Practical Part - NLP Ex3

	MLE	HMM	HMM- Smoothing	HMM-Pseudo	HMM-Pseudo-Smoothing
Known	0.0611	0.1394	0.1269	0.1534	0.1324
Unknown	0.7482	0.6442	0.6442	0.4876	0.4959
Total	0.1199	0.1823	0.1711	0.1819	0.1634

Top 10 Errors from Confusion Matrix Are:

(133, {'true_tag': 'NNS', 'predicted_tag': 'NN'})

(114, {'true_tag': 'NN', 'predicted_tag': 'NP'})

(91, {'true_tag': 'NP', 'predicted_tag': 'NN'})

(86, {'true_tag': 'JJ', 'predicted_tag': 'NN'})

(45, {'true_tag': 'NNS', 'predicted_tag': 'NP'})

(42, {'true_tag': 'JJ', 'predicted_tag': 'NP'})

(34, {'true_tag': 'VBG', 'predicted_tag': 'NN'})

(26, {'true_tag': 'NP', 'predicted_tag': 'JJ'})

(25, {'true_tag': 'NN', 'predicted_tag': 'JJ'})

(25, {'true_tag': 'JJ', 'predicted_tag': 'VBN'})

Conclusion:

While the addition of Pseudo Words generation and Laplace Smoothing techniques improves the HMM model's performance, it still doesn't surpass the MLE model for this dataset. This unexpected result might be attributed to the relatively small size of our dataset and the high proportion of unknown words, which can significantly impact model accuracy.