INFO: Epoch 030: valid\_loss 3.28 | num\_tokens 10.1 | batch\_size 500 | valid\_perplexity 26.5

INFO: Epoch 031: loss 2.631 | lr 0.0003 | num\_tokens 10.31 | batch\_size 1 | grad\_norm 47.11 | clip 1

INFO: Epoch 031: valid\_loss 3.29 | num\_tokens 10.1 | batch\_size 500 | valid\_perplexity 26.7

INFO: Epoch 032: loss 2.614 | lr 0.0003 | num\_tokens 10.31 | batch\_size 1 | grad\_norm 47.17 | clip 0.9999

INFO: Epoch 032: valid\_loss 3.27 | num\_tokens 10.1 | batch\_size 500 | valid\_perplexity 26.4

INFO: Epoch 033: loss 2.592 | lr 0.0003 | num\_tokens 10.31 | batch\_size 1 | grad\_norm 46.99 | clip 1

INFO: Epoch 033: valid\_loss 3.27 | num\_tokens 10.1 | batch\_size 500 | valid\_perplexity 26.3

INFO: Epoch 034: loss 2.643 | lr 0.0003 | num\_tokens 10.31 | batch\_size 1 | grad\_norm 47.69 | clip 1

INFO: Epoch 034: valid\_loss 3.26 | num\_tokens 10.1 | batch\_size 500 | valid\_perplexity 26

INFO: Epoch 035: loss 2.621 | lr 0.0003 | num\_tokens 10.31 | batch\_size 1 | grad\_norm 47.05 | clip 1

INFO: Epoch 035: valid\_loss 3.26 | num\_tokens 10.1 | batch\_size 500 | valid\_perplexity 26

INFO: Epoch 036: loss 2.606 | lr 0.0003 | num\_tokens 10.31 | batch\_size 1 | grad\_norm 47.29 | clip 1

INFO: Epoch 036: valid\_loss 3.27 | num\_tokens 10.1 | batch\_size 500 | valid\_perplexity 26.2

INFO: Epoch 037: loss 2.583 | lr 0.0003 | num\_tokens 10.31 | batch\_size 1 | grad\_norm 47.34 | clip 1

INFO: Epoch 037: valid\_loss 3.26 | num\_tokens 10.1 | batch\_size 500 | valid\_perplexity 25.9

INFO: Epoch 038: loss 2.577 | lr 0.0003 | num\_tokens 10.31 | batch\_size 1 | grad\_norm 47.11 | clip 1

INFO: Epoch 038: valid\_loss 3.23 | num\_tokens 10.1 | batch\_size 500 | valid\_perplexity 25.4

INFO: Epoch 039: loss 2.558 | lr 0.0003 | num\_tokens 10.31 | batch\_size 1 | grad\_norm 47.28 | clip 1

INFO: Epoch 039: valid\_loss 3.23 | num\_tokens 10.1 | batch\_size 500 | valid\_perplexity 25.3

INFO: Epoch 040: loss 2.537 | lr 0.0003 | num\_tokens 10.31 | batch\_size 1 | grad\_norm 47.16 | clip 1

INFO: Epoch 040: valid\_loss 3.23 | num\_tokens 10.1 | batch\_size 500 | valid\_perplexity 25.3

INFO: Epoch 041: loss 2.525 | lr 0.0003 | num\_tokens 10.31 | batch\_size 1 | grad\_norm 47.18 | clip 1

INFO: Epoch 041: valid\_loss 3.23 | num\_tokens 10.1 | batch\_size 500 | valid\_perplexity 25.4

INFO: Epoch 042: loss 2.51 | lr 0.0003 | num\_tokens 10.31 | batch\_size 1 | grad\_norm 47.12 | clip 1

INFO: Epoch 042: valid\_loss 3.23 | num\_tokens 10.1 | batch\_size 500 | valid\_perplexity 25.2

INFO: Epoch 043: loss 2.498 | lr 0.0003 | num\_tokens 10.31 | batch\_size 1 | grad\_norm 47.53 | clip 1

INFO: Epoch 043: valid\_loss 3.24 | num\_tokens 10.1 | batch\_size 500 | valid\_perplexity 25.6

INFO: Epoch 044: loss 2.483 | lr 0.0003 | num\_tokens 10.31 | batch\_size 1 | grad\_norm 47.38 | clip 0.9999

INFO: Epoch 044: valid\_loss 3.23 | num\_tokens 10.1 | batch\_size 500 | valid\_perplexity 25.4

INFO: Epoch 045: loss 2.469 | lr 0.0003 | num\_tokens 10.31 | batch\_size 1 | grad\_norm 47.37 | clip 1

INFO: Epoch 045: valid\_loss 3.23 | num\_tokens 10.1 | batch\_size 500 | valid\_perplexity 25.3

INFO: Epoch 046: loss 2.456 | lr 0.0003 | num\_tokens 10.31 | batch\_size 1 | grad\_norm 47.49 | clip 0.9999

INFO: Epoch 046: valid\_loss 3.22 | num\_tokens 10.1 | batch\_size 500 | valid\_perplexity 25

INFO: Epoch 047: loss 2.442 | lr 0.0003 | num\_tokens 10.31 | batch\_size 1 | grad\_norm 47.71 | clip 1

INFO: Epoch 047: valid\_loss 3.23 | num\_tokens 10.1 | batch\_size 500 | valid\_perplexity 25.2

INFO: Epoch 048: loss 2.43 | lr 0.0003 | num\_tokens 10.31 | batch\_size 1 | grad\_norm 47.59 | clip 1

INFO: Epoch 048: valid\_loss 3.24 | num\_tokens 10.1 | batch\_size 500 | valid\_perplexity 25.4

INFO: Epoch 049: loss 2.422 | lr 0.0003 | num\_tokens 10.31 | batch\_size 1 | grad\_norm 47.69 | clip 1

INFO: Epoch 049: valid\_loss 3.23 | num\_tokens 10.1 | batch\_size 500 | valid\_perplexity 25.2

INFO: Epoch 050: loss 2.404 | lr 0.0003 | num\_tokens 10.31 | batch\_size 1 | grad\_norm 47.79 | clip 1

INFO: Epoch 050: valid\_loss 3.23 | num\_tokens 10.1 | batch\_size 500 | valid\_perplexity 25.2

INFO: Epoch 051: loss 2.399 | lr 0.0003 | num\_tokens 10.31 | batch\_size 1 | grad\_norm 47.75 | clip 1

INFO: Epoch 051: valid\_loss 3.22 | num\_tokens 10.1 | batch\_size 500 | valid\_perplexity 24.9

INFO: Epoch 052: loss 2.39 | lr 0.0003 | num\_tokens 10.31 | batch\_size 1 | grad\_norm 47.63 | clip 0.9999

INFO: Epoch 052: valid\_loss 3.21 | num\_tokens 10.1 | batch\_size 500 | valid\_perplexity 24.8

INFO: Epoch 053: loss 2.376 | lr 0.0003 | num\_tokens 10.31 | batch\_size 1 | grad\_norm 47.69 | clip 0.9999

INFO: Epoch 053: valid\_loss 3.2 | num\_tokens 10.1 | batch\_size 500 | valid\_perplexity 24.6

INFO: Epoch 054: loss 2.363 | lr 0.0003 | num\_tokens 10.31 | batch\_size 1 | grad\_norm 47.98 | clip 0.9999

INFO: Epoch 054: valid\_loss 3.2 | num\_tokens 10.1 | batch\_size 500 | valid\_perplexity 24.6

INFO: Epoch 055: loss 2.356 | lr 0.0003 | num\_tokens 10.31 | batch\_size 1 | grad\_norm 47.82 | clip 0.9999

INFO: Epoch 055: valid\_loss 3.21 | num\_tokens 10.1 | batch\_size 500 | valid\_perplexity 24.7

INFO: Epoch 056: loss 2.342 | lr 0.0003 | num\_tokens 10.31 | batch\_size 1 | grad\_norm 47.61 | clip 1

INFO: Epoch 056: valid\_loss 3.22 | num\_tokens 10.1 | batch\_size 500 | valid\_perplexity 25

INFO: Epoch 057: loss 2.331 | lr 0.0003 | num\_tokens 10.31 | batch\_size 1 | grad\_norm 48.1 | clip 1

INFO: Epoch 057: valid\_loss 3.21 | num\_tokens 10.1 | batch\_size 500 | valid\_perplexity 24.7

INFO: Epoch 058: loss 2.326 | lr 0.0003 | num\_tokens 10.31 | batch\_size 1 | grad\_norm 48.06 | clip 1

INFO: Epoch 058: valid\_loss 3.2 | num\_tokens 10.1 | batch\_size 500 | valid\_perplexity 24.6

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