Tied Transformers

NMT with Shared Encoder - Decoder

https://taoqin.github.io/papers/tiedT.AAAI2019.pdf

What are transformers?

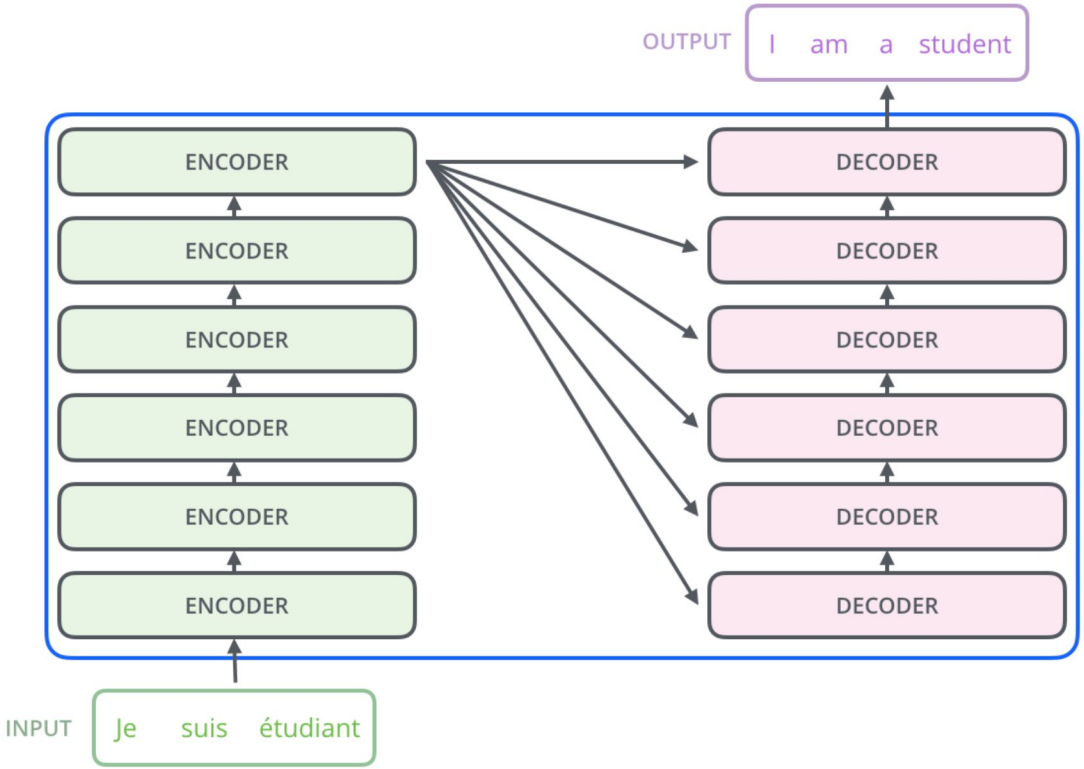
A new **Non-recurrent** approach for modeling natural language.

Previous contenders:

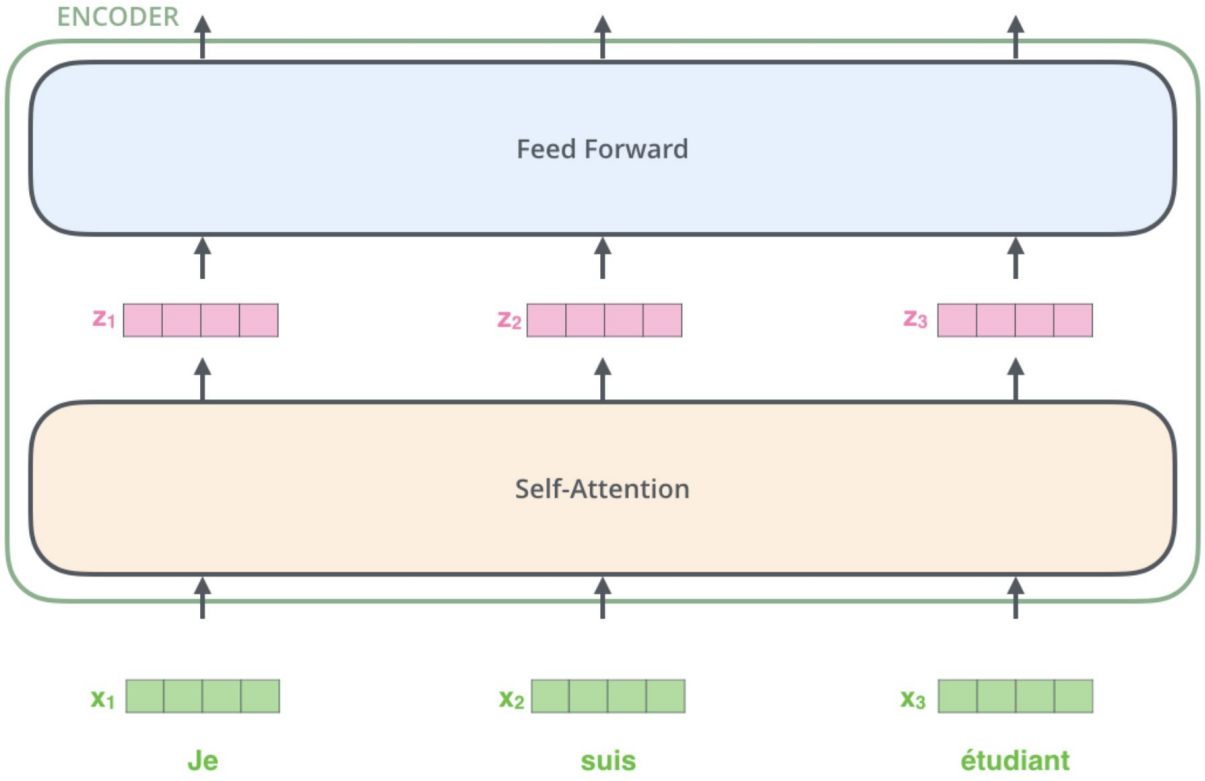
- LSTMs with attention
- CNNs with attention

Simple Example: What if we had none of the above and we wanted to build a model for a NLP task? e.g. sentiment.

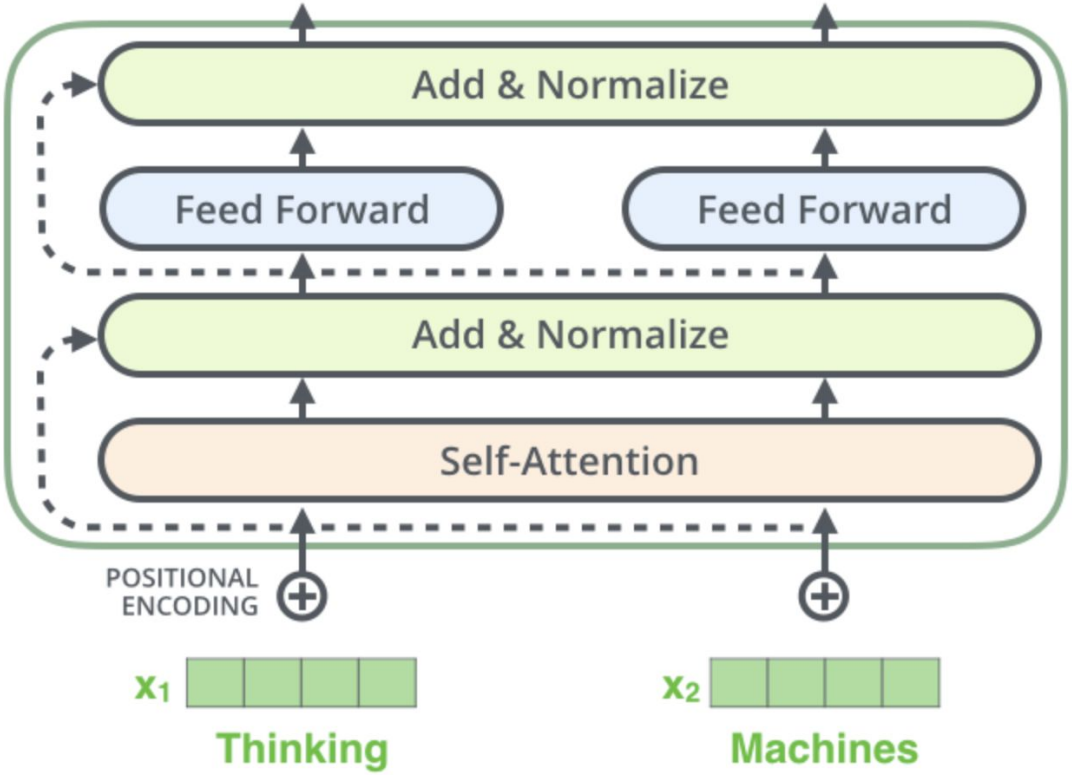
What are transformers?



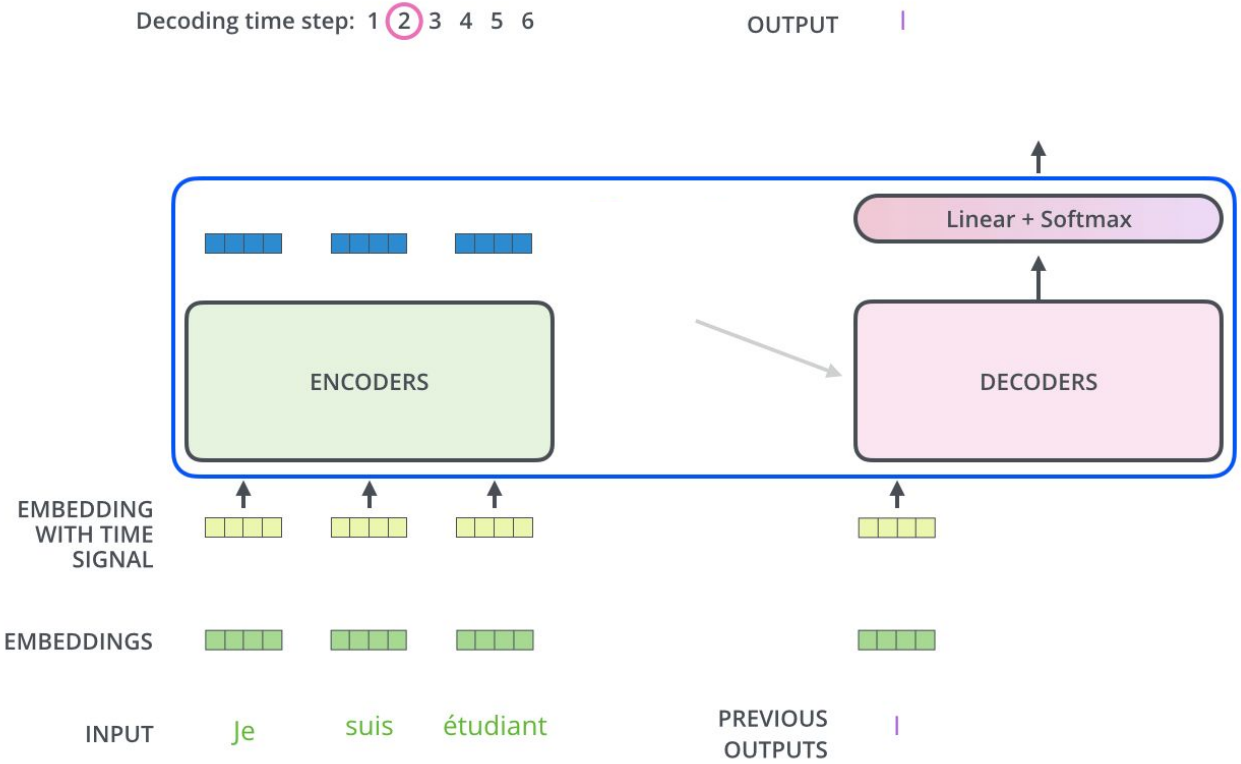
What are transformers?



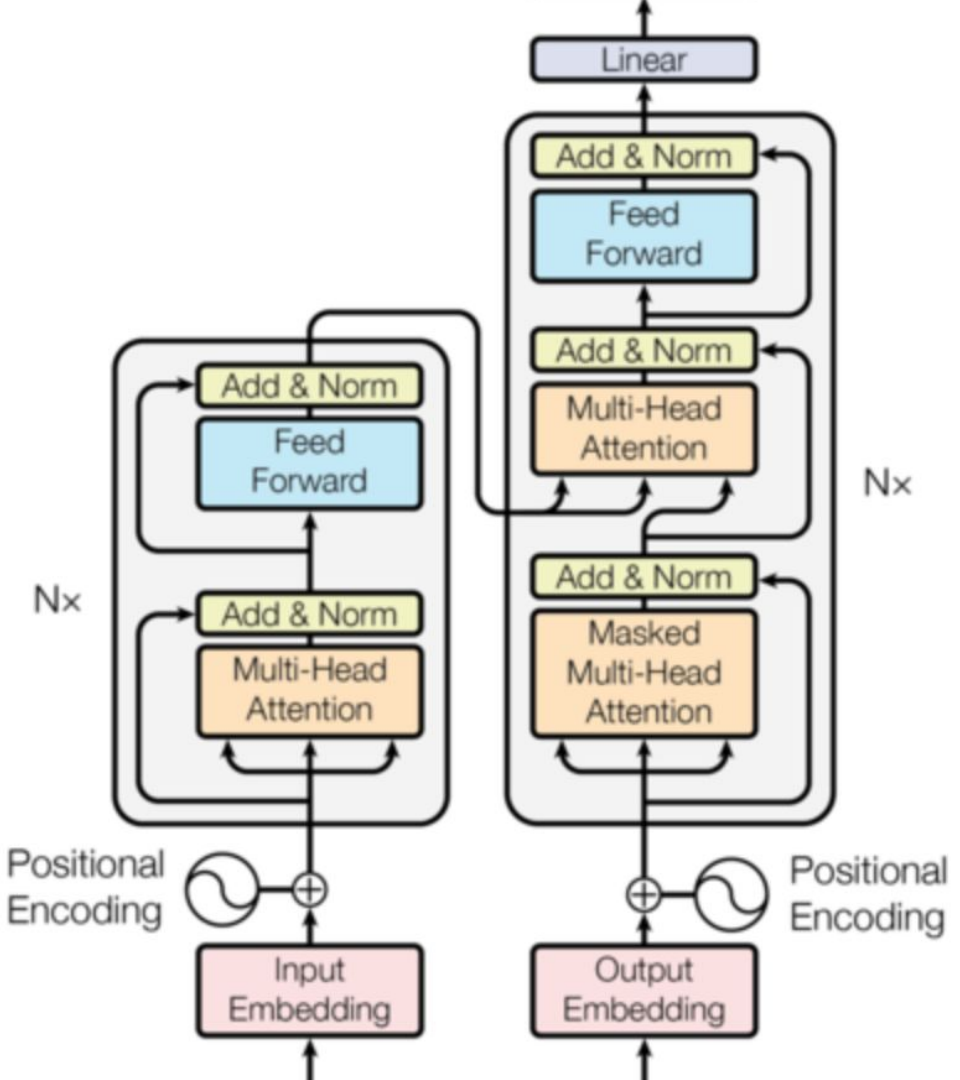
What are transformers?



What are transformers?



What are transformers?



Tied transformers?

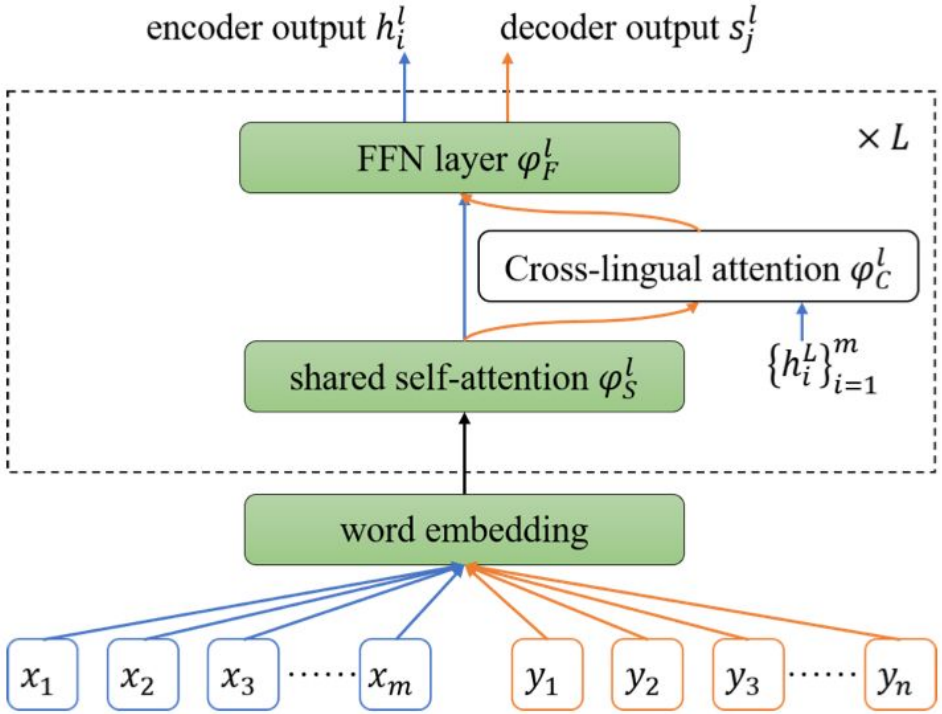
Sharing weights between encoders and decoders has been used in the past. e.g. tied autoencoders.

In MT => both languages follow similar semantics - hence sharing components between the two networks may not be a bad idea, even though lexical differences exist.

Sharing the self-attention component:

> For example, if looking at “The Cat” then decoder should also look at the same concepts in the target language.

Tied transformers?



Tied transformers

Let’s discuss datasets and their paper in detail ...