Introduction to Seq2Seq Models

t.me/cvision

Abolfazl Mahdizade

Outline

- *Introduction and Concepts*
- *High Level View*
- *Seq2Seq Different Views*
- *What is Embedding*
- *Why Seq2Seq Works*
- *Seq2Seq Training*
- *Seq2Seq with TensorFlow*

Sequence to Sequence Model: Introduction and Concepts

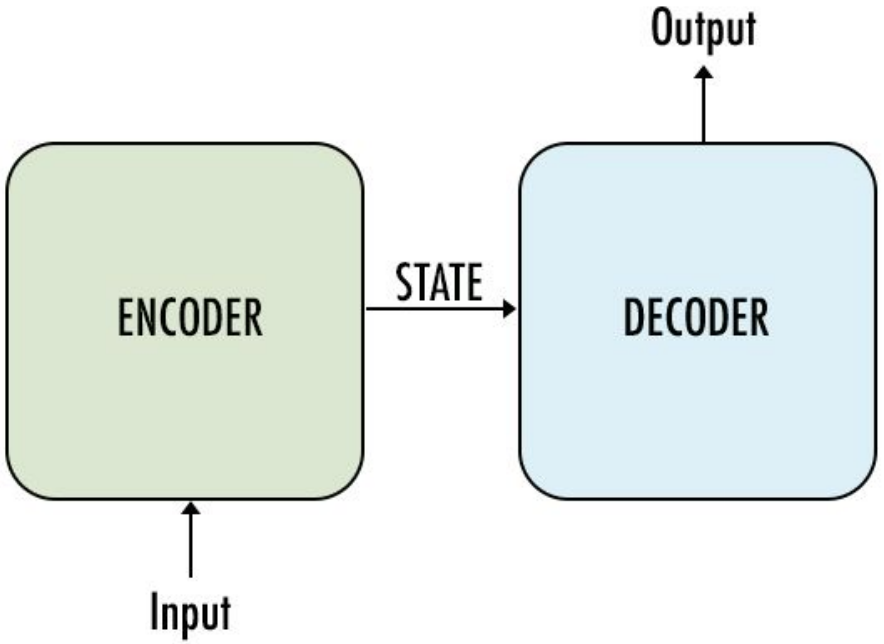
https://goo.gl/4tky9Q

https://indico.io/blog/sequence-modeling-neuralnets-part1

https://github.com/tensorflow/nmt

High Level View

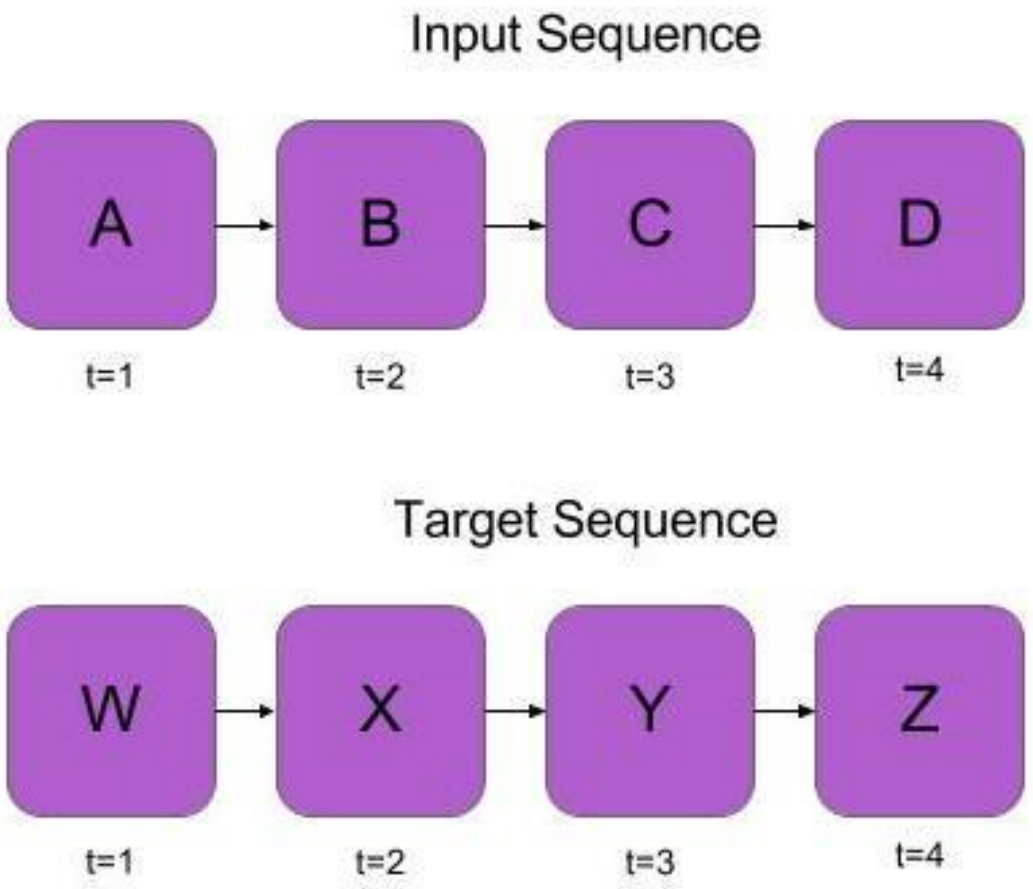
- A seq2seq model has encoder, decoder and intermediate step as its main components



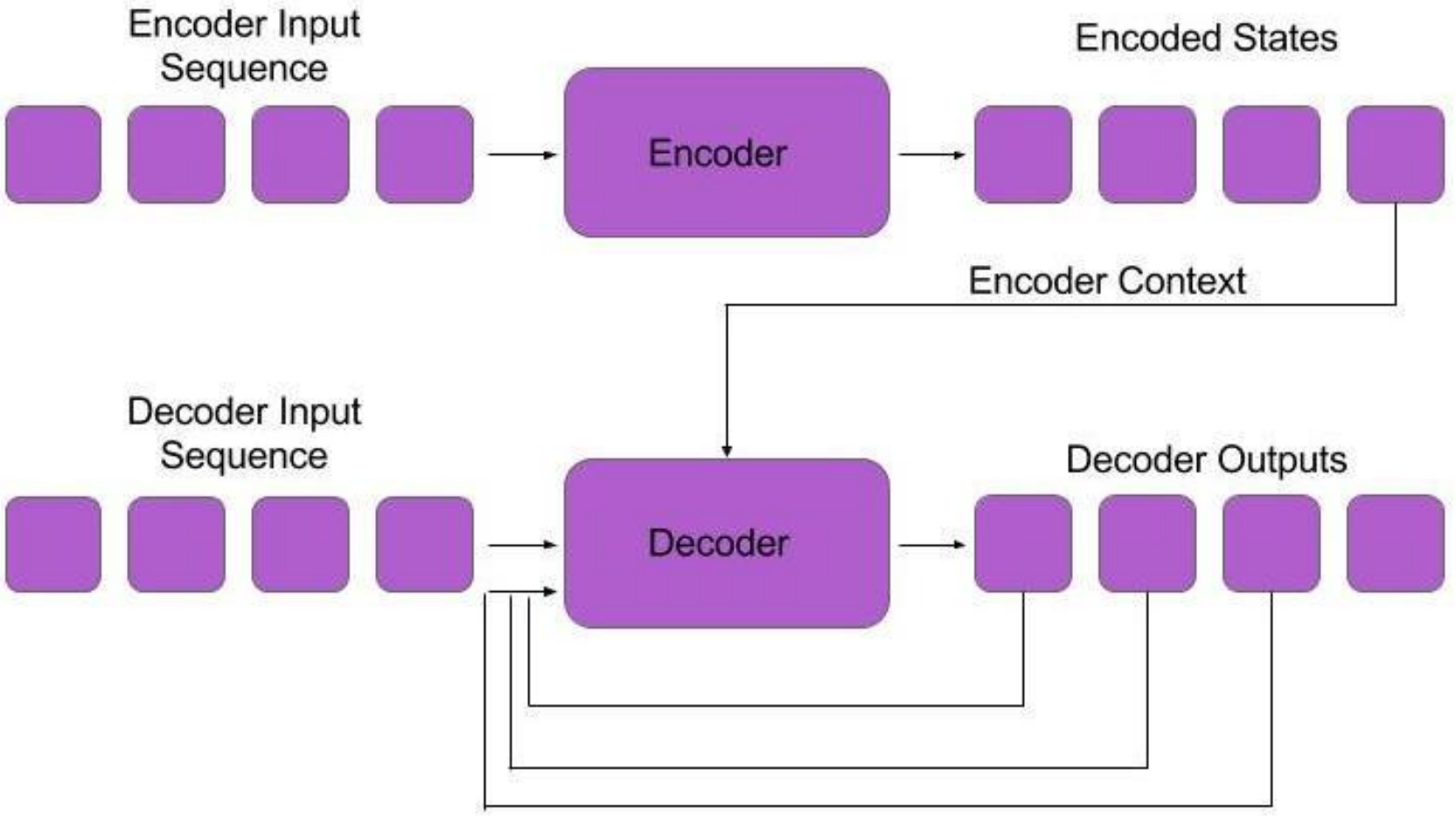
High Level View (Continue)

- *Encoder*
  - *A model (LSTM) converts an input sequence (such as an English sentence) into a fixed representation (called Context Vector)*
- *Decoder*
  - *A model (LSTM) is trained on both the output sequence (such as the translated sentence) as well as the fixed representation from the encoder*

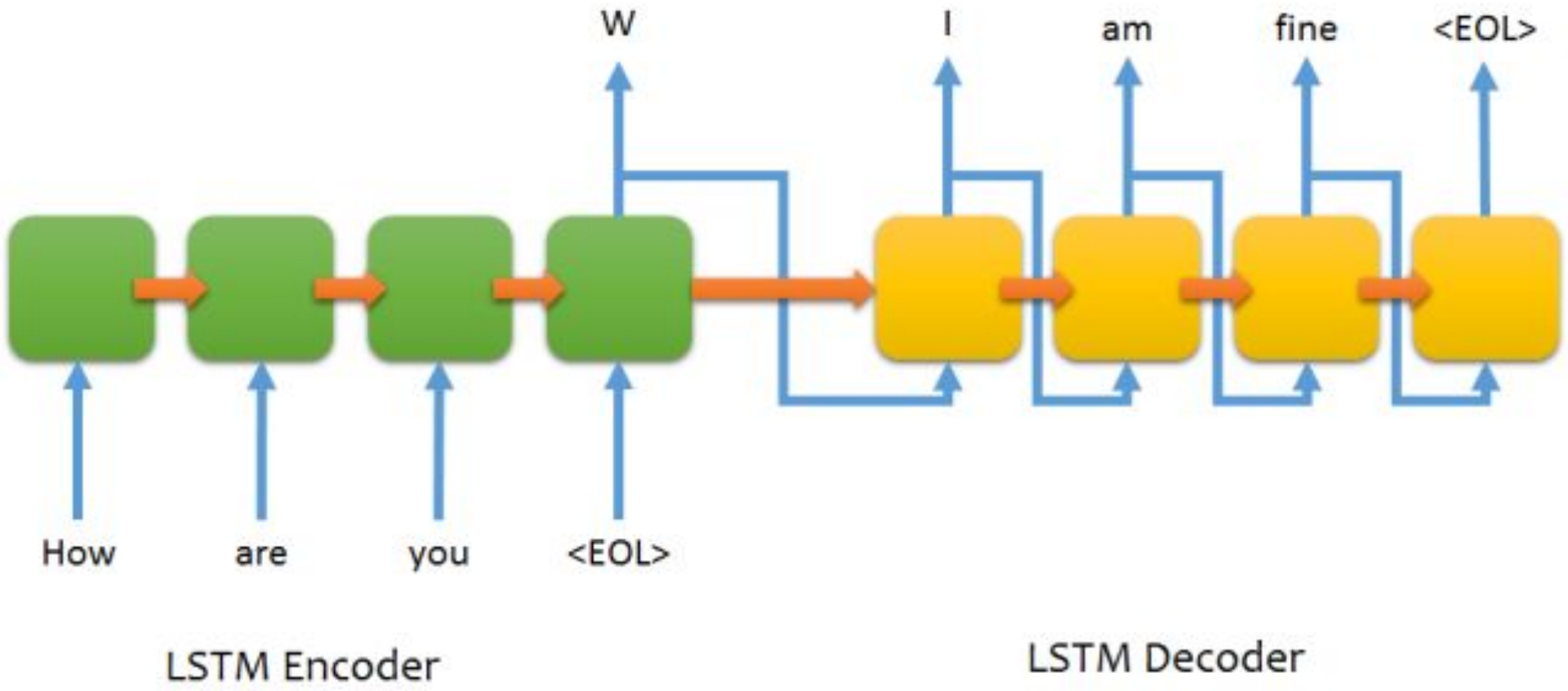
Seq2Seq Different Views



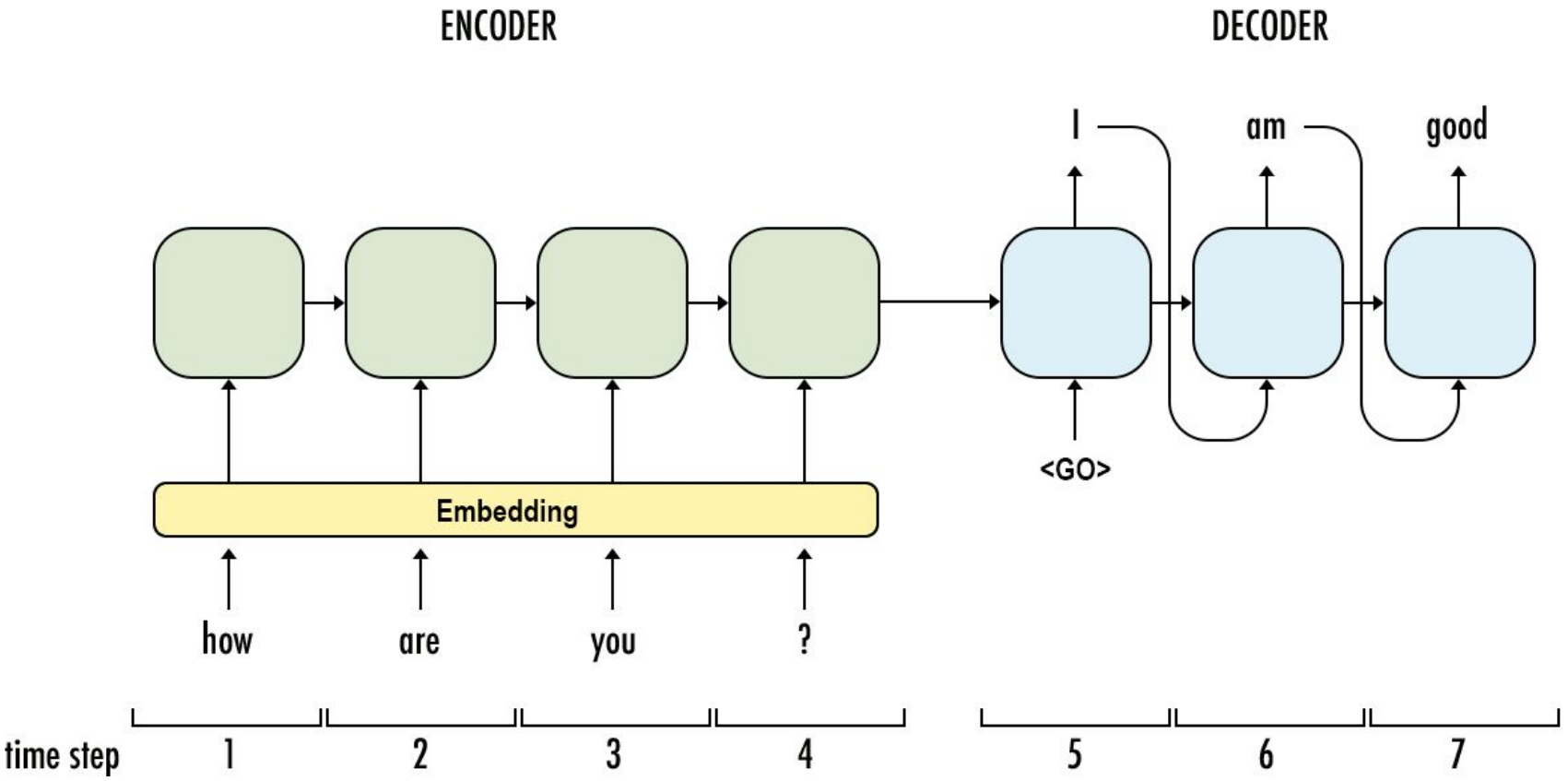
Seq2Seq Different Views (Continue)



Seq2Seq Different Views (Continue)



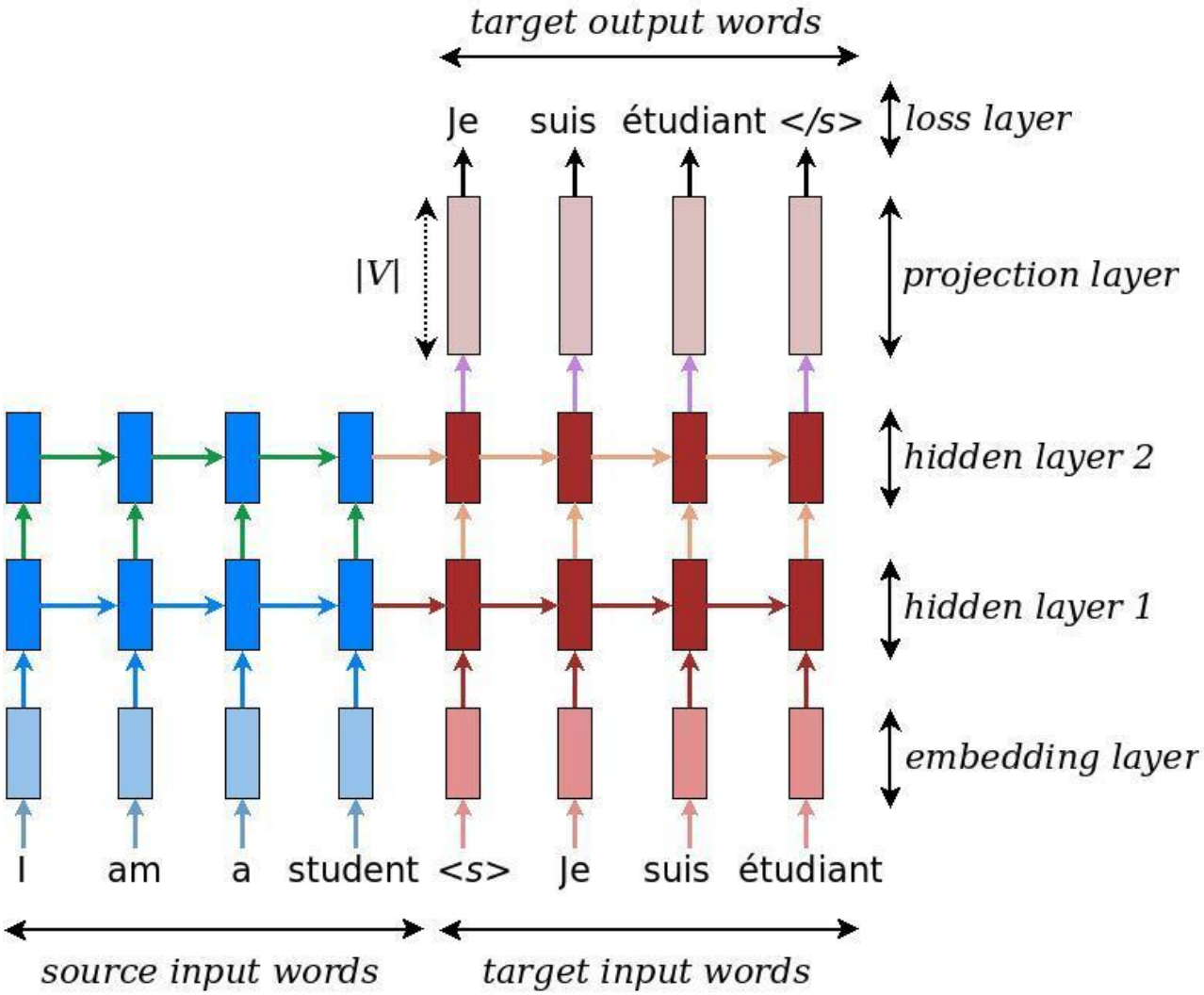
Seq2Seq Different Views (Continue)



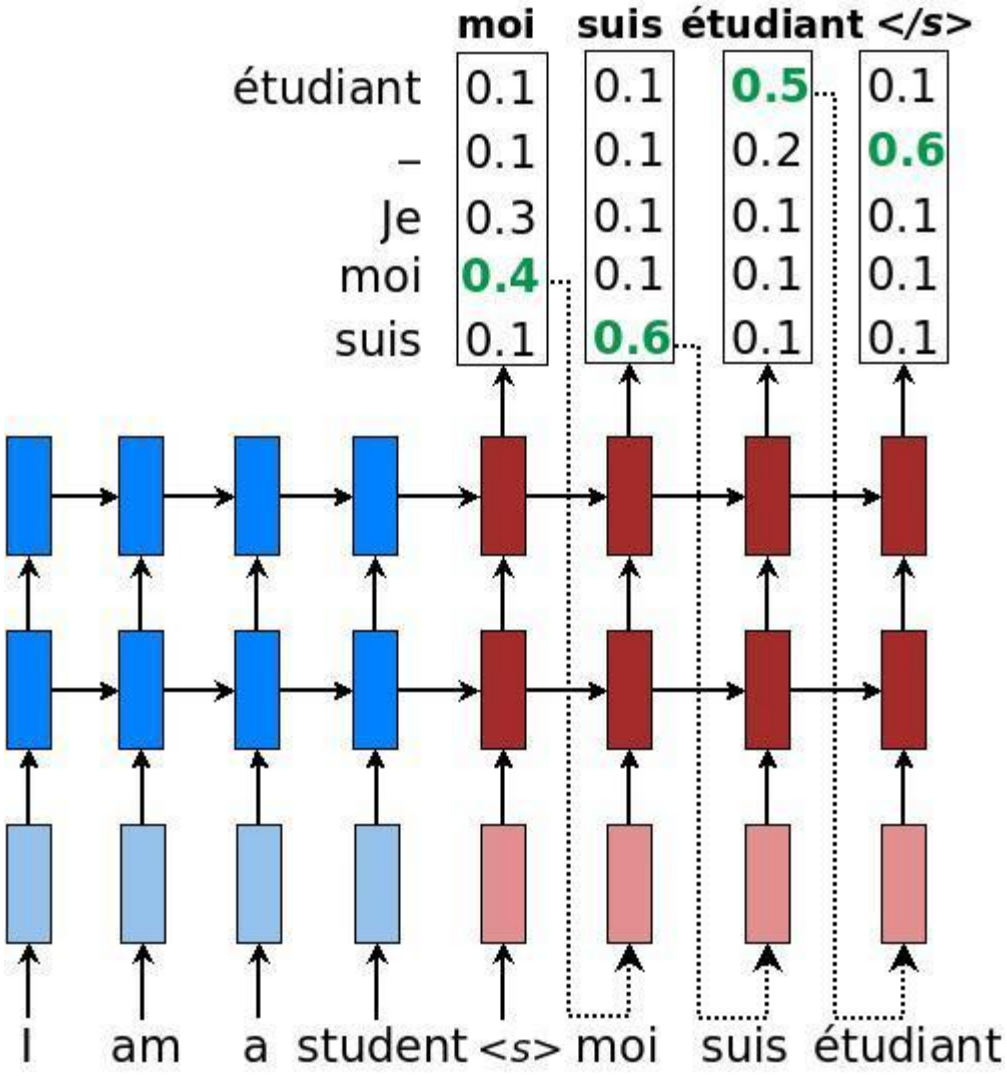
What is Embedding

- *Similar to Word2Vec*
  - *Dense Representation of Words in Vector Space*
- *Pretrain Embeddings*
  - *Word2Vec*
  - *GloVe*
- *Embedding Layer*
  - *Lookup table (2D weights)*

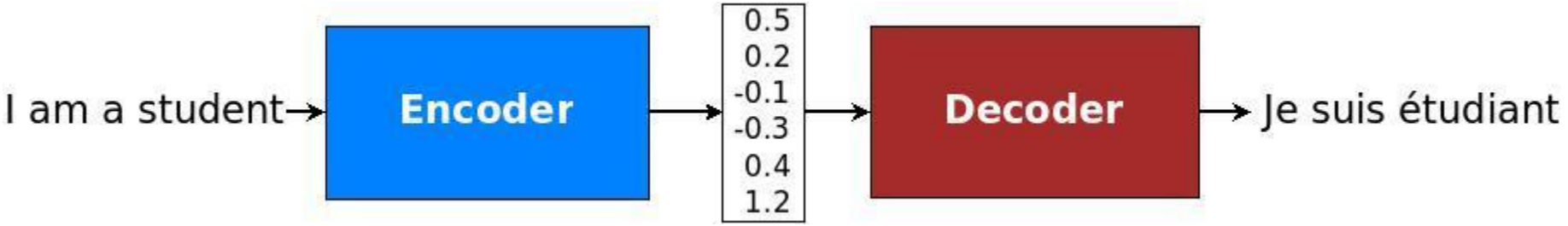
Seq2Seq Different Views (Continue)

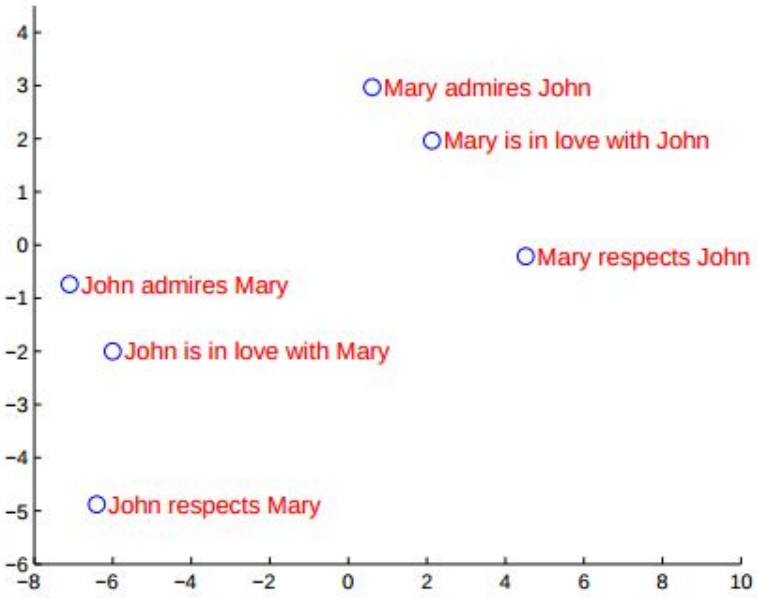


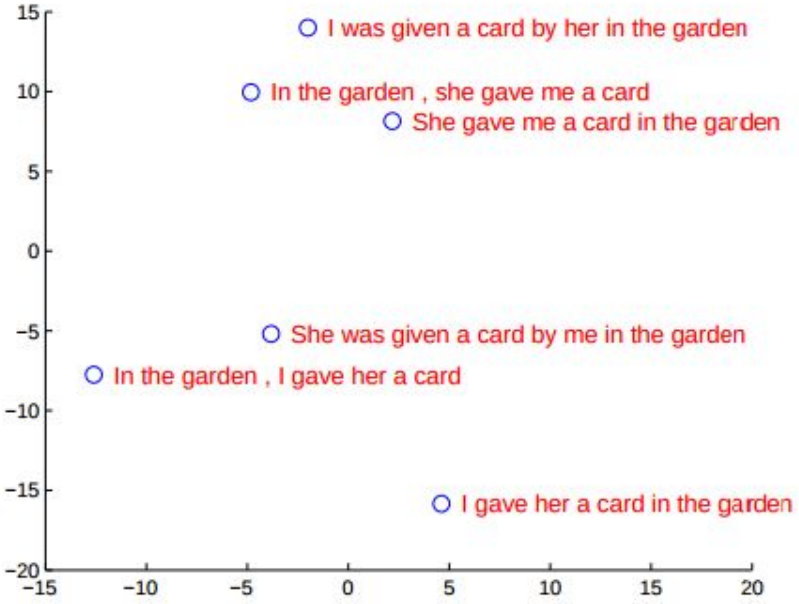
Seq2Seq Different Views (Continue)



Why Seq2Seq Works





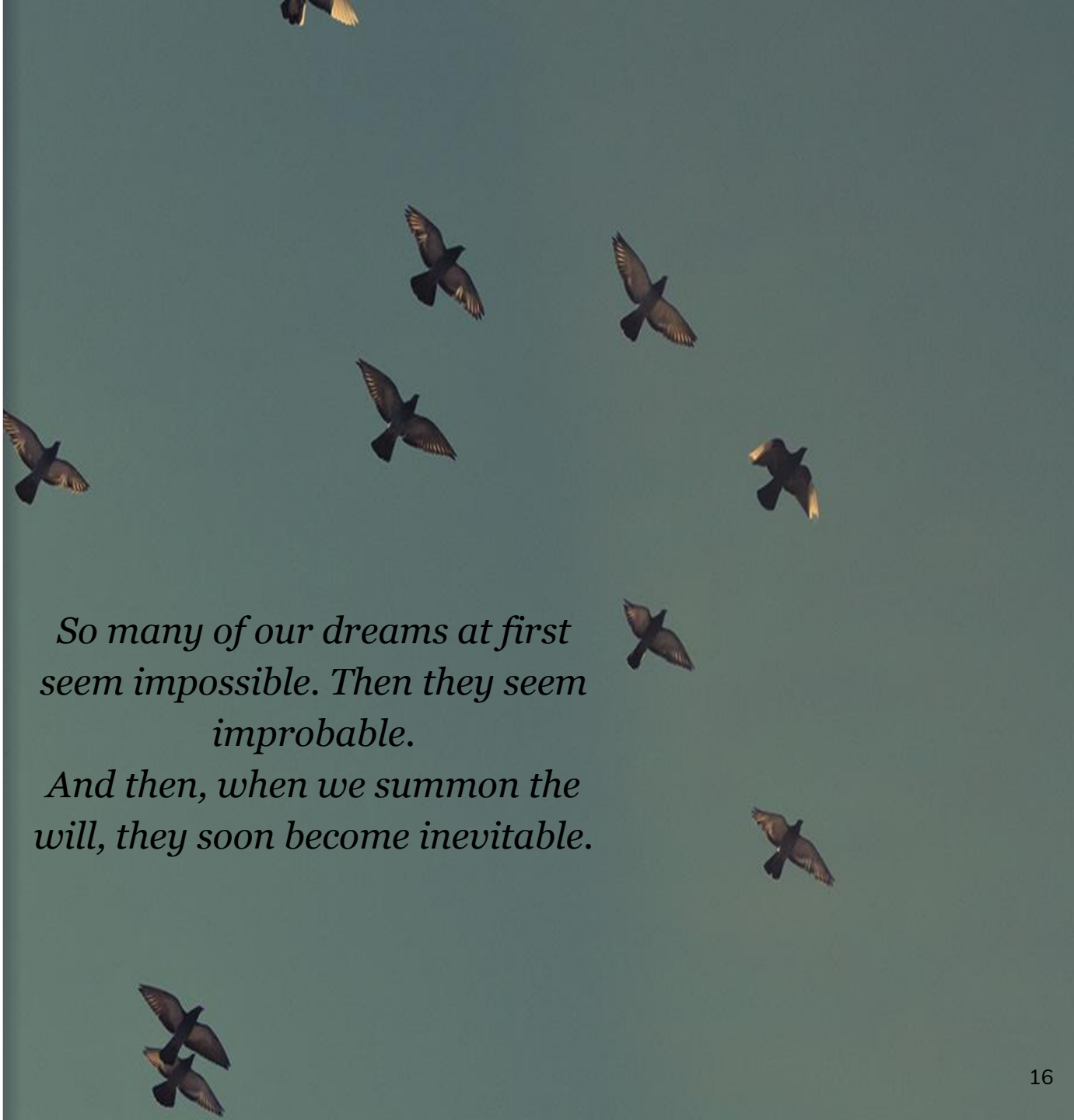


Seq2Seq Training

- *Inputs*
  - *A tensor contains IDs of the words in the sequence*
  - *IDs (numbers, 0, 1, 2, …) stands for one hot vectors*
- *Some specific symbols (Spans)*
  - <PAD> = 0
  - <EOS> = 1
  - <UNK> = 2

Seq2Seq Training (Continue)

- *<PAD>*
  - *Feed inputs in batch*
  - *The inputs in these batches all need to be the same width for the network to do its calculation*
  - *However examples are not of the same length*
  - *You should pad shorter inputs to bring them to the same width of the batch*
- *<EOS>*
  - *It is important on the decoder side*
  - *It tells the decoder where a sentence ends, and it allows the decoder to indicate the same thing in its outputs as well*
- *<UNK>*
  - *Use for unknown words*
  - *Unknown words = rare words in the corpora*


So many of our dreams at first seem impossible. Then they seem improbable.
And then, when we summon the will, they soon become inevitable.