**As stated, working in pairs is possible for this assignment.**

**The name and registration number of my partner is as follows:**

**LIU SZE CHUN (Eric) (1403713)**

**Q1) Training POS taggers on the Brown corpus** (15 marks)

1. Output of POS-tagged sentence from UnigramTagger:

**[('I', 'PPSS'), ('walked', 'VBD'), ('5', 'CD'), ('miles', 'NNS'), ('.', '.')]**

**..... It has taken 1.4241983890533447 to run the program using UnigramTagger. .....**

Output of POS-tagged sentence from HMMTagger:

**[('I', 'PPSS'), ('walked', 'VBD'), ('5', 'CD'), ('miles', 'NNS'), ('.', '.')]**

**..... It has taken 7.44858717918396 to run the program using HMMTagger. .....**

1. The UnigramTagger did produce the correct POS tags for the sentence:

* ‘I’ has been listed as other nominative personal pronoun such as (I, we, they, you).
* ‘walked’ has been listed as verb that’s past tense.
* ‘5’ has been listed as cardinal numeral such as (one, two, 2, etc.)
* ‘miles’ has been listed as plural noun.

In comparison, the HMMTagger also produce identical POS tags for the sentence:

* The tagging from UnigramTagger and HMMTagger have been the same.
* Therefore, we imported ‘time’ into the python code to determine the time it takes for the program to run.
* Results shown that even though the taggers produce the same result, UnigramTagger on average takes a lot less time to run.
* This suggests on this occasion, Unigram is slightly more effective than HMMTagger.

**Q2) Comparing taggers** (25 marks)

1. The correct POS for the word “drop” in S1 would be a ‘verb’, as in the prices has dropped/ decreased in the short term. On the other hand, in S2 ‘drop’ functions as a ‘noun’ where it is described that the extent of the drop was very great.
2. **[('The', 'AT'), ('prices', 'NNS'), ('will', 'MD'), ('drop', 'NN'), ('in', 'IN'), ('the', 'AT'), ('short', 'JJ'), ('term', 'NN'), ('.', '.')]**

**[('The', 'AT'), ('extent', 'NN'), ('of', 'IN'), ('the', 'AT'), ('drop', 'NN'), ('was', 'BEDZ'), ('very', 'QL'), ('great', 'JJ'), ('.', '.')]**

**..... It has taken 1.504648208618164 to run the program using UnigramTagger. .....**

For both sentences, the UnigramTagger has produced the same output for ‘drop’. It is listed as ‘NN’ meaning it is singular or mass noun. However, the tagger is incorrect as it could not detect ‘drop’ as a verb from S1. The tagger is unable to differentiate the word ‘drop’ when applied to different forms of meanings.

1. **[('The', 'AT'), ('prices', 'NNS'), ('will', 'MD'), ('drop', 'VB'), ('in', 'IN'), ('the', 'AT'), ('short', 'JJ'), ('term', 'NN'), ('.', '.')]**

**[('The', 'AT'), ('extent', 'NN'), ('of', 'IN'), ('the', 'AT'), ('drop', 'NN'), ('was', 'BEDZ'), ('very', 'QL'), ('great', 'JJ'), ('.', '.')]**

**..... It has taken 7.450746536254883 to run the program using HMMTagger. .....**

Using HMMTagger it has produced the different results for the word ‘drop’. It can detect from the 2 sentences different observations. For example, in S1 ‘drop’ is detected as a verb in base form, whereas in S2 ‘drop’ is perceive as a noun in singular form. Therefore, HMMTagger is more effective as a Tagger in comparison to UnigramTagger.

**Q3) POS Ambiguity** (20 marks)

1. In S1, there are more than one interpretation. There are evidence of semantic/lexical ambiguity, since there is presence of more than one meaning within a single word. Firstly, ‘eyedrops’ could refer to the cleansing liquid used to relieve eyes. This suggesting that the product of eye drops is not on the shelf for sale. In here, ‘eyedrops’ acts as one word and is a noun. On the other hand, ‘eye’ could be a noun with ‘drops’ being a verb. In this case, it would suggest that an eye ball has dropped off a shelf.

In S2, there are also more than one interpretation. In general, this sentence suggests that a squad of paramedics gave medical help to the victim of a dog bite. A noun phrase could be in ‘dog bite victim’, with dog and bite acting as qualifiers of victim. However, if ‘bite’ functions as the ‘verb’ then it could mean ‘A squad has helped a dog to bite someone’.

1. **[('Eye', None), ('drops', 'NNS'), ('off', 'RP'), ('shelf', None), ('.', '.')]**

**[('Squad', 'NN-HL'), ('helps', 'VBZ'), ('dog', 'NN'), ('bite', 'VB'), ('victim', 'NN'), ('.', '.')]**

**..... It has taken 1.4511480331420898 to run the program using UnigramTagger. .....**

For S1, the UnigramTagger has taken ‘drops’ as a plural noun. In this case, it would make the sentence to suggest that ‘the item (eyedrops) has been taken off the market’.

In S2, the UnigramTagger has taken ‘bite’ as a ‘verb’. This would create ambiguity as it would suggest that ‘A squad has helped a dog to bite someone’. Therefore, UnigramTagger is not efficient into removing sentences of lexical or semantic ambiguity.

**Q4) Text Classification** (40 marks)

Prior probabilities:

class 1 = 0.4

class 0 = 0.6

Feature likelihoods

enjoyable great just

class = 0 0.142857 0.285714 0.142857

sad movie boring and just

class = 1 0.1875 0.125 0.1875 0.125 0.125

Predictions on test data

d5 = class 0

d6 = class 1

d7 = class 0

d8 = class 0

d9 = class 0

d10 = class 1

Accuracy on test data = 67 %