INFO: Epoch 021: loss 1.921 | lr 0.0003 | num\_tokens 10.31 | batch\_size 1 | grad\_norm 61.58 | clip 0.9996

INFO: Epoch 021: valid\_loss 3.12 | num\_tokens 10.1 | batch\_size 500 | valid\_perplexity 22.8

INFO: Epoch 022: loss 1.883 | lr 0.0003 | num\_tokens 10.31 | batch\_size 1 | grad\_norm 61.56 | clip 0.9997

INFO: Epoch 022: valid\_loss 3.11 | num\_tokens 10.1 | batch\_size 500 | valid\_perplexity 22.3

INFO: Epoch 023: loss 1.85 | lr 0.0003 | num\_tokens 10.31 | batch\_size 1 | grad\_norm 61.58 | clip 0.9995

INFO: Epoch 023: valid\_loss 3.11 | num\_tokens 10.1 | batch\_size 500 | valid\_perplexity 22.5

INFO: Epoch 024: loss 1.818 | lr 0.0003 | num\_tokens 10.31 | batch\_size 1 | grad\_norm 61.61 | clip 0.9995

INFO: Epoch 024: valid\_loss 3.11 | num\_tokens 10.1 | batch\_size 500 | valid\_perplexity 22.5

INFO: Epoch 025: loss 1.786 | lr 0.0003 | num\_tokens 10.31 | batch\_size 1 | grad\_norm 61.75 | clip 0.9997

INFO: Epoch 025: valid\_loss 3.1 | num\_tokens 10.1 | batch\_size 500 | valid\_perplexity 22.3

INFO: Epoch 026: loss 1.757 | lr 0.0003 | num\_tokens 10.31 | batch\_size 1 | grad\_norm 61.74 | clip 0.9994

INFO: Epoch 026: valid\_loss 3.1 | num\_tokens 10.1 | batch\_size 500 | valid\_perplexity 22.1

INFO: Epoch 027: loss 1.728 | lr 0.0003 | num\_tokens 10.31 | batch\_size 1 | grad\_norm 61.68 | clip 0.9989

INFO: Epoch 027: valid\_loss 3.1 | num\_tokens 10.1 | batch\_size 500 | valid\_perplexity 22.3

INFO: Epoch 028: loss 1.705 | lr 0.0003 | num\_tokens 10.31 | batch\_size 1 | grad\_norm 61.77 | clip 0.9987

INFO: Epoch 028: valid\_loss 3.11 | num\_tokens 10.1 | batch\_size 500 | valid\_perplexity 22.5

INFO: Epoch 029: loss 1.676 | lr 0.0003 | num\_tokens 10.31 | batch\_size 1 | grad\_norm 61.8 | clip 0.9988

INFO: Epoch 029: valid\_loss 3.13 | num\_tokens 10.1 | batch\_size 500 | valid\_perplexity 22.9

INFO: Epoch 030: loss 1.656 | lr 0.0003 | num\_tokens 10.31 | batch\_size 1 | grad\_norm 61.79 | clip 0.9985

INFO: Epoch 030: valid\_loss 3.11 | num\_tokens 10.1 | batch\_size 500 | valid\_perplexity 22.3

INFO: Epoch 031: loss 1.629 | lr 0.0003 | num\_tokens 10.31 | batch\_size 1 | grad\_norm 61.69 | clip 0.9984

INFO: Epoch 031: valid\_loss 3.1 | num\_tokens 10.1 | batch\_size 500 | valid\_perplexity 22.3

INFO: No validation set improvements observed for 5 epochs. Early stop!