Test Output:

1. Sentence: ['Those', 'coming', 'from', 'other', 'denominations', 'will', 'welcome', 'the', 'opportunity', 'to', 'become', 'informed', '.']
   Output: ['DET', 'NOUN', 'ADP', 'ADJ', 'NOUN', 'VERB', 'VERB', 'DET', 'NOUN', 'PRT', 'VERB', 'VERB', '.']
   Ground Truth: ['DET', 'VERB', 'ADP', 'ADJ', 'NOUN', 'VERB', 'VERB', 'DET', 'NOUN', 'PRT', 'VERB', 'VERB', '.']
2. Sentence: ['The', 'preparatory', 'class', 'is', 'an', 'introductory', 'face-to-face', 'group', 'in', 'which', 'new', 'members', 'become', 'acquainted', 'with', 'one', 'another', '.']
   Output: ['DET', 'ADJ', 'NOUN', 'VERB', 'DET', 'NOUN', 'ADP', 'NOUN', 'ADP', 'DET', 'ADJ', 'NOUN', 'VERB', 'VERB', 'ADP', 'NUM', 'DET', '.']
   Ground Truth: ['DET', 'ADJ', 'NOUN', 'VERB', 'DET', 'ADJ', 'ADJ', 'NOUN', 'ADP', 'DET', 'ADJ', 'NOUN', 'VERB', 'VERB', 'ADP', 'NUM', 'DET', '.']
3. Sentence: ['It', 'provides', 'a', 'natural', 'transition', 'into', 'the', 'life', 'of', 'the', 'local', 'church', 'and', 'its', 'organizations', '.']
   Output: ['PRON', 'VERB', 'DET', 'ADJ', 'NOUN', 'ADP', 'DET', 'NOUN', 'ADP', 'DET', 'ADJ', 'NOUN', 'CONJ', 'DET', 'NOUN', '.']
   Ground Truth: ['PRON', 'VERB', 'DET', 'ADJ', 'NOUN', 'ADP', 'DET', 'NOUN', 'ADP', 'DET', 'ADJ', 'NOUN', 'CONJ', 'DET', 'NOUN', '.']

Performance of the POS Tagger:

Average Precision for all tags: 0.798

Average Recall for all tags: 0.792

Accuracy: 0.915

Precision for each tag:

'DET', 1.0
'NOUN', 0.75
'ADJ', 1.0
'VERB', 1.0
'ADP', 0.83
'.', 1.0
'ADV', 0.0
'CONJ', 1.0
'PRT', 1.0
'PRON', 1.0
'NUM', 0.0
'X', 0.0

Recall for each tag:

'DET', 1.0
'NOUN', 1.0

'ADJ', 0.714
'VERB', 0.889
'ADP', 1.0
'.', 1.0
'ADV', 0.0
'CONJ', 1.0
'PRT', 1.0
'PRON', 1.0
'NUM', 1.0
'X', 0.0

Why it does perform well – the training corpus appears to have a similar distribution of POS tags for words when compared to the test data, thus the assumptions of the hidden Markov model hold.

Why it fails in some cases –

- It may fail in cases where the word is unseen in the training corpus. The emission probability of word|state would probably be incorrect. This is the case for the failures for words – 'introductory' and 'face-to-face'
- The model tags 'coming' as a noun in the first case because p(NOUN|DET) = 0.63 >> p(VERB|DET) = 0.061 – as per the transition matrix created from out training corpus