Token Type Distribution by Utterance Length Full corpus Pruned corpus [Sequence Length ≤ 128] **Token Type** None Claim Premise (0,100](300,400][200,600][600,700] [800,900] (400,500][400,500][700,800]200,300] 500,600] [600,700][700,800] [800,900][900, 1000]1000,1100] 1100,1200] 1200,1300] 1300,1400] 1400,1500] 1500,1600] 1600,1700] 100,200] 200,300] [900,1000] 1000,1100] 1100,1200] 1200,1300] 1300,1400] 1400,1500] 1500,1600] 300,400]

150000

100000

50000

Token Count

Binned Utterance Length