307285

	Exponential decay	Inverse Sigmoid decay
Train Loss		
Valid Losses		

Valid BLEU Score	 <p>A line graph titled 'BLEU score' showing the performance of a model over 40 steps. The y-axis represents the BLEU score, ranging from 0.25 to 0.50. The x-axis represents the number of steps, ranging from 0 to 40. The score starts at approximately 0.22 at step 0 and increases steadily, reaching about 0.48 by step 40.</p>	 <p>A line graph titled 'BLEU score' showing the performance of a model over 40 steps. The y-axis represents the BLEU score, ranging from 0.20 to 0.45. The x-axis represents the number of steps, ranging from 0 to 40. The score starts at approximately 0.22 at step 0 and increases steadily, reaching about 0.47 by step 40.</p>
Test Loss:	8.242138063780997	8.24098326548459
Test BLEU Score	0.006969266301045213	0.005498857274351591