Training with bert-base-uncased:

Data split in train/dev/test set, no shuffling
Epochs: **3**
Duration: **10:47**
Sequence length: **64 (words), 128 (tokens)**
Batch size: **16**

Train samples: **15259**
Test samples: **1405**
Dev samples: **1374**

	Precision	Recall	F1-score	support
BOS	0.89264	0.90279	0.89769	12056
O	0.98453	0.98275	0.98364	75880
accuracy			0.97179	87936
Macro avg	0.93858	0.94277	0.94066	87936
Weighted avg	0.97193	0.97179	0.97185	87936

Confusion matrix:

	BOS	10884	1172
True	O	1309	74571
		BOS	O
		Predicted	

Accuracy: **90.28%**

Sentences in test data: **12057**
Number of 1:1 correctly predicted sentences: **9883**
Accuracy: **81.97%**