Language model	PPL	WER
4-gram LM (baseline)	69.4	8.6
+ RNNLM, CTS data only	62.6	7.6
+ Web data training	60.9	7.4
+ 2nd hidden layer	59.0	7.4
+ 2-RNNLM interpolation	57.2	7.3
+ backward RNNLMs	-	6.9
+ LSTM-LM, CTS + Web data	51.4	6.9
+ 2-LSTM-LM interpolation	50.5	6.8
+ backward LSTM-LM	-	6.6