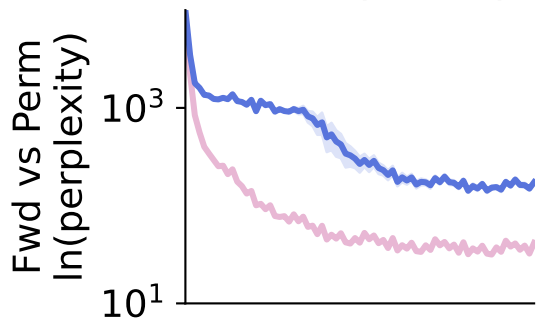
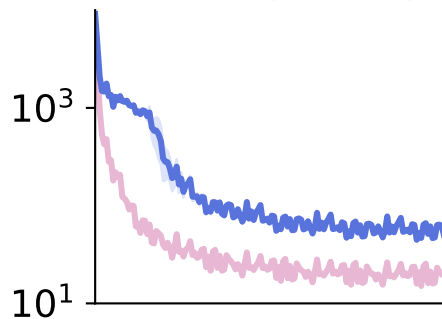


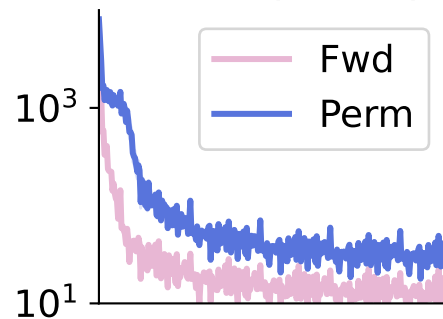
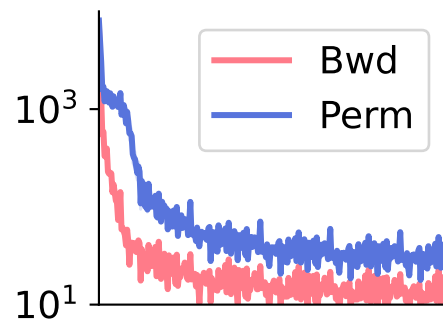
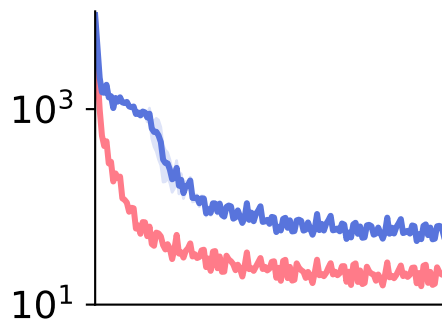
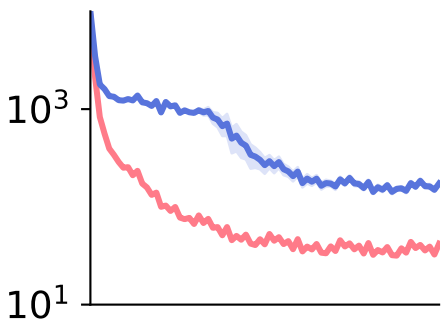
GPT-2 (124M)



GPT-2 (355M)



GPT-2 (774M)

Bwd vs Perm
 $\ln(\text{perplexity})$ Fwd - Bwd
Difference