Ex3 nlp coding part:   
the results are:

Baseline Error Rate: {'total\_error\_rate': 0.08342469849496659, 'known\_error\_rate': 0.04590975582311241, 'unknown\_error\_rate': 0.3743455497382199}

HMM Error Rate: {'total\_error\_rate': 0.07604903817402575, 'known\_error\_rate': 0.029368740857432206, 'unknown\_error\_rate': 0.43804537521815007}

Smoothed HMM Error Rate: {'total\_error\_rate': 0.08153094787202231, 'known\_error\_rate': 0.0349949364239901, 'unknown\_error\_rate': 0.4424083769633508}

Pseudo-Words HMM Error Rate: {'total\_error\_rate': 0.11123293132662215, 'known\_error\_rate': 0.027628649993452926, 'unknown\_error\_rate': 0.3777128547579299}

Smoothed Pseudo-Words HMM Error Rate: {'total\_error\_rate': 0.11452207714542006, 'known\_error\_rate': 0.029985596438392038, 'unknown\_error\_rate': 0.38397328881469117}

Confusion Matrix:

[[1301 0 0 0 0 0 0 0 0 0 0 0 0]

[ 0 0 0 0 0 0 0 0 0 0 0 0 0]

[ 0 0 448 2 17 0 46 178 0 0 2 34 0]

[ 0 0 0 1250 14 1 1 8 0 0 17 0 0]

[ 0 0 32 23 256 0 4 59 0 3 5 25 0]

[ 0 0 0 0 0 286 1 0 0 0 0 0 0]

[ 0 0 0 6 0 0 1133 3 0 6 0 0 0]

[ 1 0 40 15 0 0 75 2487 0 24 1 70 0]

[ 0 0 3 2 0 0 10 32 143 0 0 3 0]

[ 0 0 1 13 0 0 0 0 0 209 0 0 0]

[ 0 0 0 43 0 0 1 0 0 1 222 0 0]

[ 6 0 29 31 0 0 21 221 0 1 1 1144 0]

[ 0 0 0 0 0 0 5 12 0 0 0 0 5]]

Most Frequent Errors:

True Tag: VERB, Predicted Tag: NOUN, Count: 221

True Tag: ADJ, Predicted Tag: NOUN, Count: 178

True Tag: NOUN, Predicted Tag: DET, Count: 75

True Tag: NOUN, Predicted Tag: VERB, Count: 70

True Tag: ADV, Predicted Tag: NOUN, Count: 59

True Tag: ADJ, Predicted Tag: DET, Count: 46

True Tag: PRT, Predicted Tag: ADP, Count: 43

True Tag: NOUN, Predicted Tag: ADJ, Count: 40

True Tag: ADJ, Predicted Tag: VERB, Count: 34

True Tag: ADV, Predicted Tag: ADJ, Count: 32

All **unknown words** are tagged as **NOUN**, which can contribute to errors when the true tag is something else. We do see that many of the errors are mis-tagging of words as nouns, which might be because they were not seen before and therefore tagged as noun  
  
 **VERB → NOUN**: Ambiguity in words like "play" or "test" (can be both verb and noun).

**ADJ → NOUN**: Adjectives like "cold" used as nouns ("the cold").

**NOUN → DET**: Words like "that" or "which" mistaken for determiners due to their dual roles.