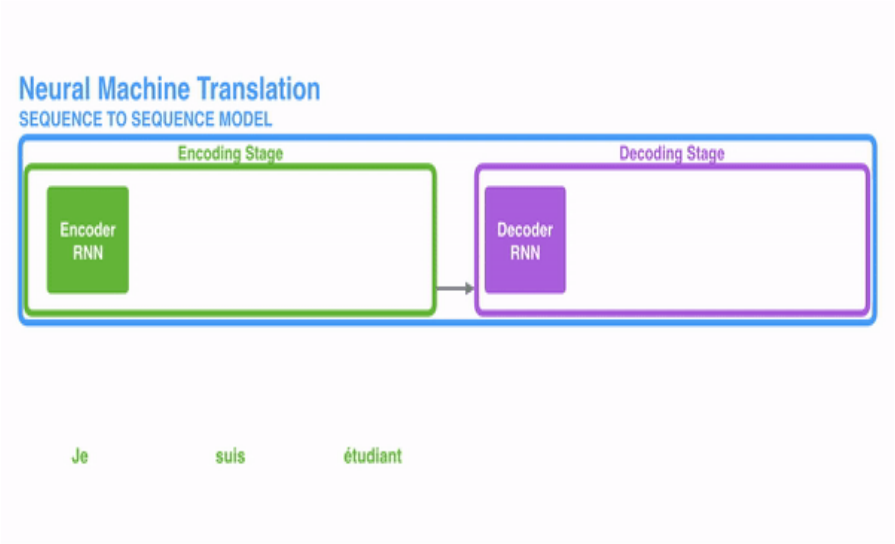
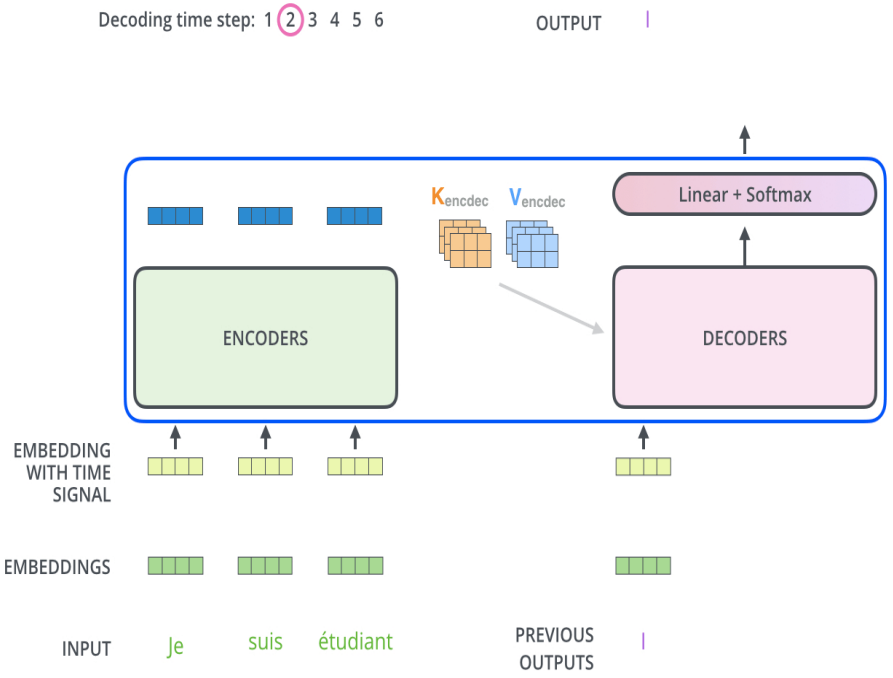
Transformer Networks vs RNN / LSTM

RNN & LSTM:



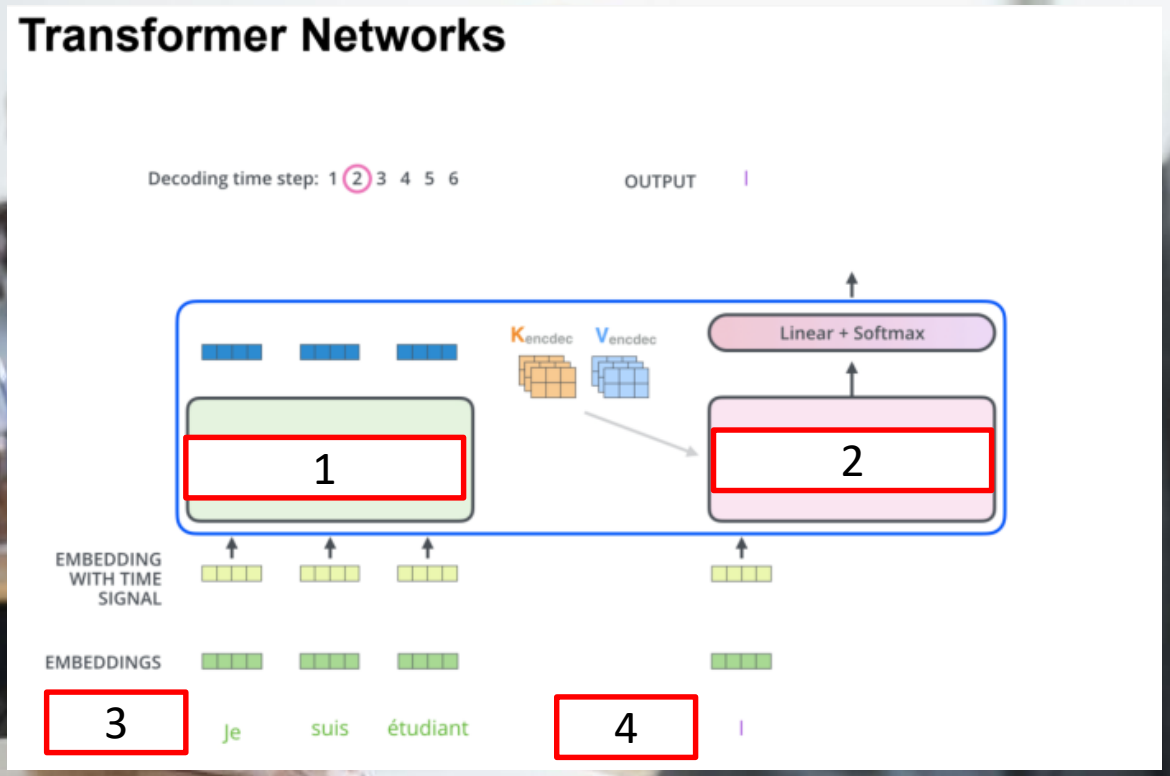
Transformer:



The Illustrated Transformer,
jalammar.github.io/illustrated-transformer/

Fill in the Blank

Art des Wissens / Schwierigkeitsgrad	Abfragewissen (Vorlesung)	Anwendungswissen (Literatur)
Einfach		
Mittel		
Schwierig		



a) Decoder | Encoder | Output|Input
b) Decoder | Encoder | Input |Previous Output
c) Encoder| Decoder | Output|Input
d) Encoder| Decoder | Input |Previous Output
e) None

4. Transformer Networks

Transformer Networks (machine learning model)

Transformer (machine learning model)

From Wikipedia, the free encyclopedia

The **Transformer** is a deep learning model introduced in 2017, used primarily in the field of natural language processing (NLP).[1]

Like recurrent neural networks (RNNs), Transformers are designed to handle sequential data, such as natural language, for tasks such as translation and text summarization. However, unlike RNNs, Transformers do not require that the sequential data be processed in order. For example, if the input data is a natural language sentence, the Transformer does not need to process the beginning of it before the end. Due to this feature, the Transformer allows for much more parallelization than RNNs and therefore reduced training times.[1]

Transformers have rapidly become the model of choice for NLP problems,[2] replacing older recurrent neural network models such as the long short-term memory (LSTM). Since the Transformer model facilitates more parallelization during training, it has enabled training on larger datasets than was possible before it was introduced. This has led to the development of pretrained systems such as BERT (Bidirectional Encoder Representations from Transformers) and GPT (Generative Pre-trained Transformer), which have been trained with huge general language datasets, such as Wikipedia Corpus, and can be fine-tuned to specific language tasks.[3][4]