1.3

	embed_dim	hidden_size	Num_layers	bidirectional	batch	dropout	Val_accuracy
Exp1	64	64	1	false	32	0	0.77
Exp2	150	150	1	false	32	0	0.79
Exp3	300	150	1	false	32	0.2	0.79
Exp4	300	256	2	true	32	0.2	0.78
Exp5	256	256	1	true	16	0.2	0.82

Finally we choose exp5 for the high eval accuracy

1.4

Test accuracy 0.818

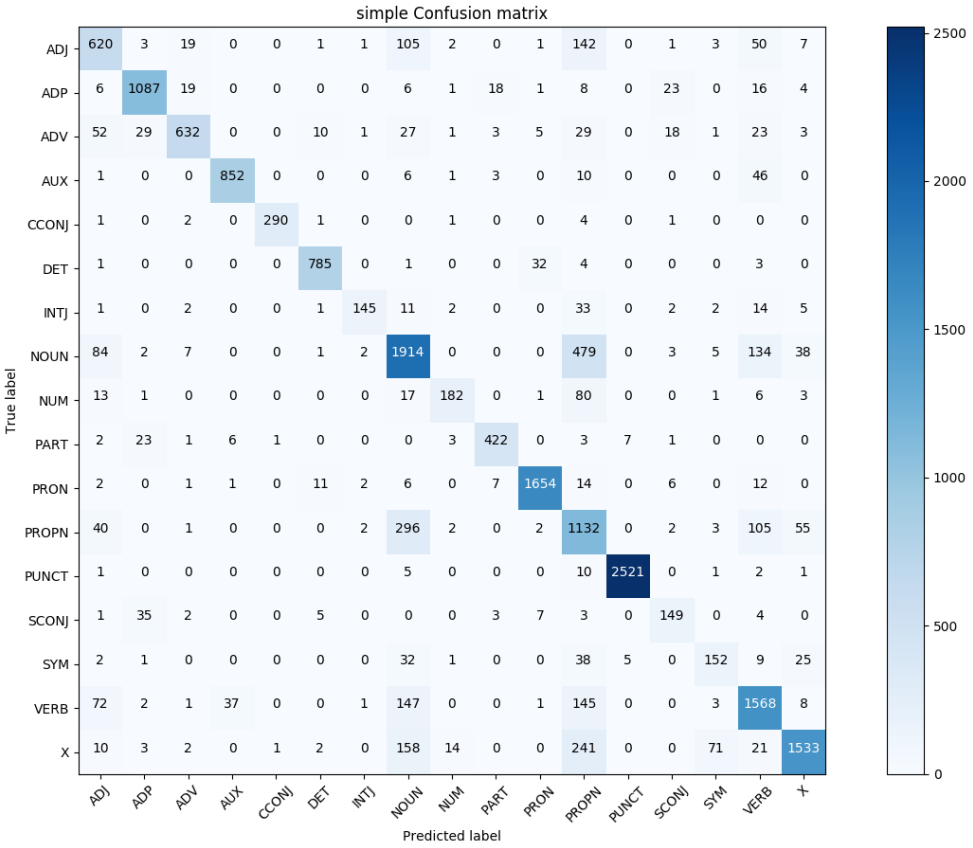
2.3

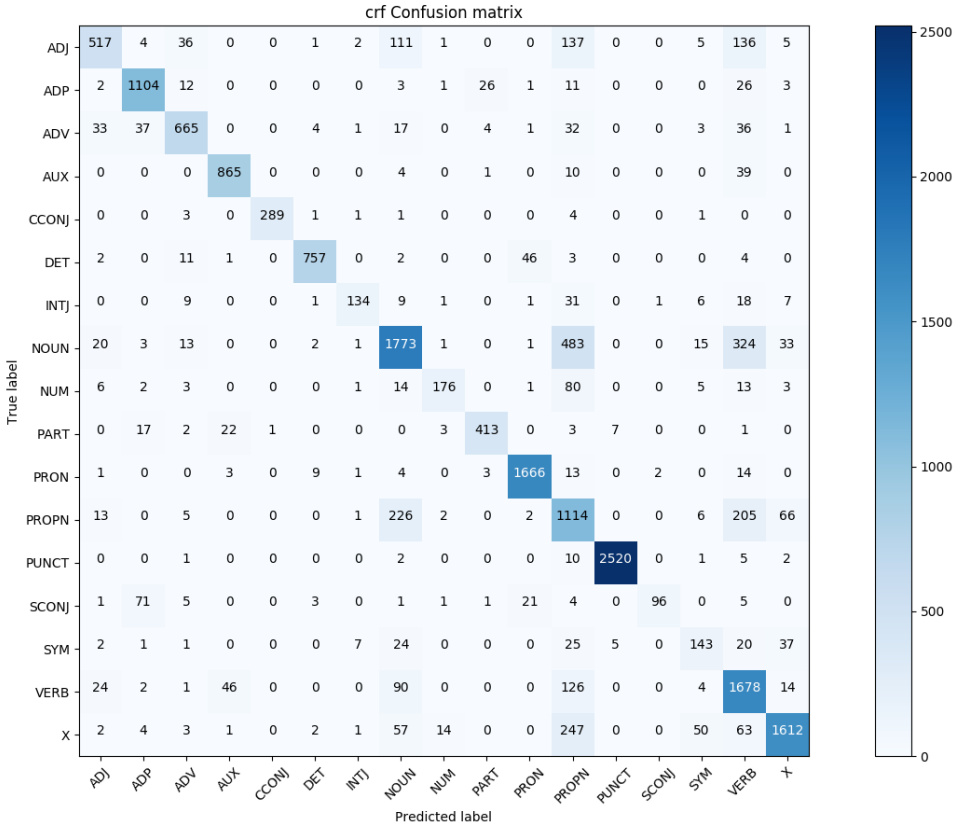
	embed_dim	hidden_size	Num_layers	bidirectional	batch	dropout	Val_accuracy
Exp1	64	64	1	false	32	0	0.784
Exp2	150	150	1	false	32	0	0.784
Exp3	300	150	1	false	32	0.2	0.798
Exp4	300	256	2	true	32	0.2	0.786
Exp5	256	256	1	true	16	0.2	0.812

So we use the exp5

2.4
The test set accuracy is 0.8128

3.1 See next page


simple Confusion matrix
True label
Predicted label
ADJ ADP ADV AUX CCONJ DET INTJ NOUN NUM PART PRON PROPN PUNCT SCONJ SYM VERB X
0 500 1000 1500 2000 2500


crf Confusion matrix
True label
Predicted label
ADJ ADP ADV AUX CCONJ DET INTJ NOUN NUM PART PRON PROPN PUNCT SCONJ SYM VERB X
0 500 1000 1500 2000 2500

3.2
Both of the models label punctuation really well. I think this is because of the fact that punctuations are of their distinct characters.

A difference I noticed is that the simple model tend to label the x and verbs into nouns but the other model do this way less.

A common place that they make mistakes is that they tend to label pronoun as noun incorrectly and vice versa.

4.1

$$\heartsuit_i(y', y) = \max_{y'' \in \mathcal{L}} s(x, l-2, y'', y', y) + \heartsuit_{l-1}(y'', y')$$
$$b_l(y', y) = \operatorname*{argmax}_{y'' \in \mathcal{L}} s(x, l-2, y'', y', y) + \heartsuit_{l-1}(y'', y')$$

4.2
Similar to the slide but we have one more layer of L so we have Space:$O(|\mathcal{L}|^2 \ell)$ and Runtime $O(|\mathcal{L}|^3 \ell)$