Introduction to Transformer

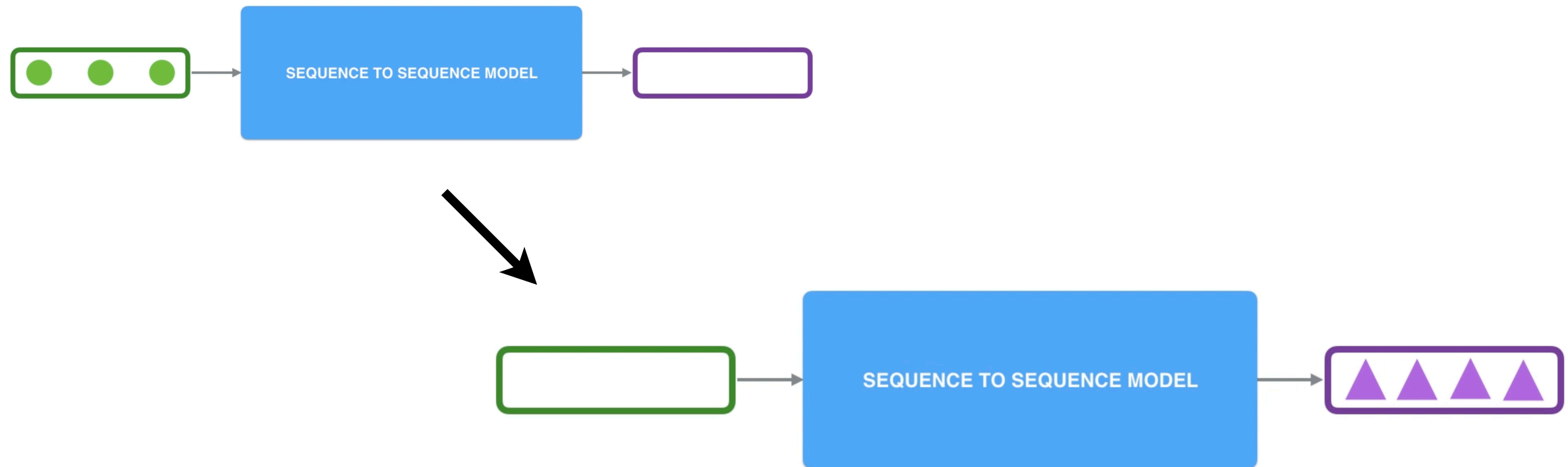Sanmay Ganguly
(sanmay.ganguly@cern.ch)

Statistical Methods and Machine Learning in High Energy Physics, ICTS 2023

28/08/2023

Sequence to Sequence mapping

A sequence is an ordered set, can be words, letters, images, ... any data-structure .. but the positional order has to have a right Human interpretable meaning.
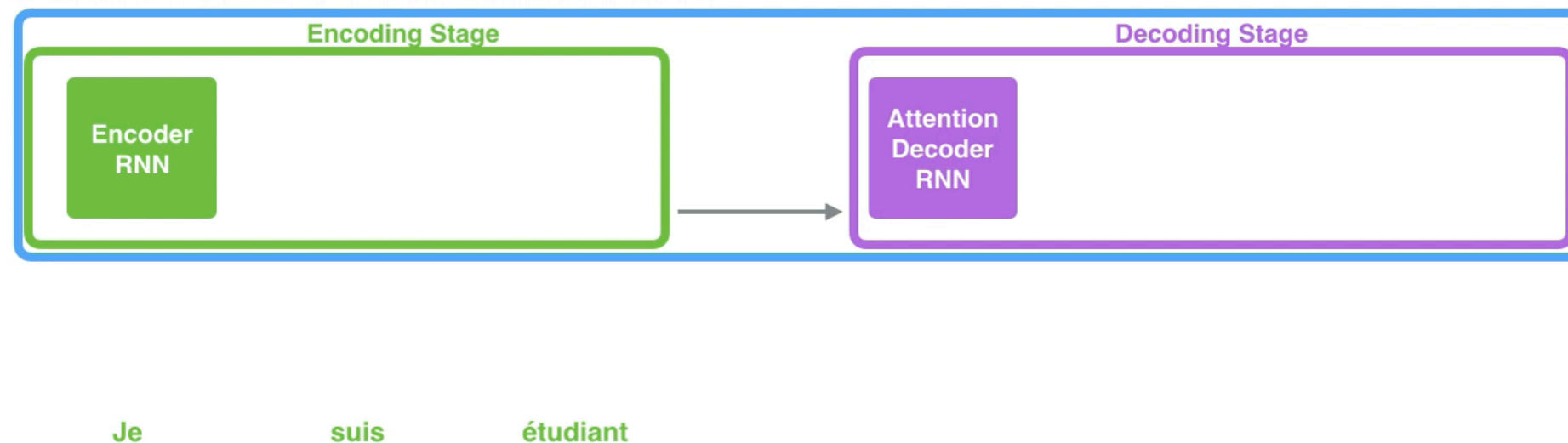

For a neural machine translation, in NLP context this sequence is a sentence with the right ordering of words.

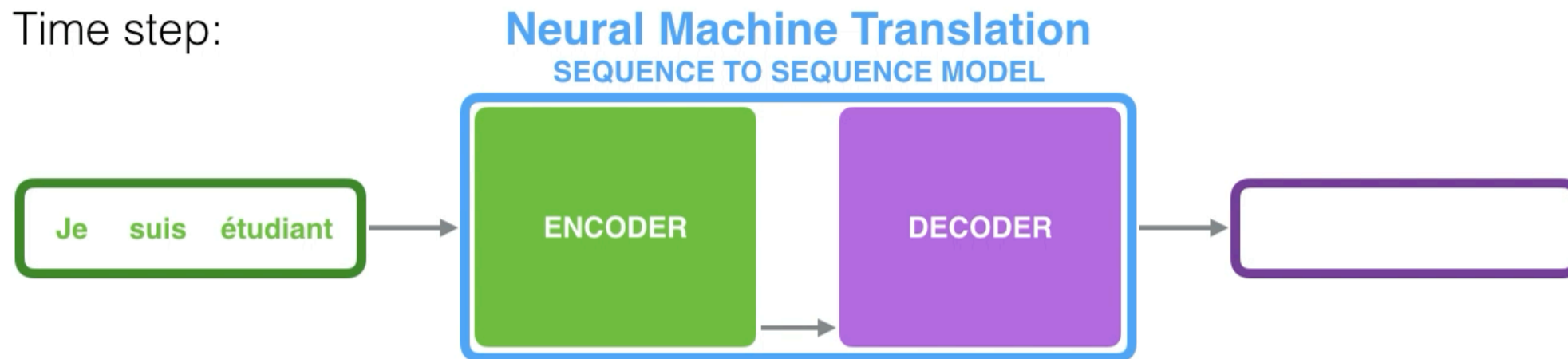# Sequence to Sequence mapping

Time step:

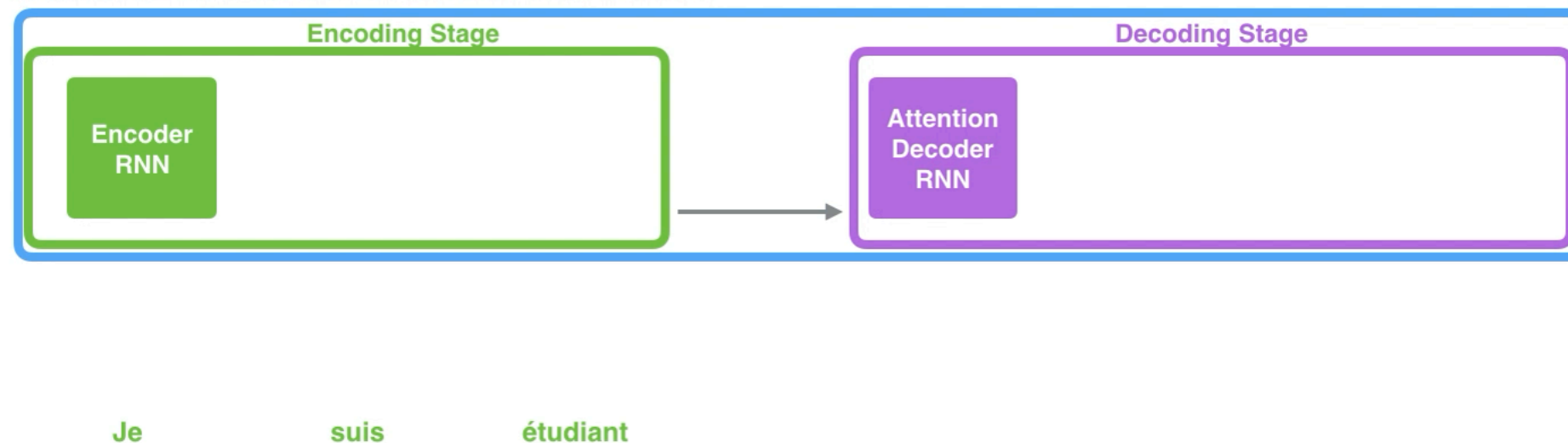**Neural Machine Translation**
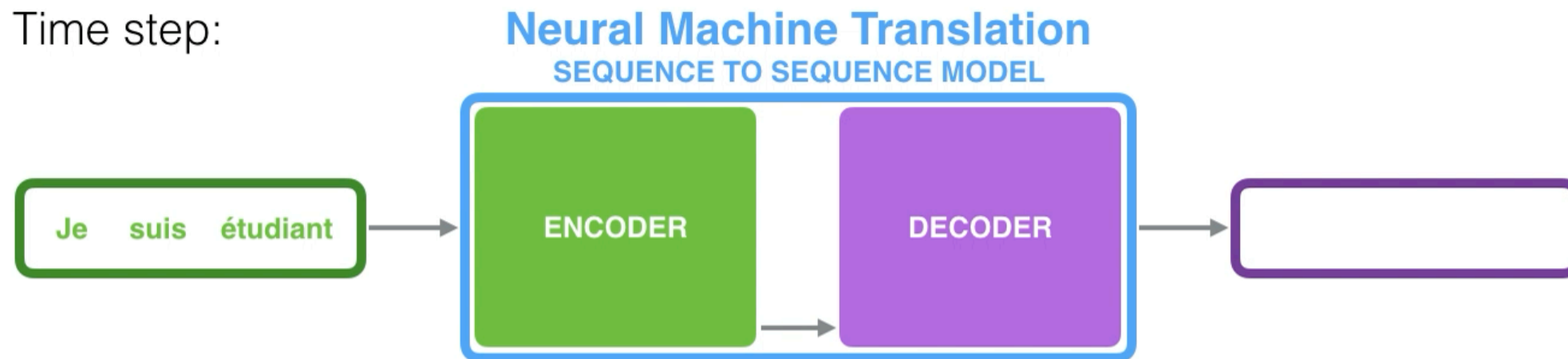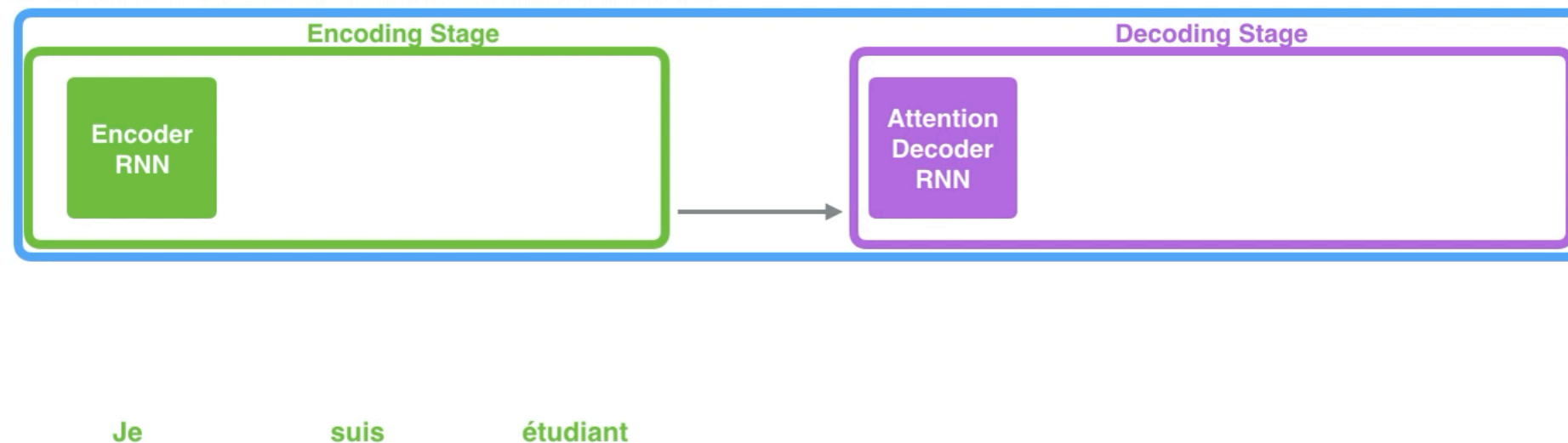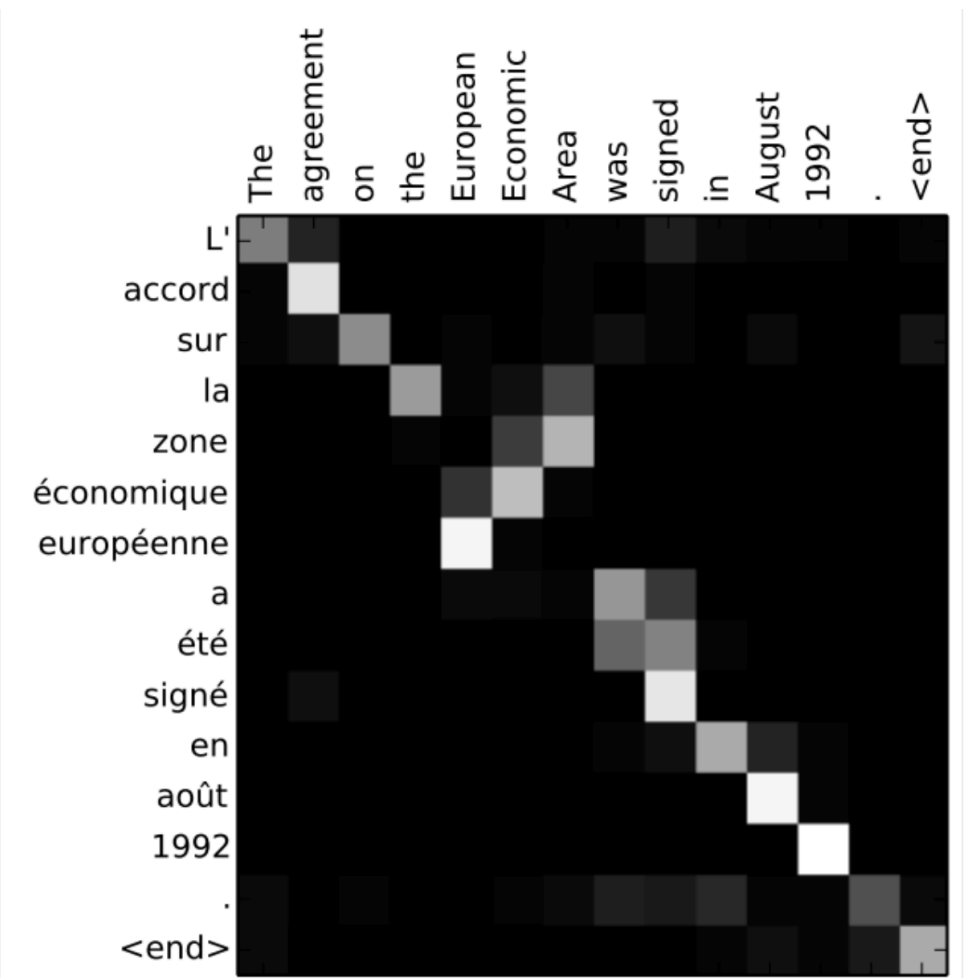SEQUENCE TO SEQUENCE MODEL

Je   suis   étudiant   →   ENCODER   →   DECODER   →
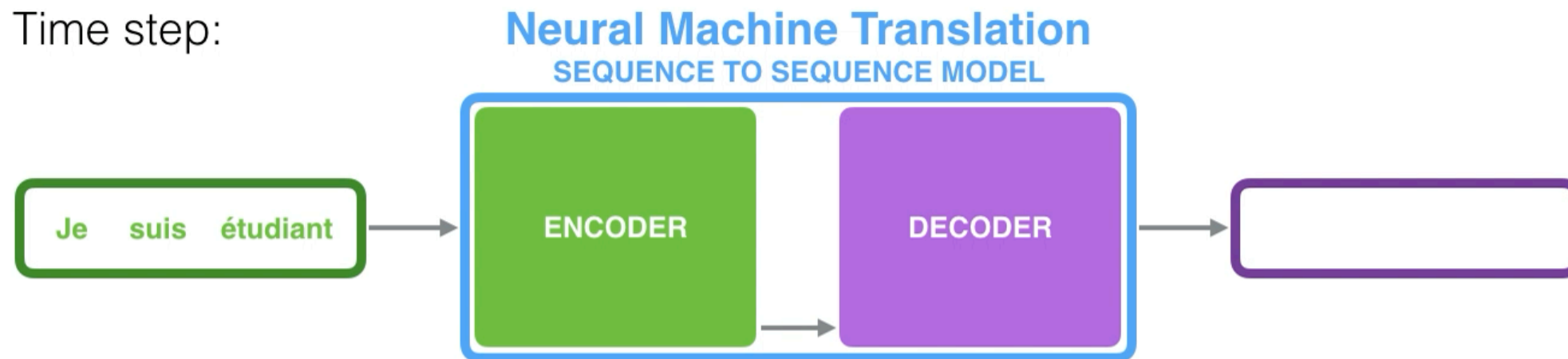
Neural Machine Translation
SEQUENCE TO SEQUENCE MODEL WITH ATTENTION

Encoding Stage   Decoding Stage

Encoder RNN   →   Attention Decoder RNN

Je   suis   étudiant

Sequence to Sequence mapping

Sequence to Sequence mapping

Sequence to Sequence mapping
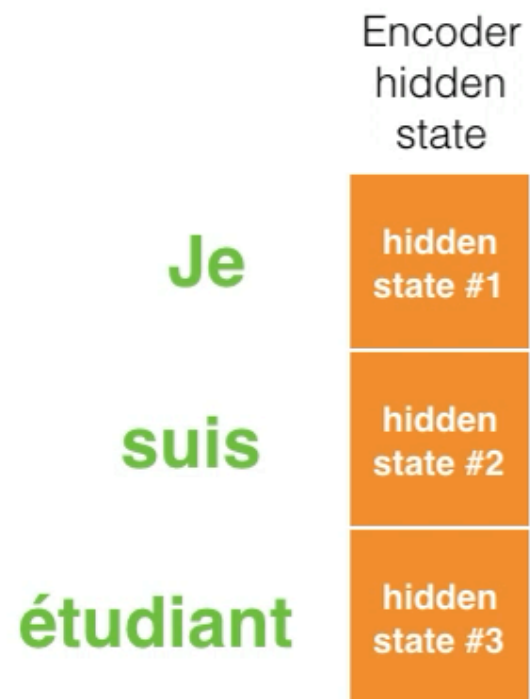
Sequence to Sequence mapping
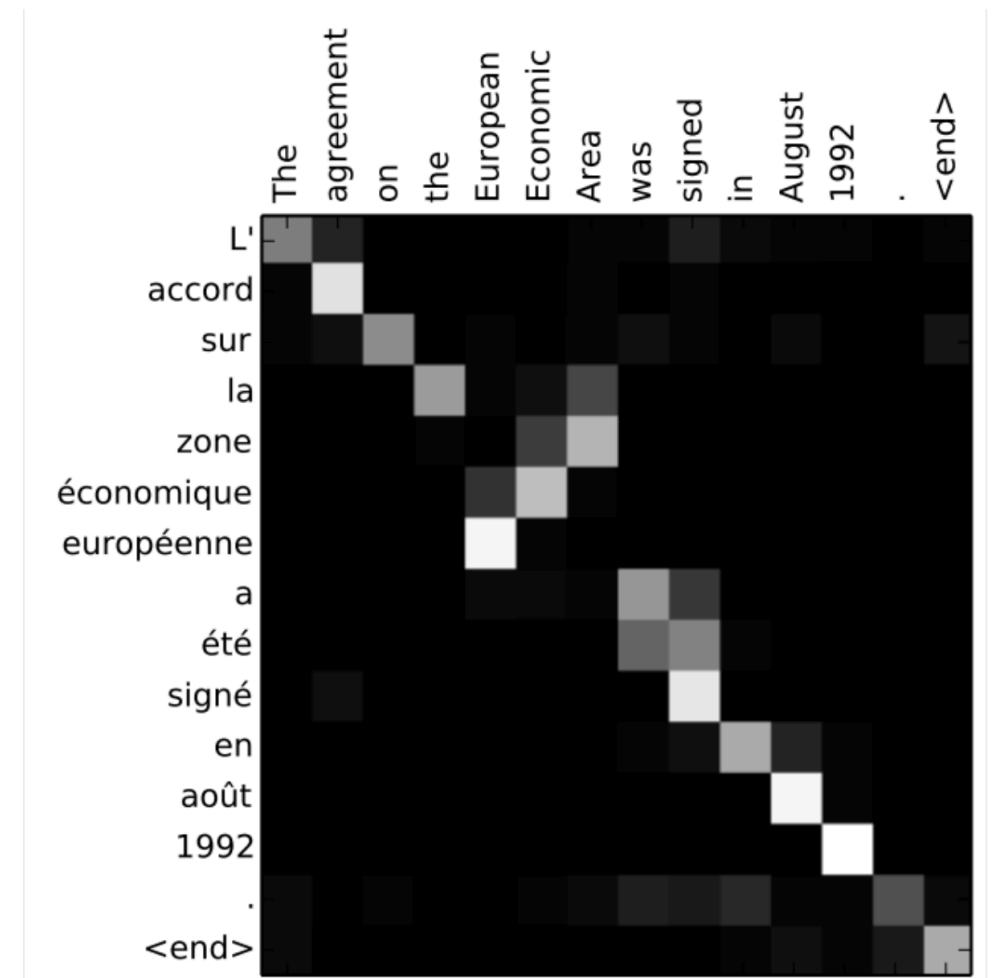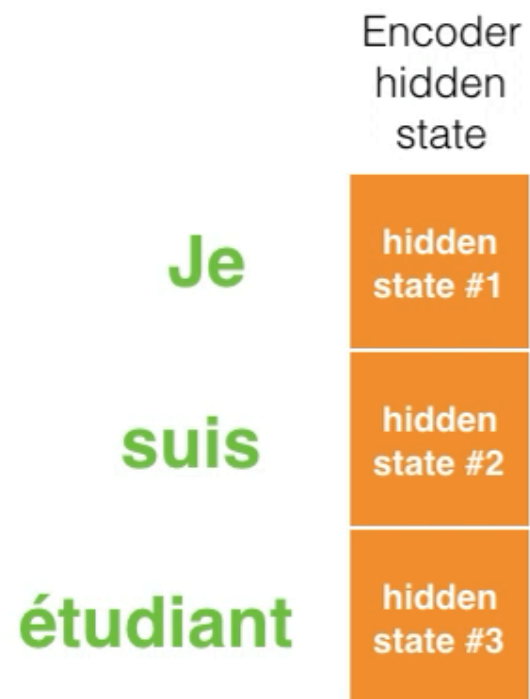
Transformer architecture