

# HW4 Part-of-speech tagging

Yuxuan Zhang

We will first run our code and compare the inferred result with the ground truth.

```
In [ ]: !python ./HW4.py
```

```
Sentence: ['Those', 'coming', 'from', 'other', 'denominations', 'will', 'welcome', 'the', 'opportunity', 'to', 'become', 'informed', '.']
Predicted Tags: ['DET', 'NOUN', 'ADP', 'ADJ', 'NOUN', 'VERB', 'VERB', 'DET', 'NOUN', 'PRT', 'VERB', 'VERB', '.']
True Tags: ['DET', 'VERB', 'ADP', 'ADJ', 'NOUN', 'VERB', 'VERB', 'DET', 'NOUN', 'PRT', 'VERB', 'VERB', '.']
Sentence: ['The', 'preparatory', 'class', 'is', 'an', 'introductory', 'face-to-face', 'group', 'in', 'which', 'new', 'members', 'become', 'acquainted', 'with', 'one', 'another', '.']
Predicted Tags: ['DET', 'ADJ', 'NOUN', 'VERB', 'DET', 'ADJ', 'NOUN', 'NOUN', 'ADP', 'DET', 'ADJ', 'NOUN', 'VERB', 'VERB', 'ADP', 'NUM', 'NOUN', '.']
True Tags: ['DET', 'ADJ', 'NOUN', 'VERB', 'DET', 'ADJ', 'ADJ', 'NOUN', 'ADP', 'DET', 'ADJ', 'NOUN', 'VERB', 'VERB', 'ADP', 'NUM', 'DET', '.']
Sentence: ['It', 'provides', 'a', 'natural', 'transition', 'into', 'the', 'life', 'of', 'the', 'local', 'church', 'and', 'its', 'organizations', '.']
Predicted Tags: ['PRON', 'VERB', 'DET', 'ADJ', 'NOUN', 'ADP', 'DET', 'NOUN', 'ADP', 'DET', 'ADJ', 'NOUN', 'CONJ', 'DET', 'NOUN', '.']
True Tags: ['PRON', 'VERB', 'DET', 'ADJ', 'NOUN', 'ADP', 'DET', 'NOUN', 'ADP', 'DET', 'ADJ', 'NOUN', 'CONJ', 'DET', 'NOUN', '.']
[nltk_data] Downloading package brown to
[nltk_data] C:\Users\LEGION\AppData\Roaming\nltk_data...
[nltk_data] Package brown is already up-to-date!
[nltk_data] Downloading package universal_tagset to
[nltk_data] C:\Users\LEGION\AppData\Roaming\nltk_data...
[nltk_data] Package universal_tagset is already up-to-date!
```

Here are the reasons why our tagger does produce the correct tags:

- The Hidden Markov models rely on transition probabilities between tags. For example, knowing that an adjective is often followed by a noun helps the model predict the correct tag sequence. Thus, sentence with standard grammatical structures are more likely to have correct predictions.
- The use of add-1 smoothing ensures that unseen or rare transitions and emissions in the training data do not have zero probability. This allows the tagger to consider corner cases for infrequent structures.

Here are the reasons why our tagger does not produce the correct tags:

- In the first sentence, we see that our model incorrectly tagged 'coming' as NOUN instead of VERB. The mistake with "coming" likely arises from part-of-speech ambiguity. The word "coming" can function as both a verb and an adjective in English. In this case, "coming" is used as a verb in the phrase "coming from," but the model may have incorrectly tagged it due to its frequent use as an adjective, as in "the coming event." This ambiguity is

challenging for models like HMMs that rely on limited context and may not fully capture the verb usage in this particular sentence structure.

- In the second sentence, we see that 'face-to-face' is incorrectly tagged as ADJ instead of NOUN. It is possible that the word 'face-to-face' is rarely seen in the training corpus. The absence of these words can lead to incorrect predictions