

1000		1050	
1001		1051	
1002		1052	
1003		1053	
1004		1054	
1005		1055	
1006		1056	
1007		1057	
1008		1058	
1009		1059	
1010		1060	
1011		1061	
1012		1062	
1013		1063	
1014		1064	
1015		1065	
1016		1066	
1017	Hyperparameter Name	Description	Values Tested
1018	rand_reset	whether activations in the model reset to a random state (True), or to zero (False) after each word	True , False
1019	lr	initial learning rate	0.1, 0.5, 1.0 , 2.0
1020	anneal_factor	amount by which to anneal learning rate, if no improvement found	0, 0.25 , 0.5, 1.0
1021	patience	number of training epochs to wait for validation loss to improve before updating weights	0 , 2, 4
1022	dropout	proportion of weights to keep fixed	0, 0.2 , 0.5
1023	epochs	number of epochs (complete passes through the data) to train for	25 , 50, 100
1024	nlayers	number of recurrent layers	1, 2 , 4
1025	nhid	number of nodes in each recurrent layer	128, 256, 512 , 1250

1030 Table A1: Particulars of hyperparameter testing. Hyperparameters were optimized for speed and likelihood as-
 1031 signed to the validation set. Optimal parameters for the validation set are bolded, and were used in the experiments
 1032 reported here.

1033		1083
1034		1084
1035		1085
1036		1086
1037		1087
1038		1088
1039		1089
1040		1090
1041		1091
1042		1092
1043		1093
1044		1094
1045		1095
1046		1096
1047		1097
1048		1098
1049		1099