

Small model	Perplexity
Baseline w175/w200	76.78/76
c10 with $n = 2$	75.23
c10 with $n = 3$	75.04
Large model	Perplexity
Baseline w475/w650	70.88/70.69
c25 with $n = 4$	68.79
c25 with $n = 6(b)$	67.64