```
+1)] = block1
            attention_weights['decoder_layer{}_block2_decenc_att'.format
(i+1)] = block2


        # END CODE HERE

        # x.shape == (batch_size, target_seq_len, fully_connected_dim)
        return x, attention_weights
```

In [26]:
```
# UNIT TEST
Decoder_test(Decoder, create_look_ahead_mask, create_padding_mask)
```

```
All tests passed
```

6 - Transformer

Phew! This has been quite the assignment, and now you've made it to your last exercise of the Deep Learning Specialization. Congratulations! You've done all the hard work, now it's time to put it all together.

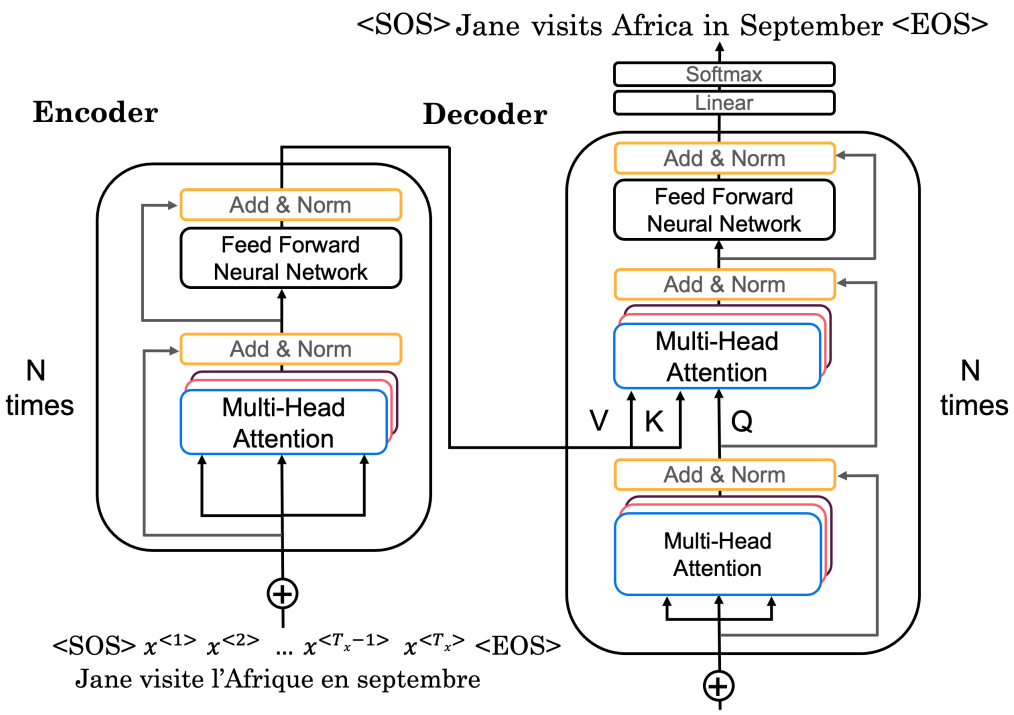

Figure 4: Transformer

The flow of data through the Transformer Architecture is as follows:

- **First your input passes through an Encoder, which is just repeated Encoder layers that you implemented:**
  - **embedding and positional encoding of your input**
  - **multi-head attention on your input**