**Transformer Results**

1 sentence = about 32 tokens

1 batch = 52 sentences

1 step = 15 batches = 780 sentences = about 24960 tokens (per update)

1 epoch = 84,193 batches 🡪 6 hours per epoch with a 2080ti

18 epochs = 1,515,474 batches = about 101,031 steps (updates)

{"max\_sen\_len": 128, "max\_epoch": 18, "step\_batch": 15, "batch\_size": 52, "random\_seed": 42}

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| epoch | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 |
| sacre | . | . | 26.2 | 27.0 | 27.3 | 28.10 | 28.18 | 28.22 | 28.62 | 28.60 | . | . | . | 28.82 | 29.10 | . | . | 29.11 |
| multi | . | . | 26.5 | 27.2 | 27.5 | 28.29 | 28.33 | 28.39 | 28.76 | 28.77 | . | . | . | 28.93 | 29.24 | . | . | 29.27 |