Base Case Avg Accy	0.660	0.681	0.708	0.762	
0.5 Threshold over Base Case	0.758	0.734	0.706	0.657	
	Avg Accy Ratio Comparison (vs base case)				
Method \ Seq len	5	10	25	50	
Inhibit	0.98	0.96	0.97	0.92	NOUN, VERB ,ADJ, (ADJ,NOUN)
Random Permute Token	0.98	0.964	0.935	0.832	
Random Merge Token	0.9	0.91	0.94	0.906	
Merge ADJ Token	0.905	0.88	0.634	0.523	
Merge VERB Token	0.902	0.885	0.863	0.854	
PermuteToken+Random Merge Token	0.95	0.86	0.673	?	

Cells in yellow color are under 0.5 avg accy

5 methods are used to corrupt the structure

Method	Description
Inhibit	Change loss value (cross entropy) to mean=0.5 ,var =0.05 for specific POS tag position (NOUN , VERB, ADJ, ADJ+NOUN)
Random Permute Token	Random permutation of input token sequence
Random Merge Token	Merge token randomly with min2, max 3
Merge <POS> Token	Only Merge surrounding token of specificy POS tag
Permute+Merge	Combine randome permute token and random merge

Remark :

The top 10 merge count are different between seq 5,10 and 25,50

Random Merge Seq 5, 10	Random Merge Seq 25, 50	Merge Only ADJ
('NOUN_NOUN', 1299403), ('NOUN_ADJ', 794882), ('VERB_NOUN', 782272), ('NOUN_VERB', 684696), ('ADJ_NOUN', 501560), ('NOUN_NOUN_NOUN', 433874), ('NOUN_NOUN_ADJ', 388947), ('NOUN_VERB_NOUN', 296444), ('ADJ_DET', 191582), ('VERB_NOUN_NOUN', 148550)	('ADJ_ADJ_ADJ', 4748367), ('ADJ_ADJ', 3725680), ('NOUN_NOUN', 3126343), ('NOUN_ADJ', 1781641), ('VERB_NOUN', 1771679), ('NOUN_VERB', 1507486), ('NOUN_NOUN_NOUN', 1481755), ('NOUN_NOUN_ADJ', 1118719), ('NOUN_VERB_NOUN', 979389), ('ADJ_NOUN', 872810)	('ADJ_ADJ_ADJ', 2131074), ('ADJ_ADJ', 1516601), ('ADJ_NOUN', 260508), ('ADJ_ADJ_NOUN', 156022), ('NOUN_NOUN', 136952), ('NOUN_ADJ_NOUN', 105874), ('NOUN_ADJ', 93044), ('NOUN_VERB', 87431), ('DET_ADJ', 79607), ('NOUN_NOUN_NOUN', 56365)

2. For random token permutation, a measurment between the original POS pattern and after permutation is add to track the difference

For seq 5, 10 , training and validation is rougthly 0.5-0.6
For seq 25 , 50 training and validation are different , training : 0.6 , validation : 0.7-0.8