Pre-trained language models received extensive attention in recent years. However, it is still challenging to incorporate a pre-trained model such as BERT into natural language generation tasks. This work investigates a recent method called adapters as an alternative to fine-tuning the whole model in machine translation. Adapters are a promising approach that allows fine-tuning only a tiny fraction of a pre-trained network. We show that with proper initialization, adapters can help achieve better performance than training models from scratch while training substantially fewer weights than the original model. We further show that even with randomly set weights used as the base models for fine-tuning, we can achieve similar performance to one of the baseline models, bypassing the need to train hundreds of millions of weights in the pre-training phase. Furthermore, we study the effectiveness of adapters in the Transformer model for machine translation. We put adapters either in the encoder or the decoder only, and we also attempt to down-scale the pre-trained model size to decrease GPU memory demands. We found that incorporating adapters in the encoder alone matches the setup's performance when we include the adapters on both the encoder and decoder. Finally, our down-scaling study found that using only half of the original pre-trained weights can positively impact the performance when fine-tuned with adapters. Our experiments show that we can get almost the same performance as the original BERT model after fine-tuning the cross-attention layer.