Transformer test loss on random tasks with different regular alphabet sizes 2.00 regular alphabet size 0 regular alphabet size 1 1.75 regular alphabet size 2 1.50 1.25 Test loss 1.00 0.75 0.50 0.25 0.00 10^{1} 10^{2} 10³ 10^{4} n = train alphabet size = # samples