

Hyper-parameter	CoNLL-2003 (Round 2)		OntoNotes 5.0 (Round 1)	
	Final	Range	Final	Range
Convolution width	3	[3, 7]	3	[3, 9]
CNN output size	53	[15, 84]	20	[15, 100]
LSTM state size	275	[100, 500]	200	[100, 400] ¹
LSTM layers	1	[1, 4]	2	[2, 4]
Learning rate	0.0105	$[10^{-3}, 10^{-1.8}]$	0.008	$[10^{-3.5}, 10^{-1.5}]$
Epochs ²	80	-	18	-
Dropout ³	0.68	[0.25, 0.75]	0.63	[0, 1]
Mini-batch size	9	- ⁴	9	[5, 14]