Performing Language Translation Using Sequence-to-Sequence Models

Janani Ravi
CO-FOUNDER, LOONYCORN

www.loonycorn.com

Overview

Exploring the different kinds of models built using neural networks

Contrasting seq2seq and seq2vec models

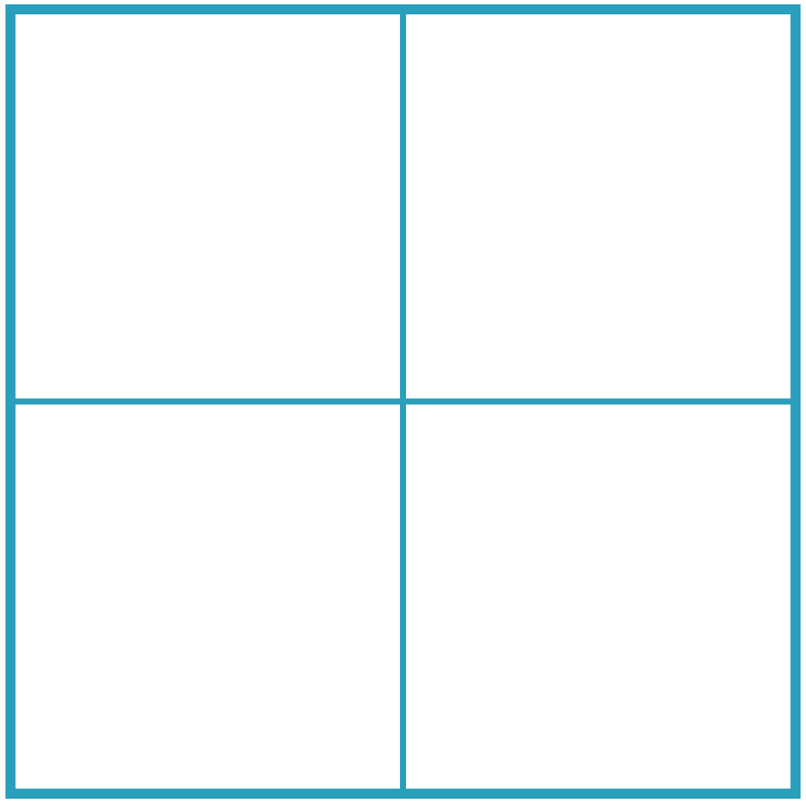
Understanding encoder-decoder architectures for language translation

Implementing language translation using RNNs

Working with Sequences

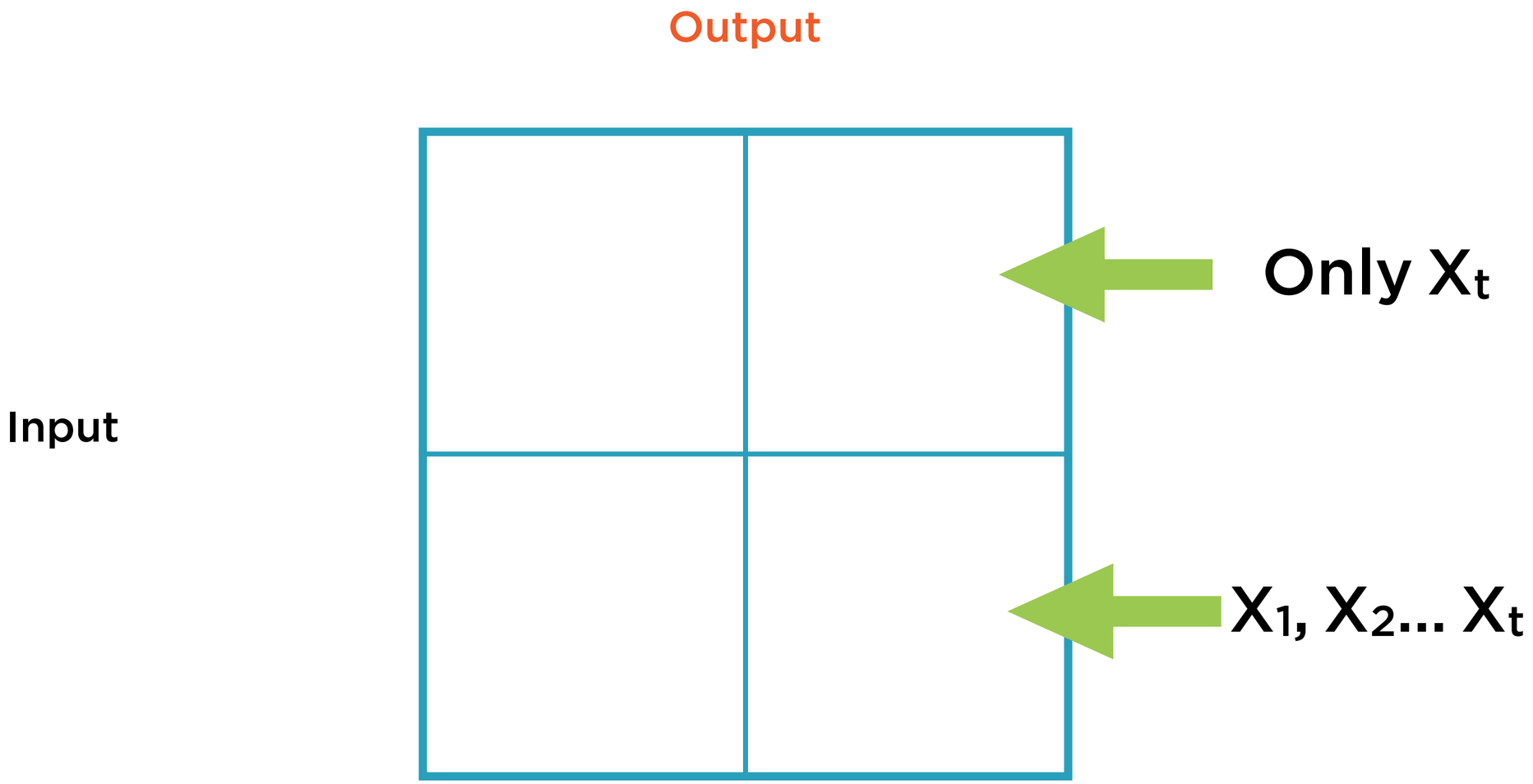
Sequential Data and RNNs

Output



Input

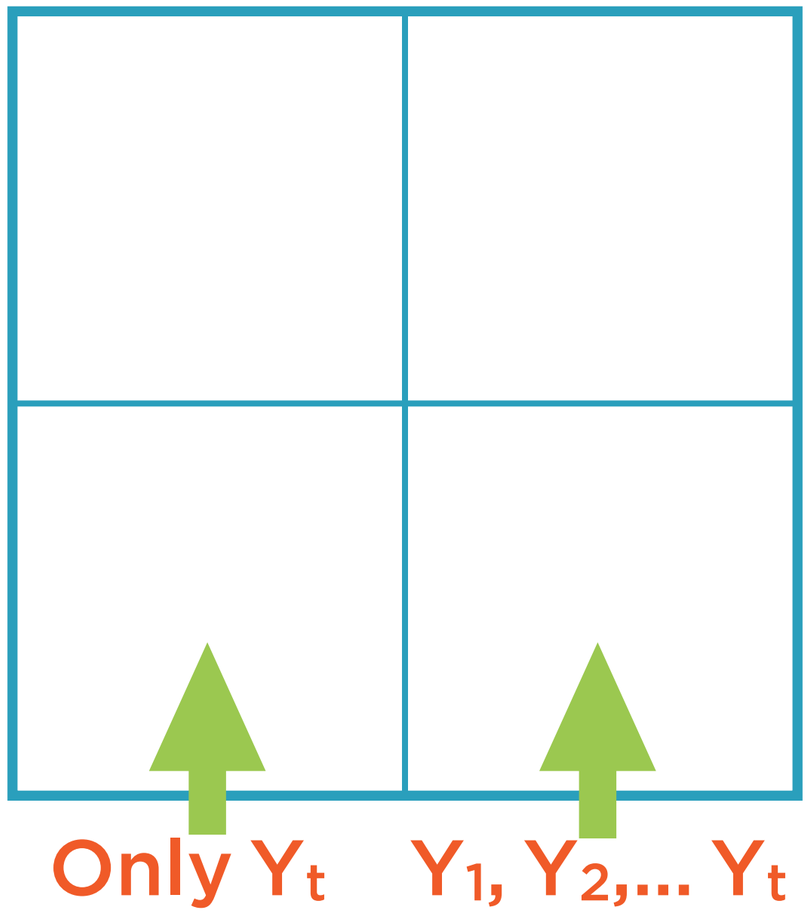
Sequential Data and RNNs



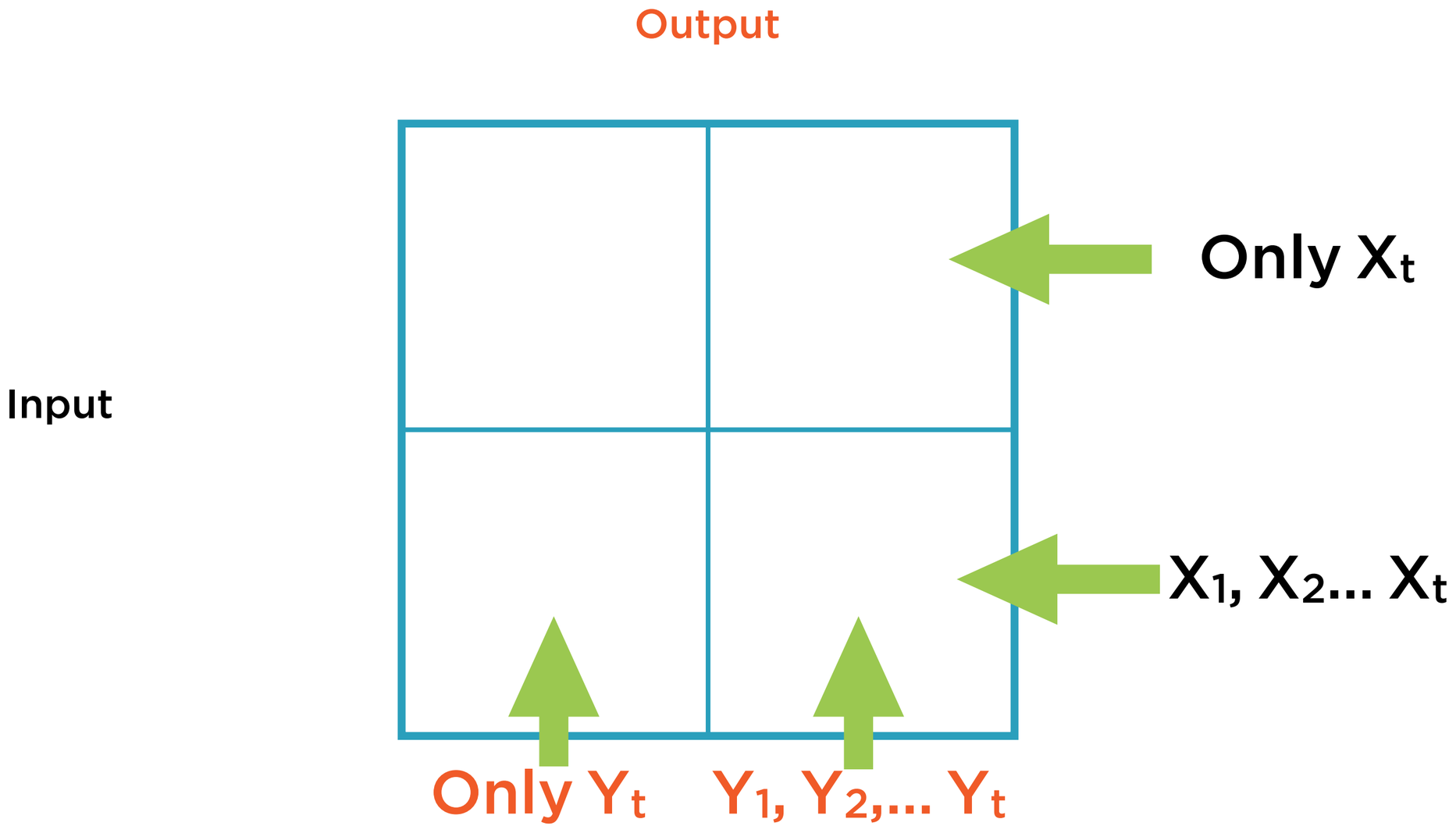
Sequential Data and RNNs

Output

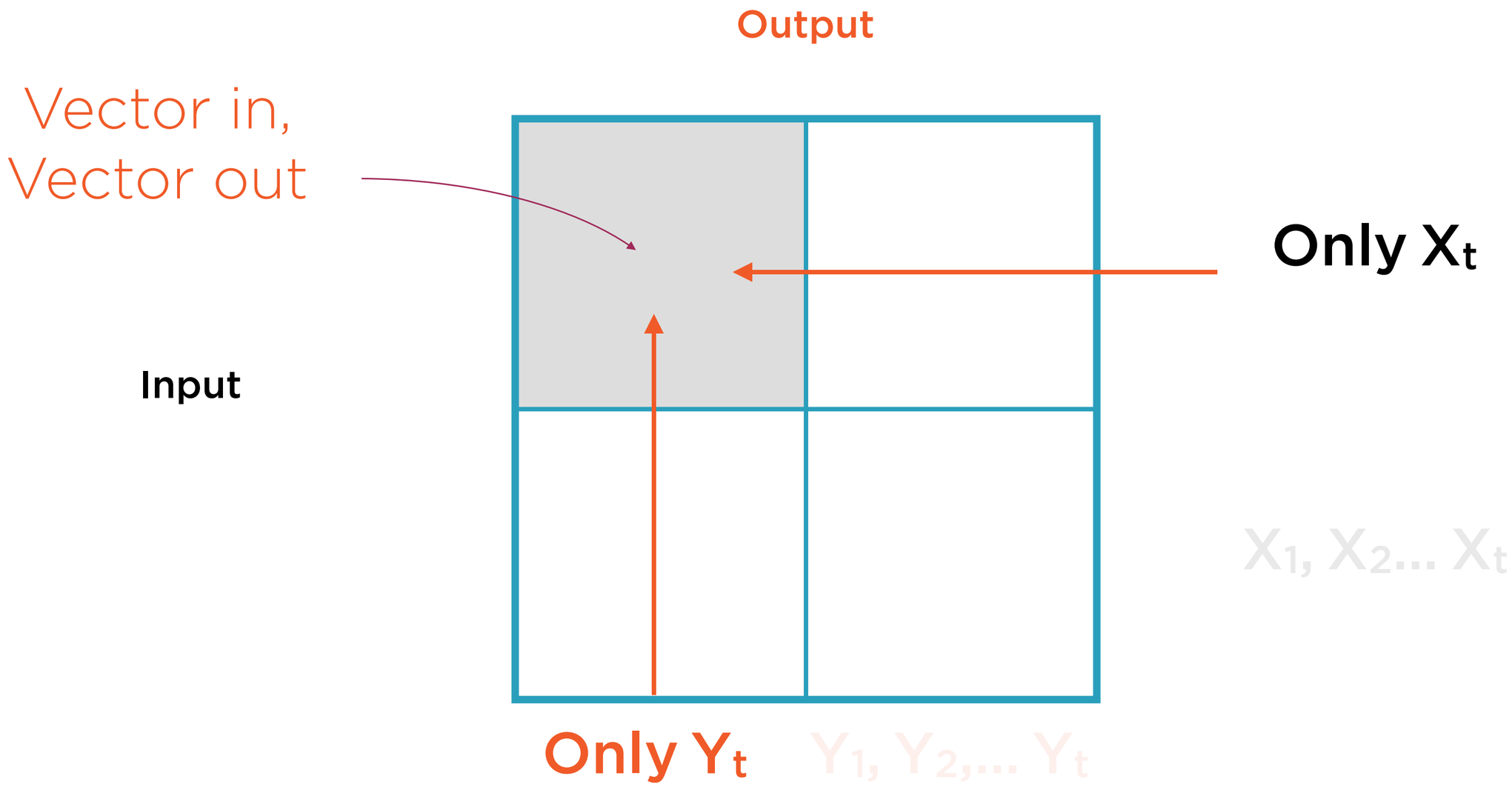
Input



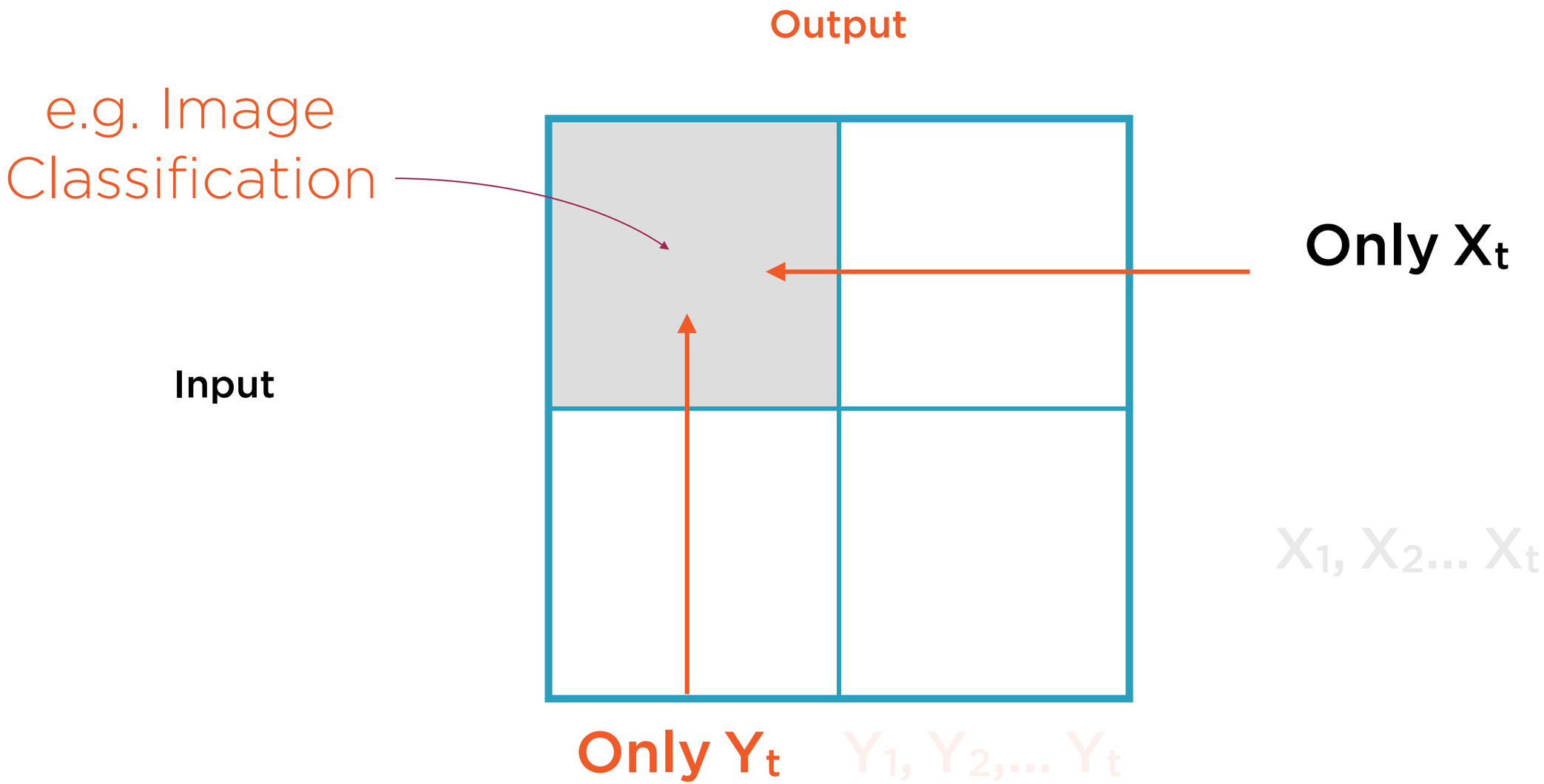
Sequential Data and RNNs



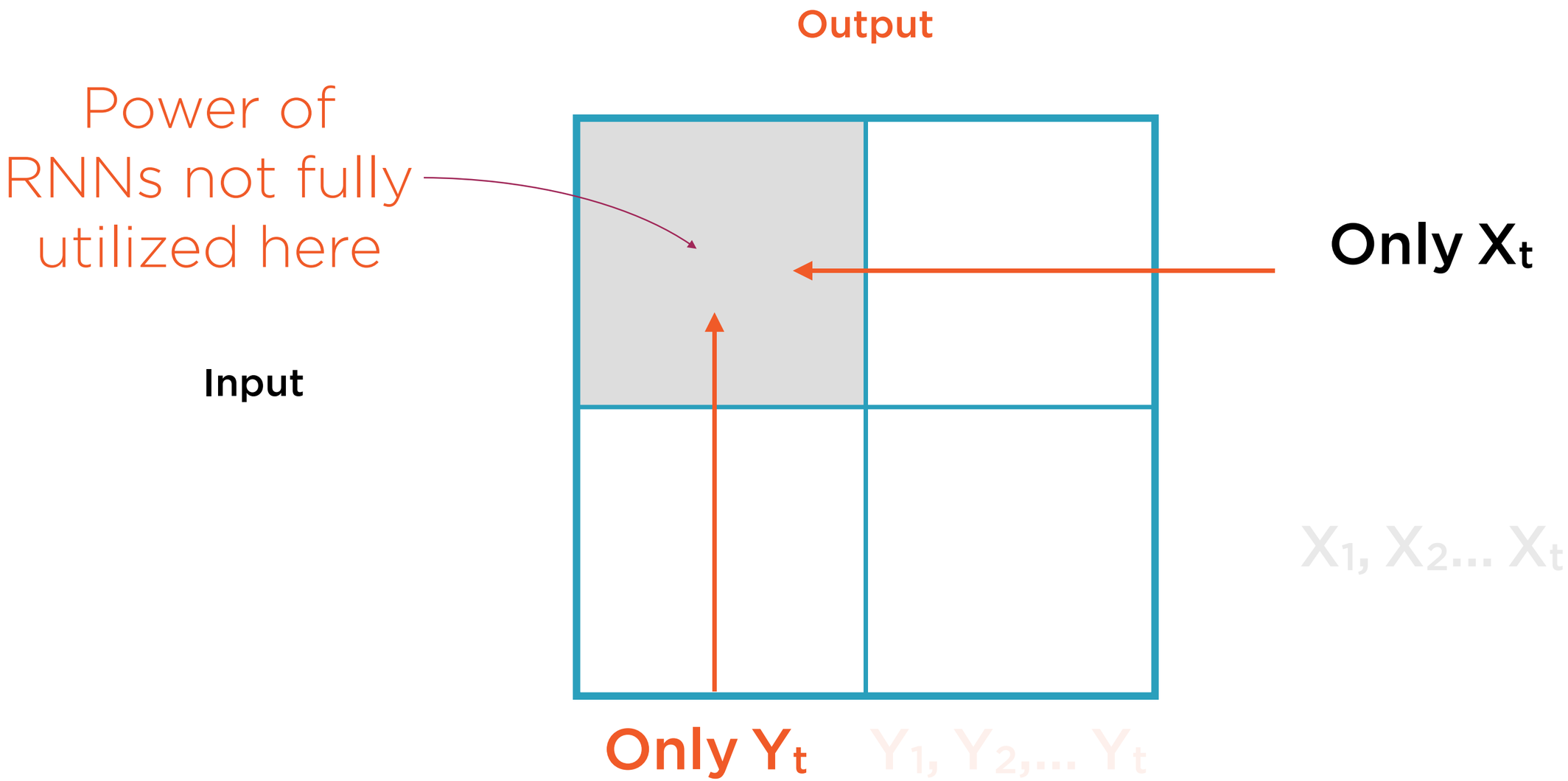
Non-sequential Data and RNNs



Non-sequential Data and RNNs



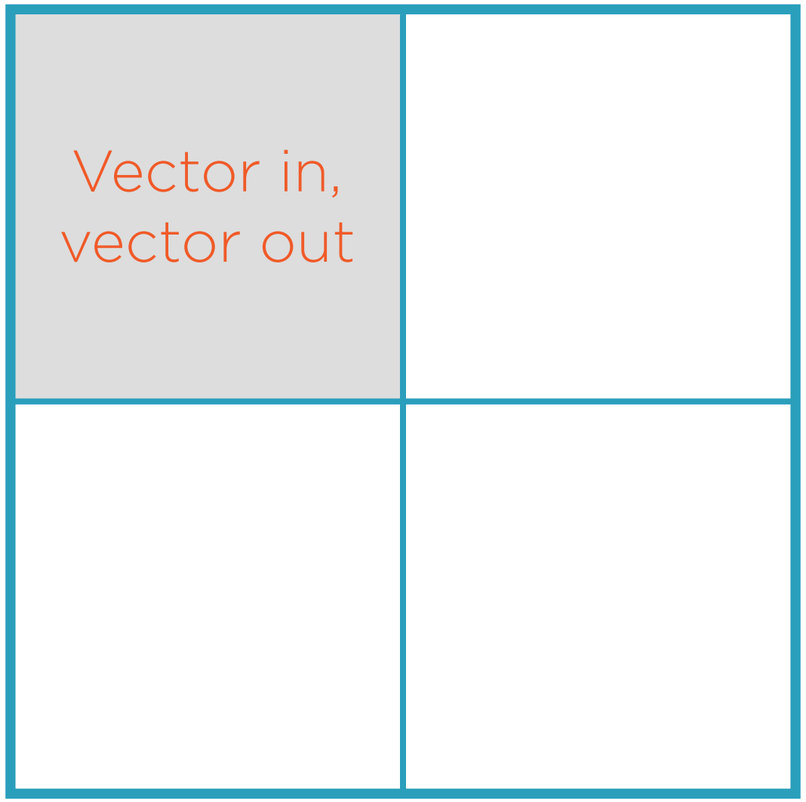
Non-sequential Data and RNNs



Sequential Data and RNNs

Output

Input



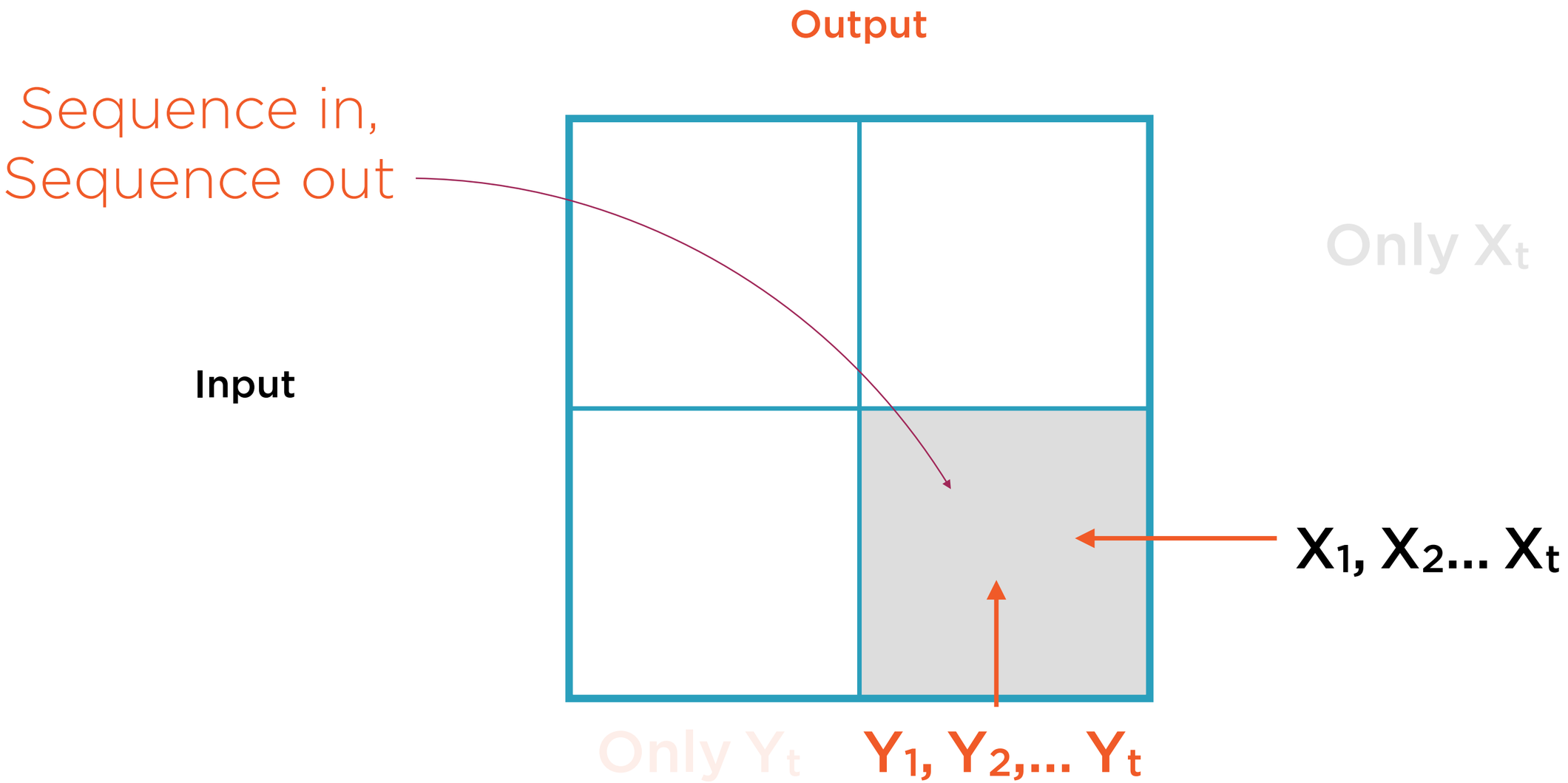
Only X_t

$X_1, X_2 \ldots X_t$

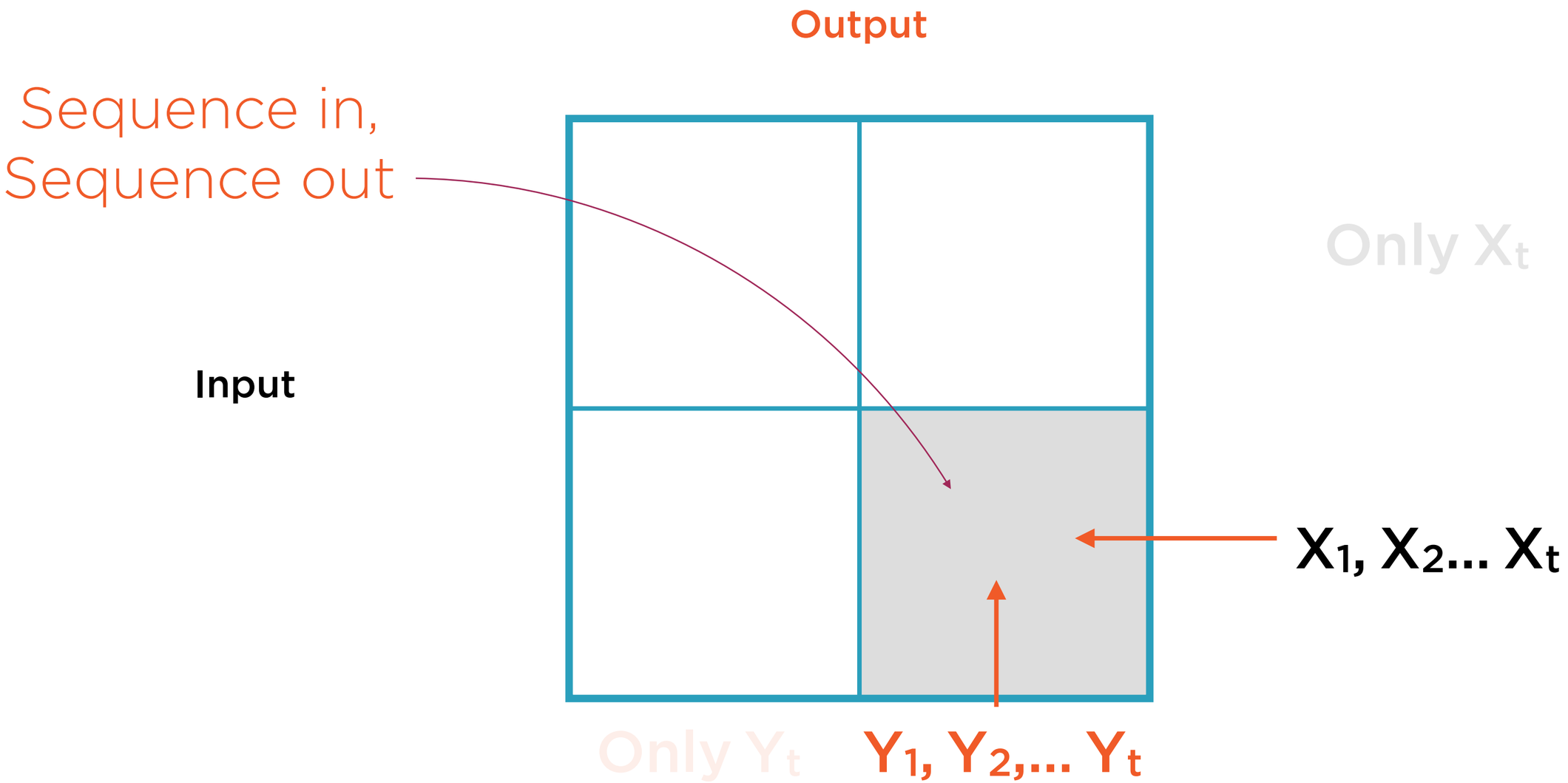
Only Y_t

$Y_1, Y_2, \ldots Y_t$

Sequential Data and RNNs



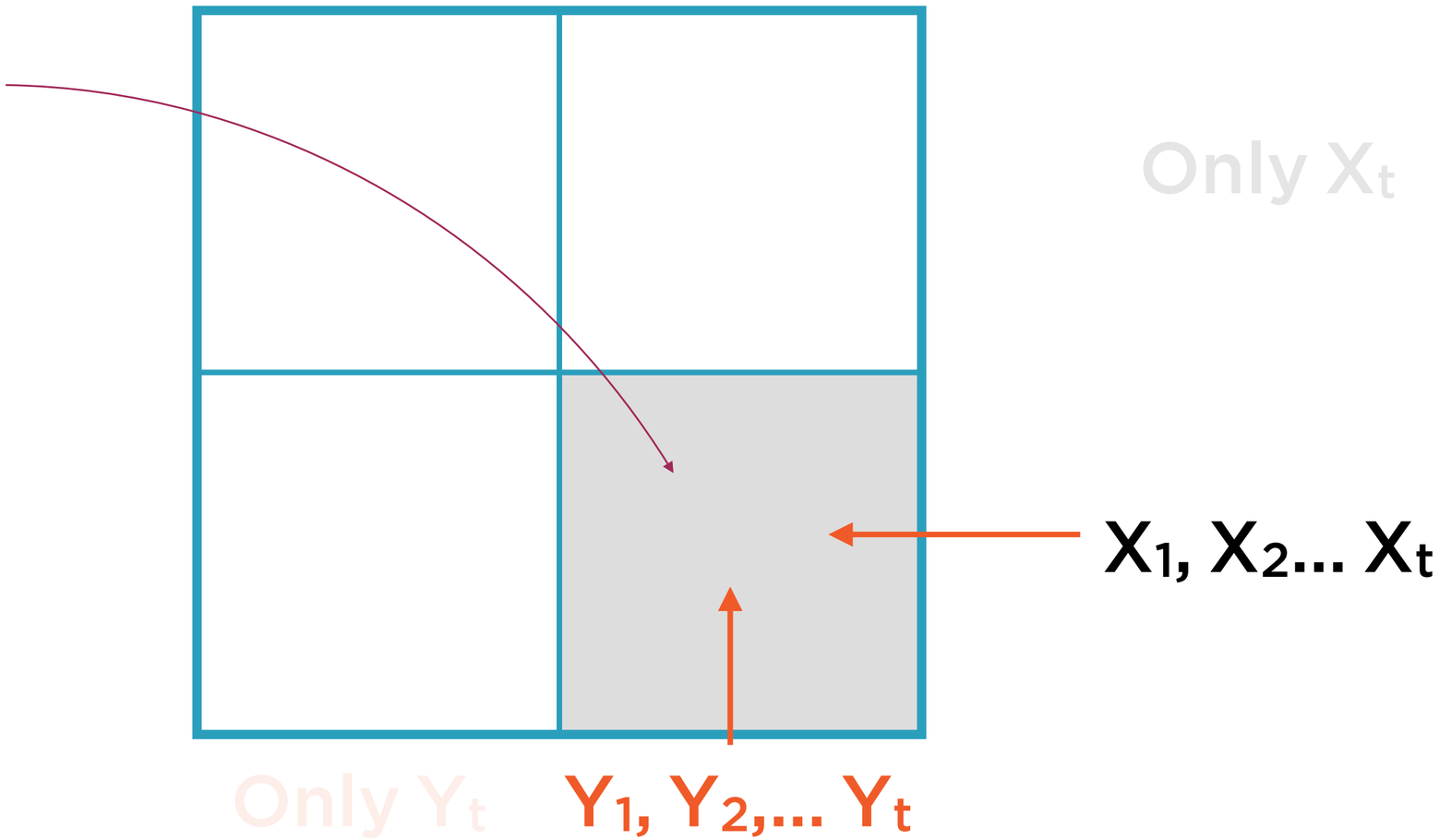
Sequence-to-Sequence



Sequence-to-Sequence

Output

e.g. language translation



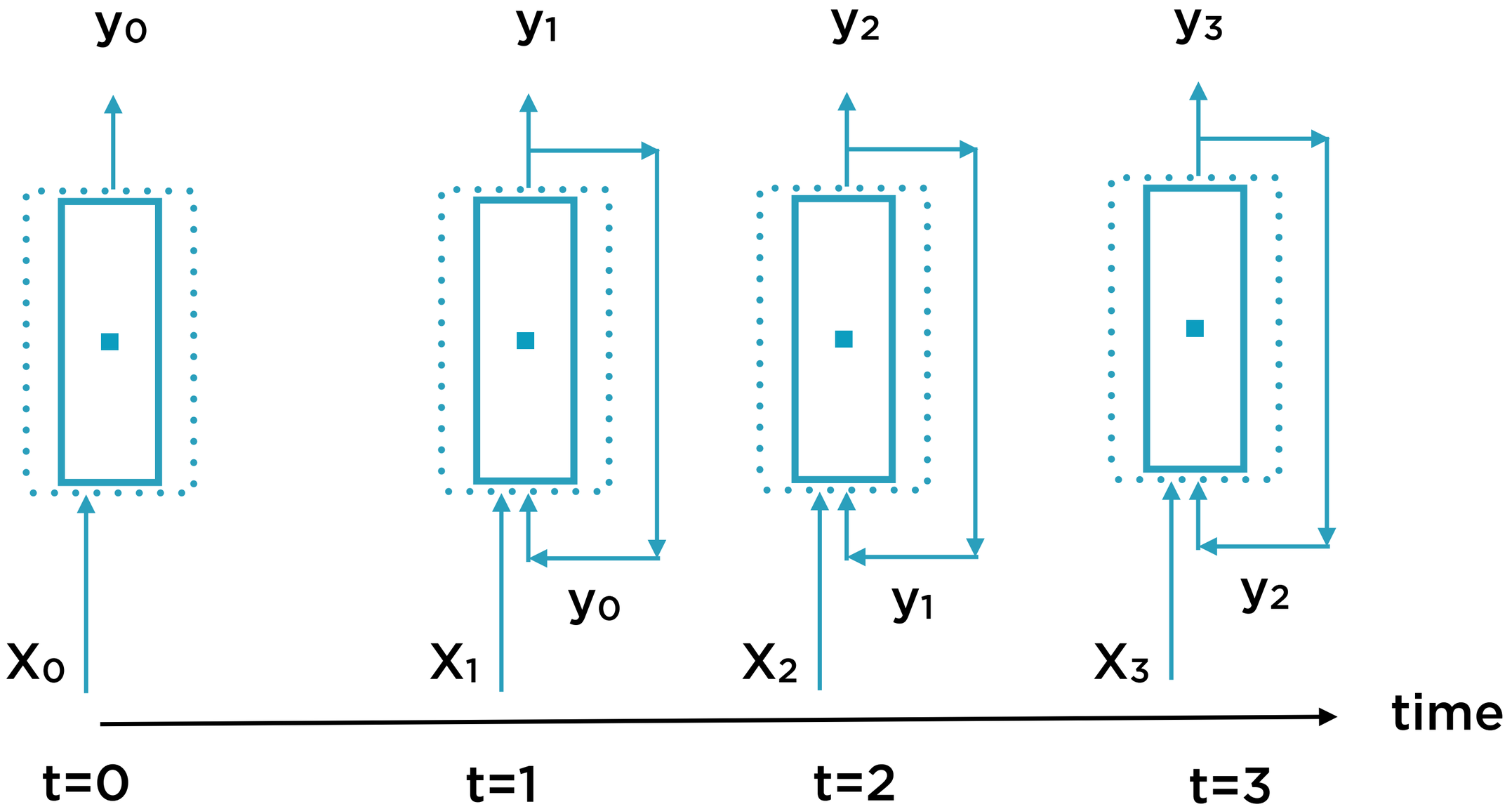
Only X_t

Input

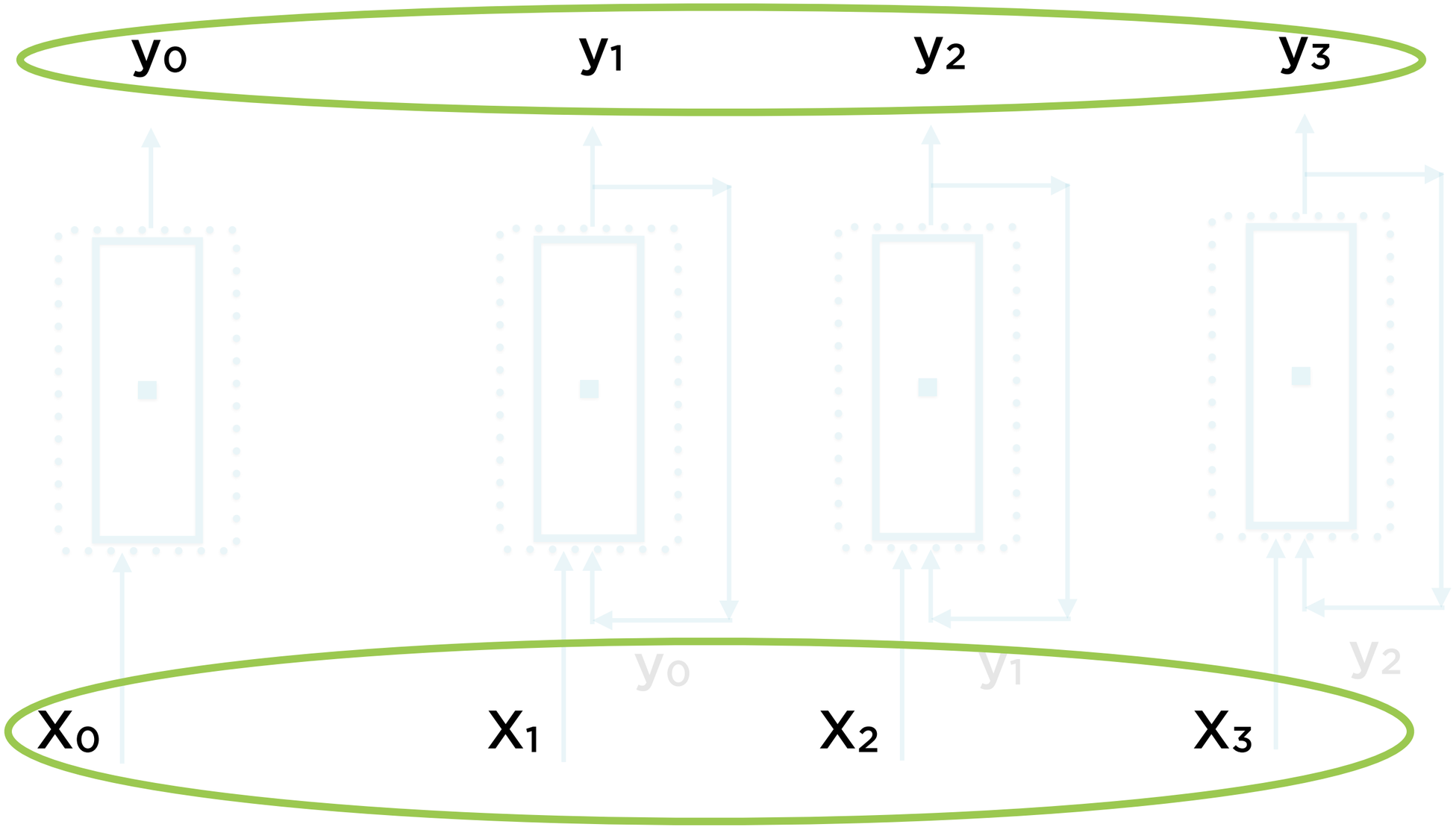
$X_1, X_2 \ldots X_t$

Only Y_t $Y_1, Y_2, \ldots Y_t$

Sequence-to-Sequence



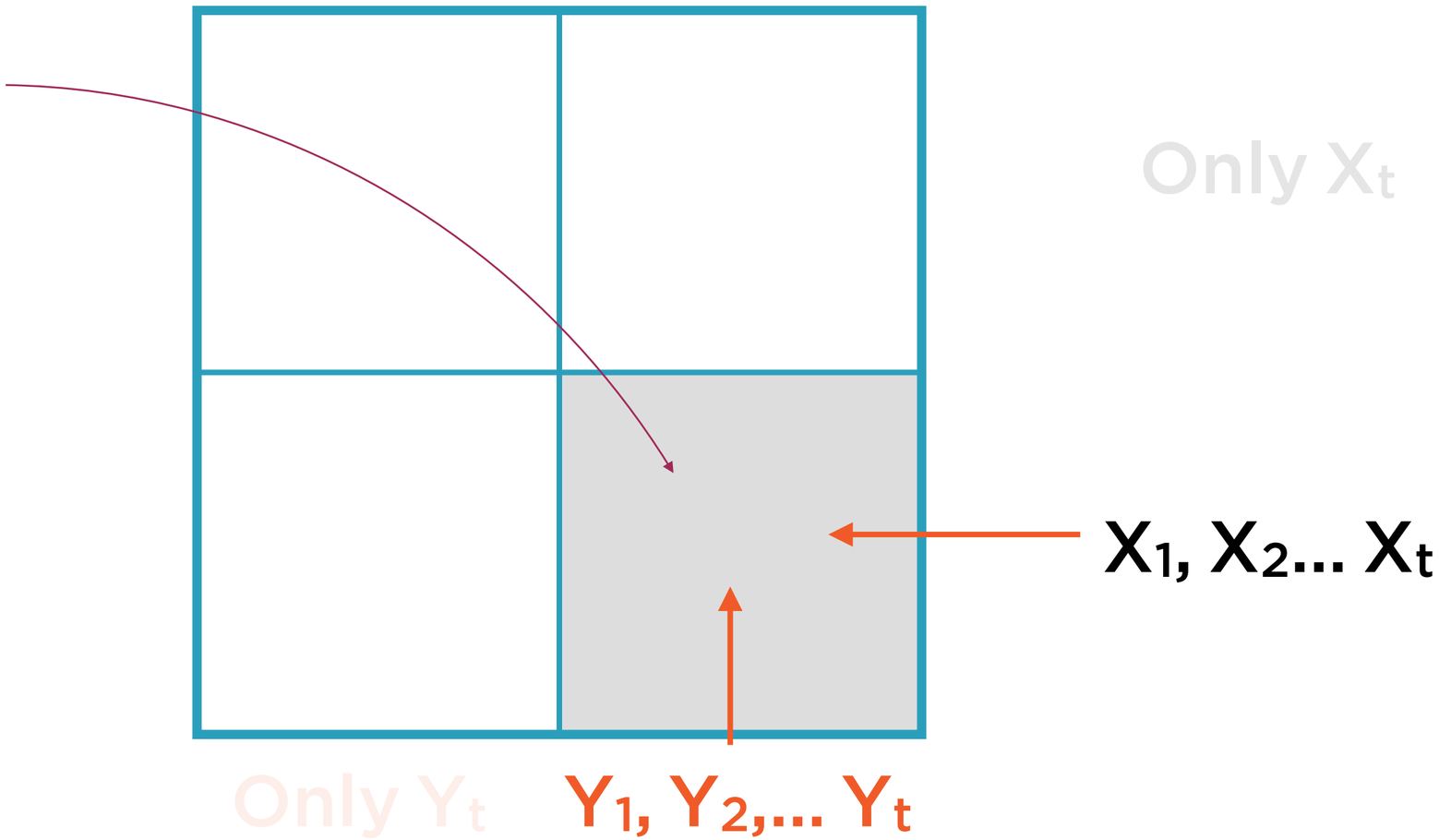
Sequence-to-Sequence



Sequence-to-Sequence

Output

e.g. language translation



Only X_t

Input

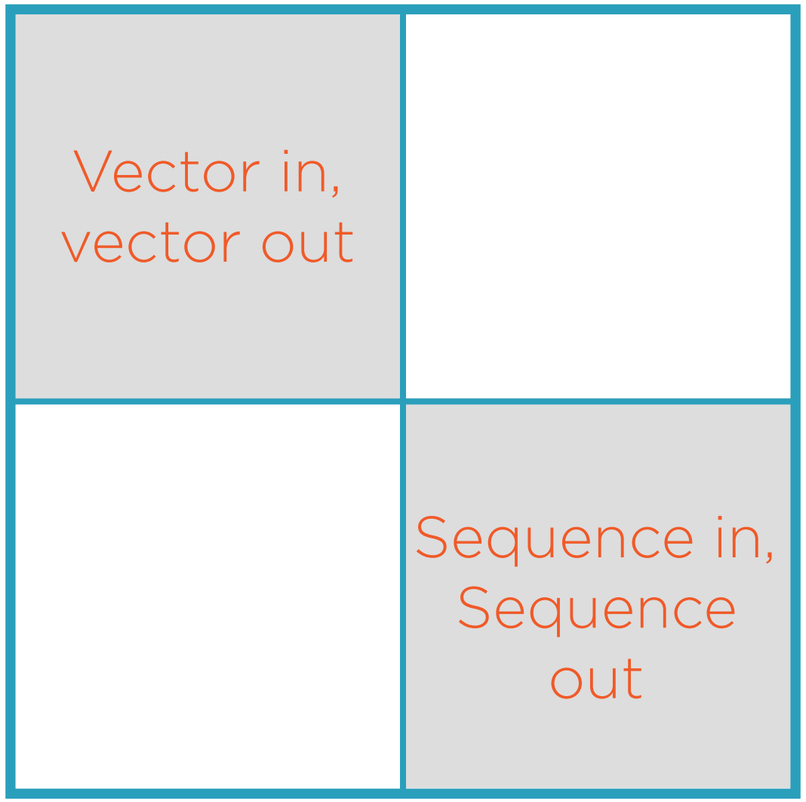
$X_1, X_2 \ldots X_t$

Only Y_t $Y_1, Y_2, \ldots Y_t$

Sequential Data and RNNs

Output

Input



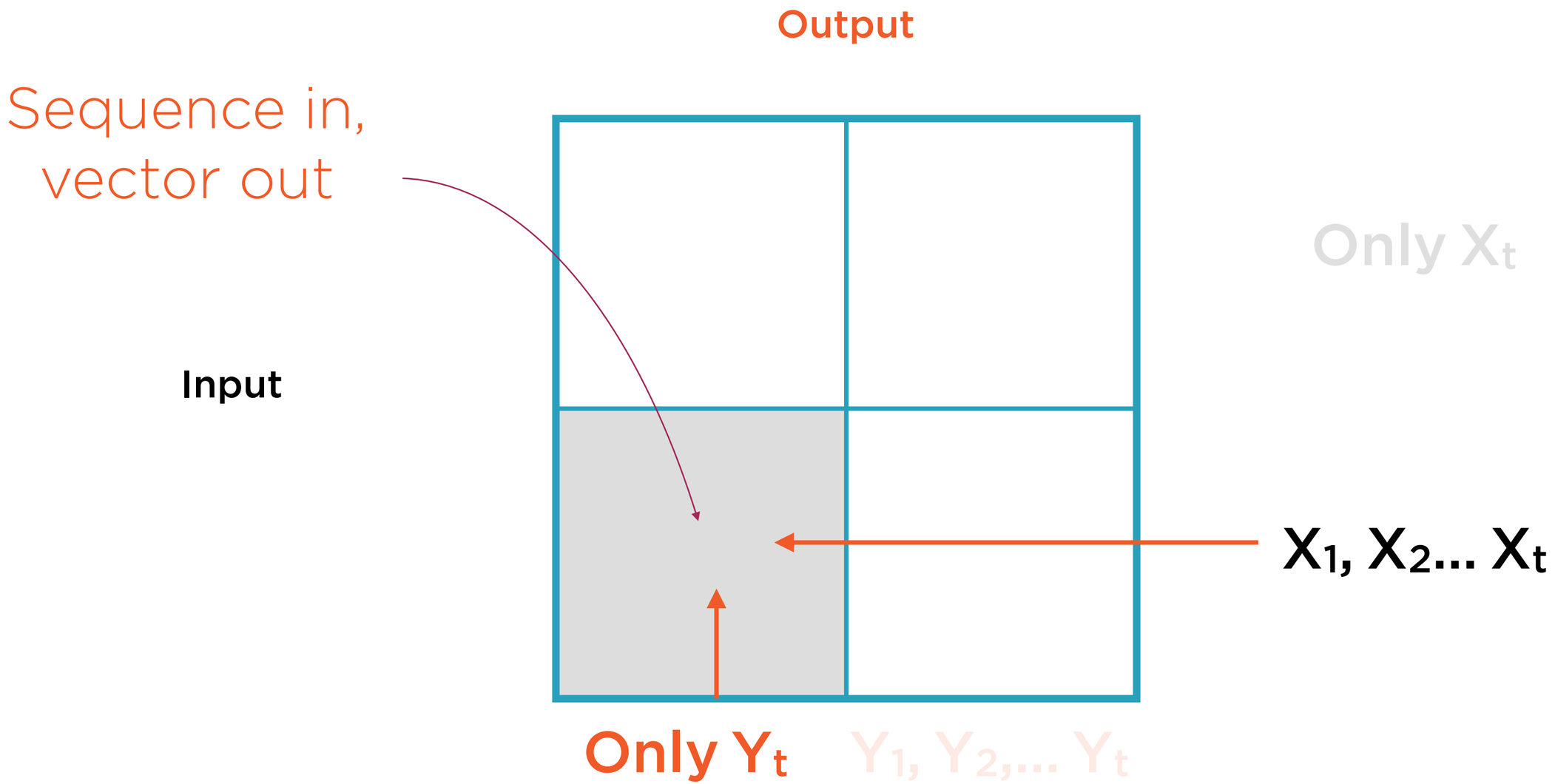
Only X_t

$X_1, X_2 \ldots X_t$

Only Y_t

$Y_1, Y_2, \ldots Y_t$

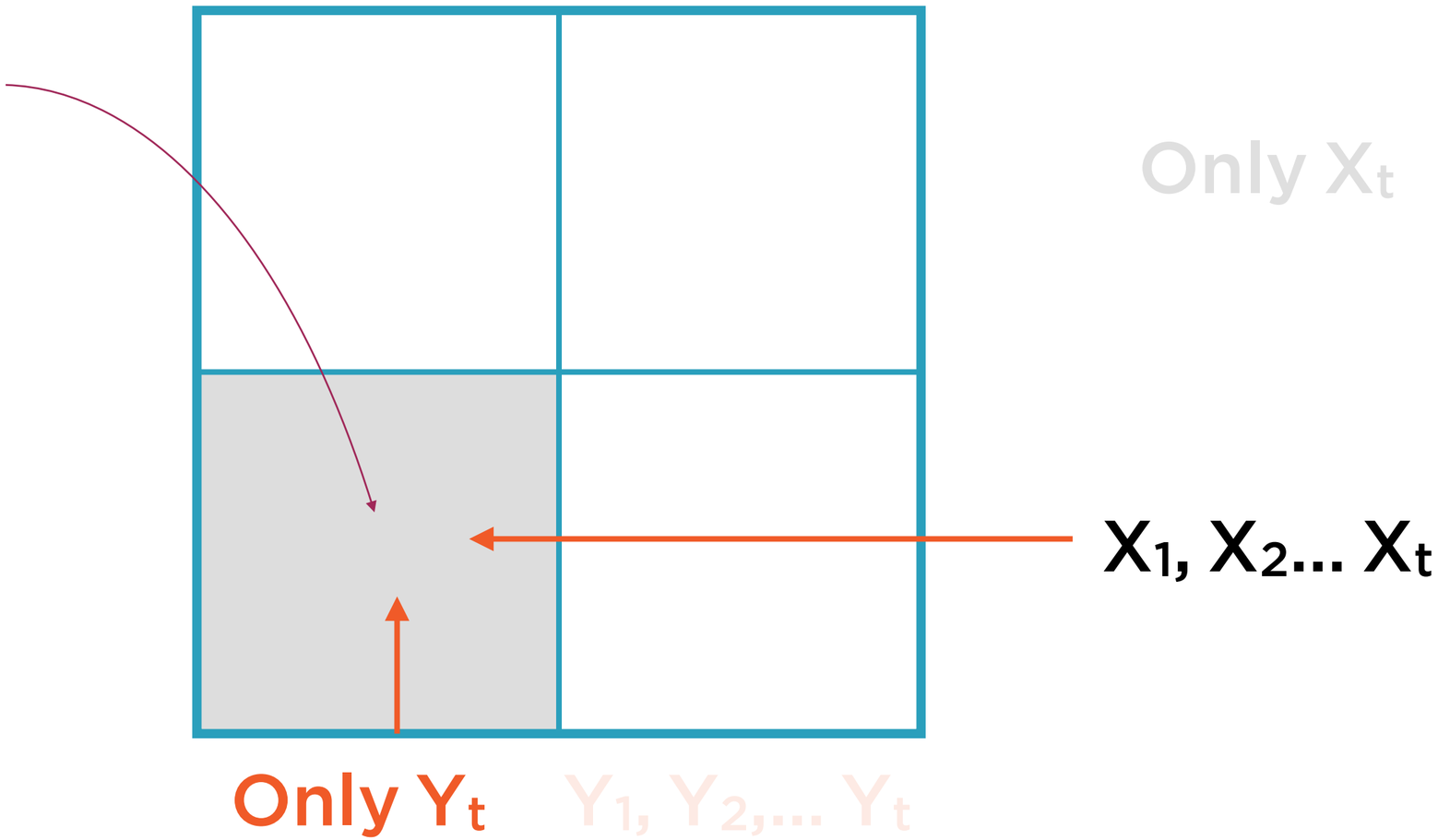
Sequential Data and RNNs



Sequence-to-Vector

Output

e.g. stock price prediction



Only X_t

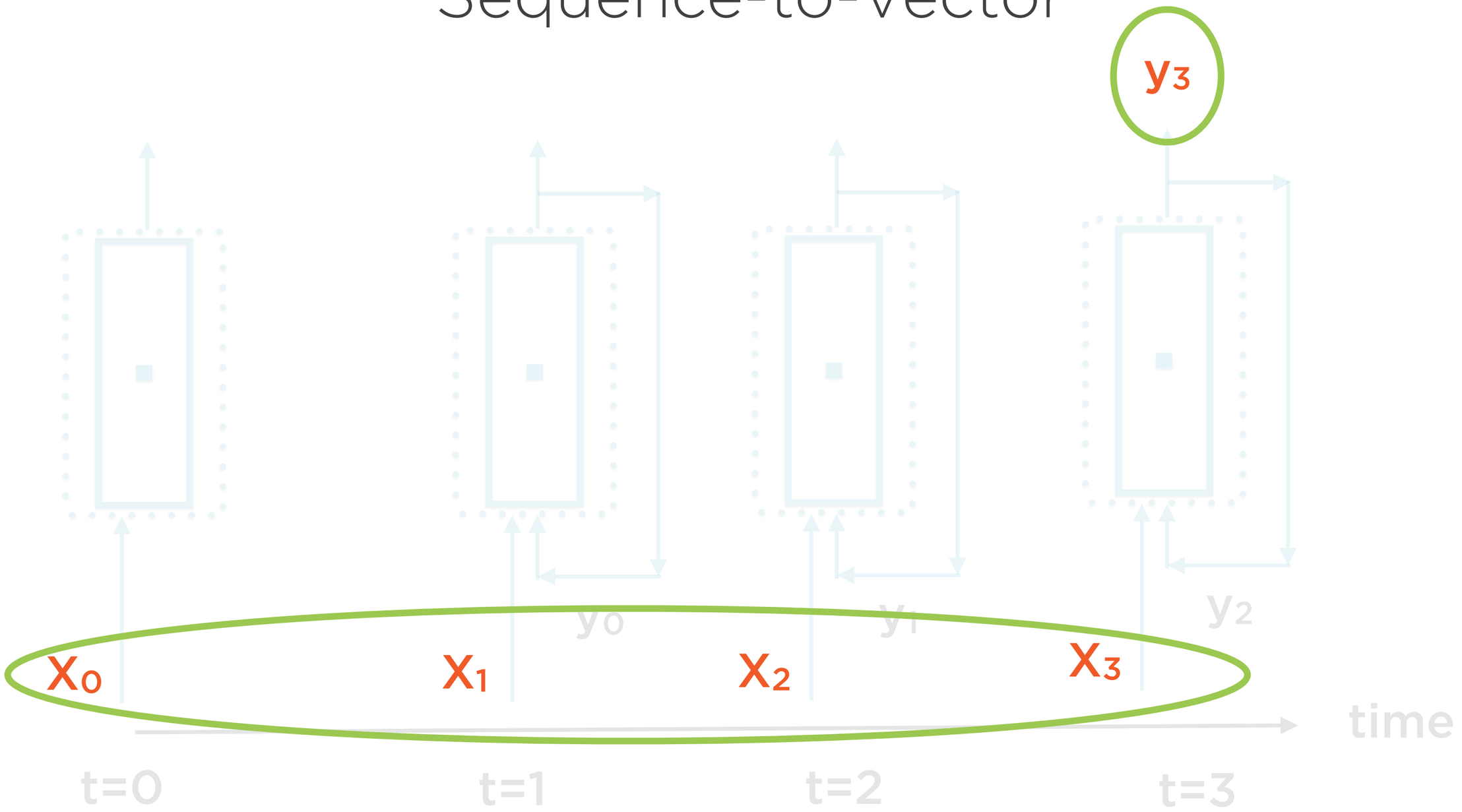
Input

$X_1, X_2 \ldots X_t$

Only Y_t

$Y_1, Y_2, \ldots Y_t$

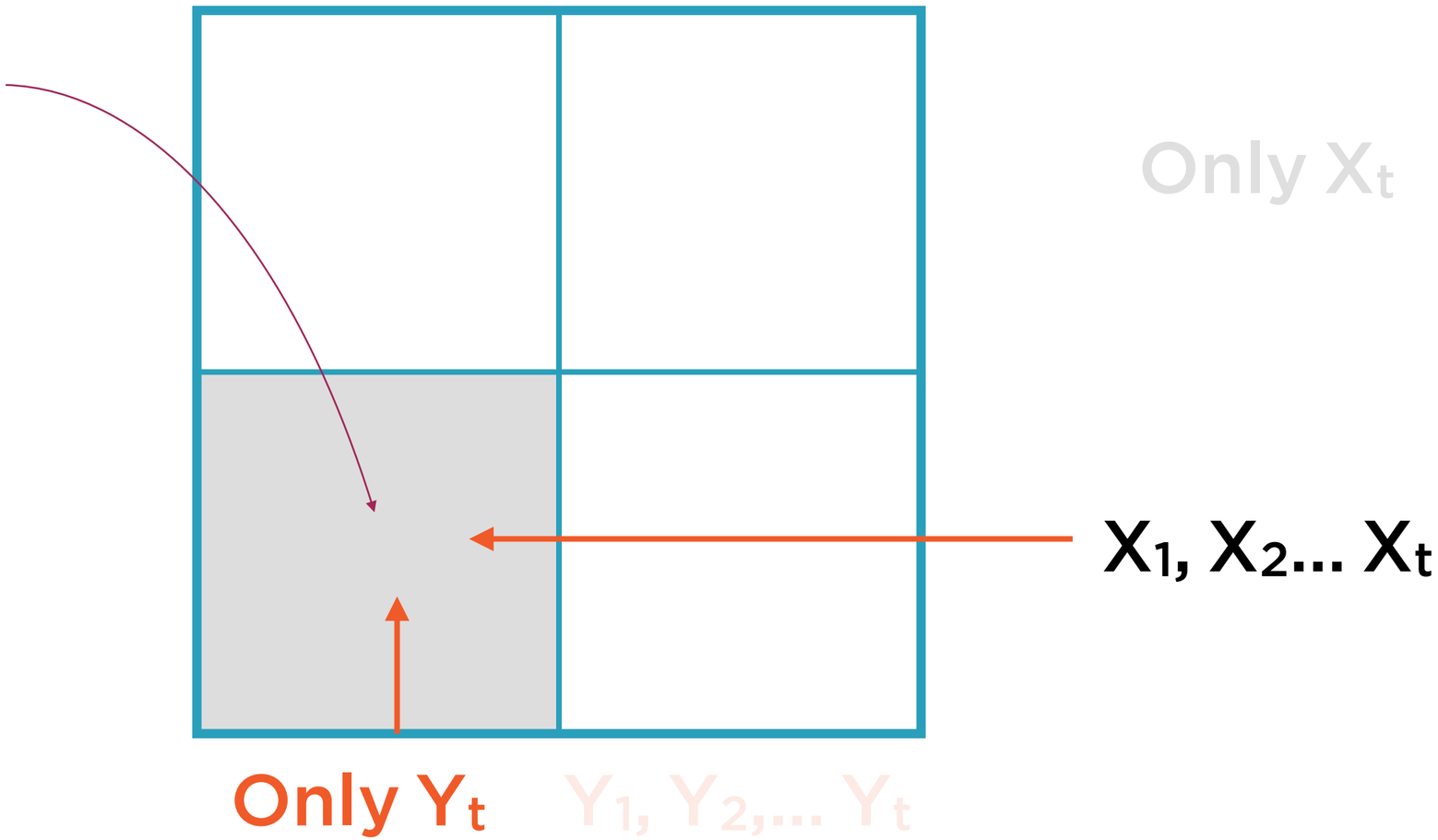
Sequence-to-Vector



Sequence-to-Vector

Output

e.g. stock price prediction



Only X_t

Input

$X_1, X_2 \ldots X_t$

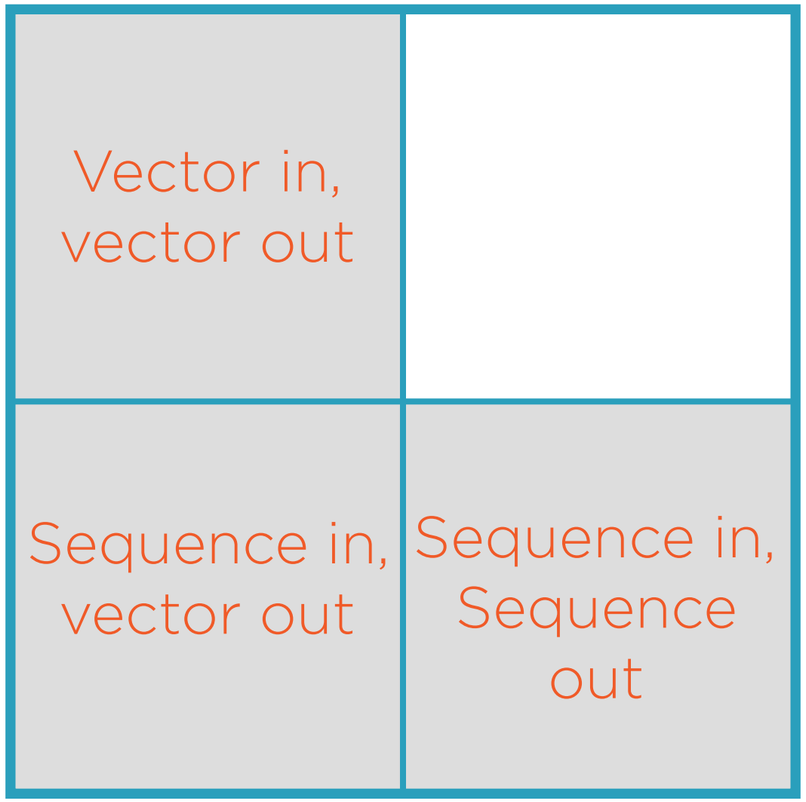
Only Y_t

$Y_1, Y_2, \ldots Y_t$

Sequential Data and RNNs

Output

Input



Only X_t

$X_1, X_2 \ldots X_t$

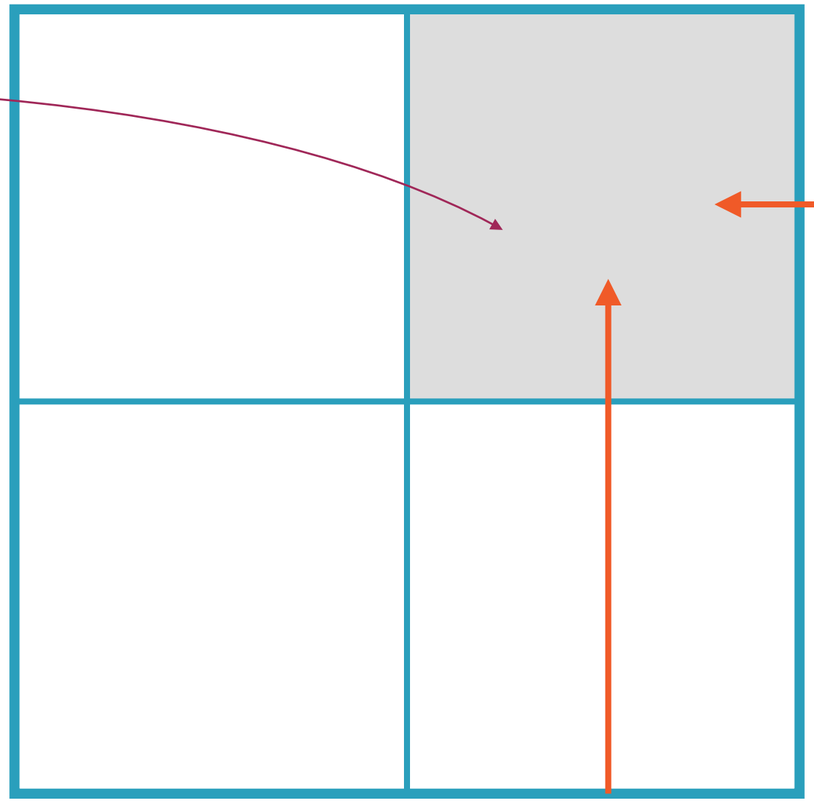
Only Y_t

$Y_1, Y_2, \ldots Y_t$

Sequential Data and RNNs

Output

Vector in, sequence out



Only X_t

Input

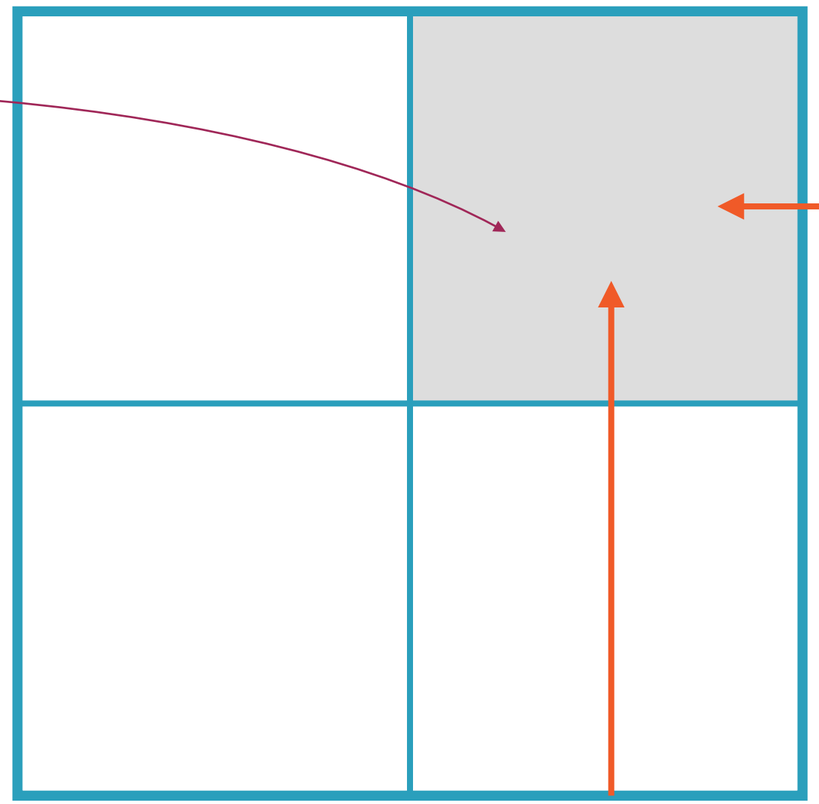
$X_1, X_2 \ldots X_t$

Only Y_t $Y_1, Y_2, \ldots Y_t$

Vector-to-Sequence

Output

e.g. text autocomplete



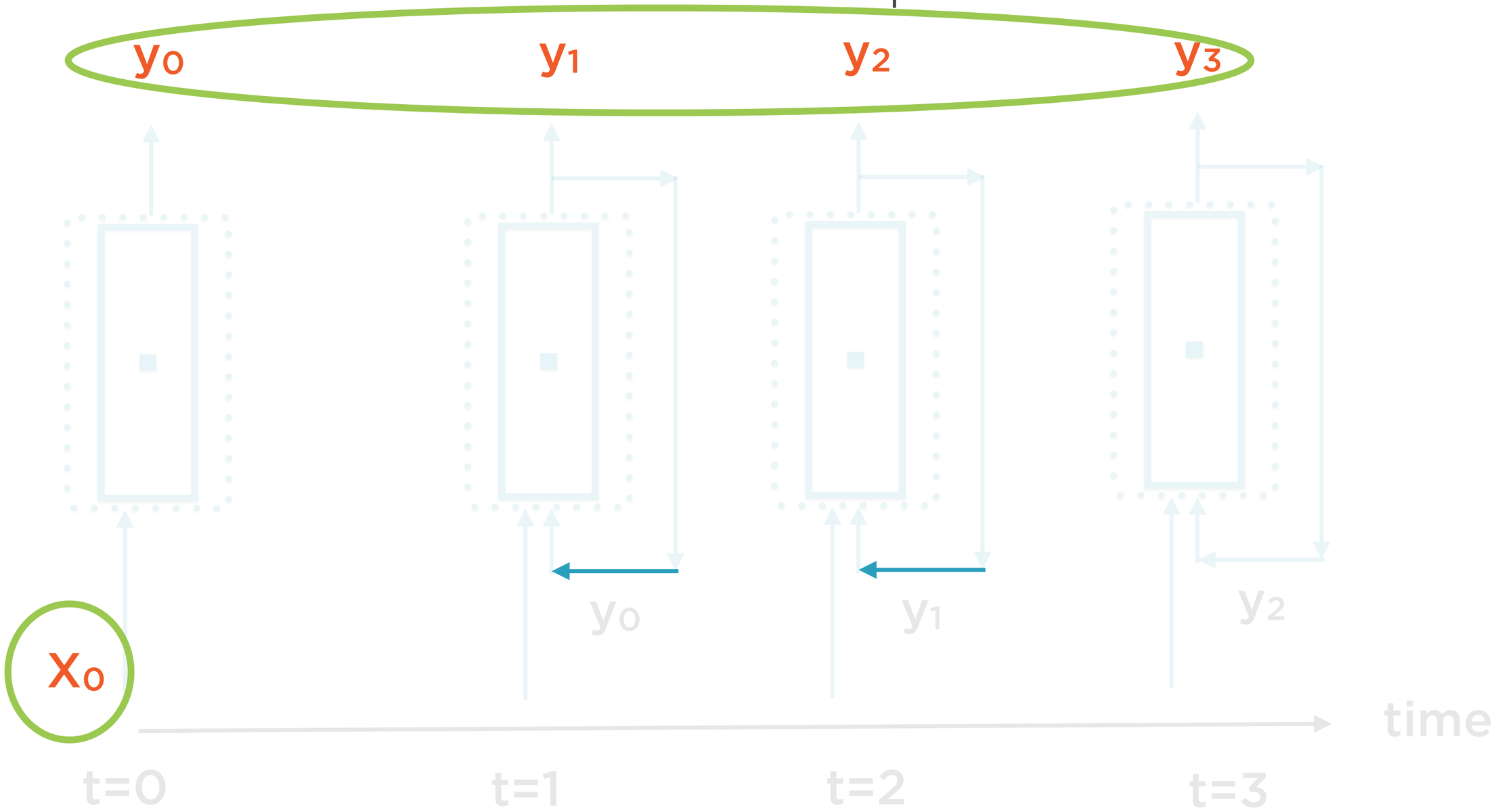
Only X_t

Input

$X_1, X_2 \ldots X_t$

Only Y_t $Y_1, Y_2, \ldots Y_t$

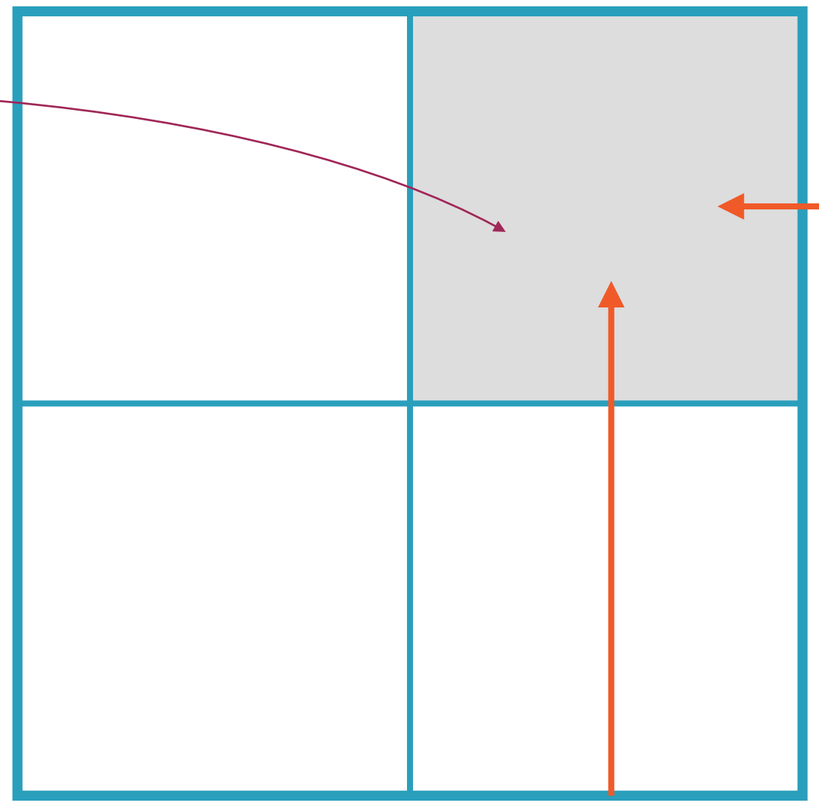
Vector-to-Sequence



Vector-to-Sequence

Output

e.g. text autocomplete



Only X_t

Input

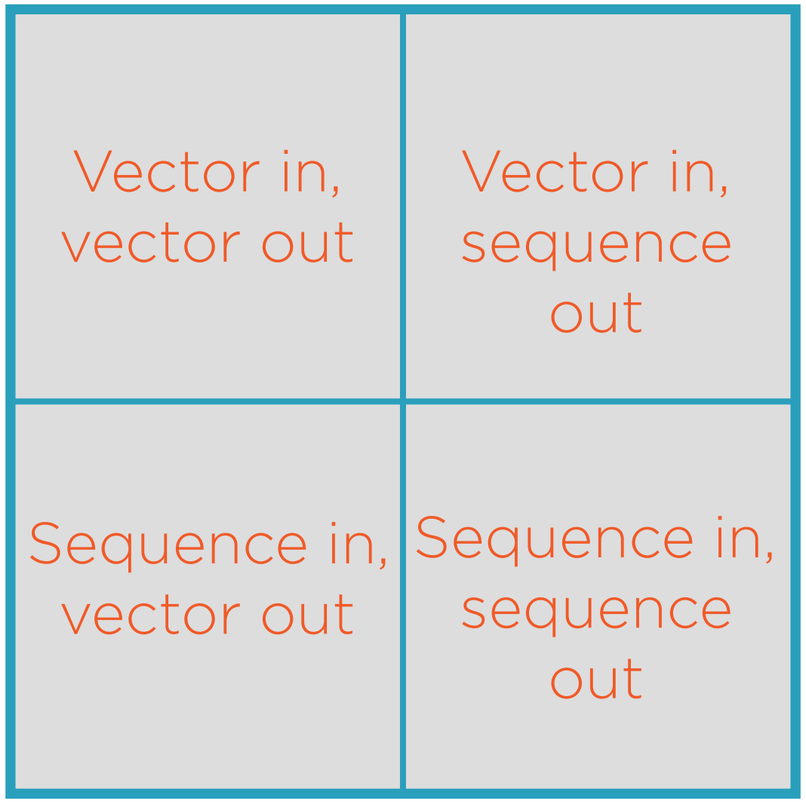
$X_1, X_2 \ldots X_t$

Only Y_t $Y_1, Y_2, \ldots Y_t$

Sequential Data and RNNs

Output

Input



Only X_t

$X_1, X_2 \ldots X_t$

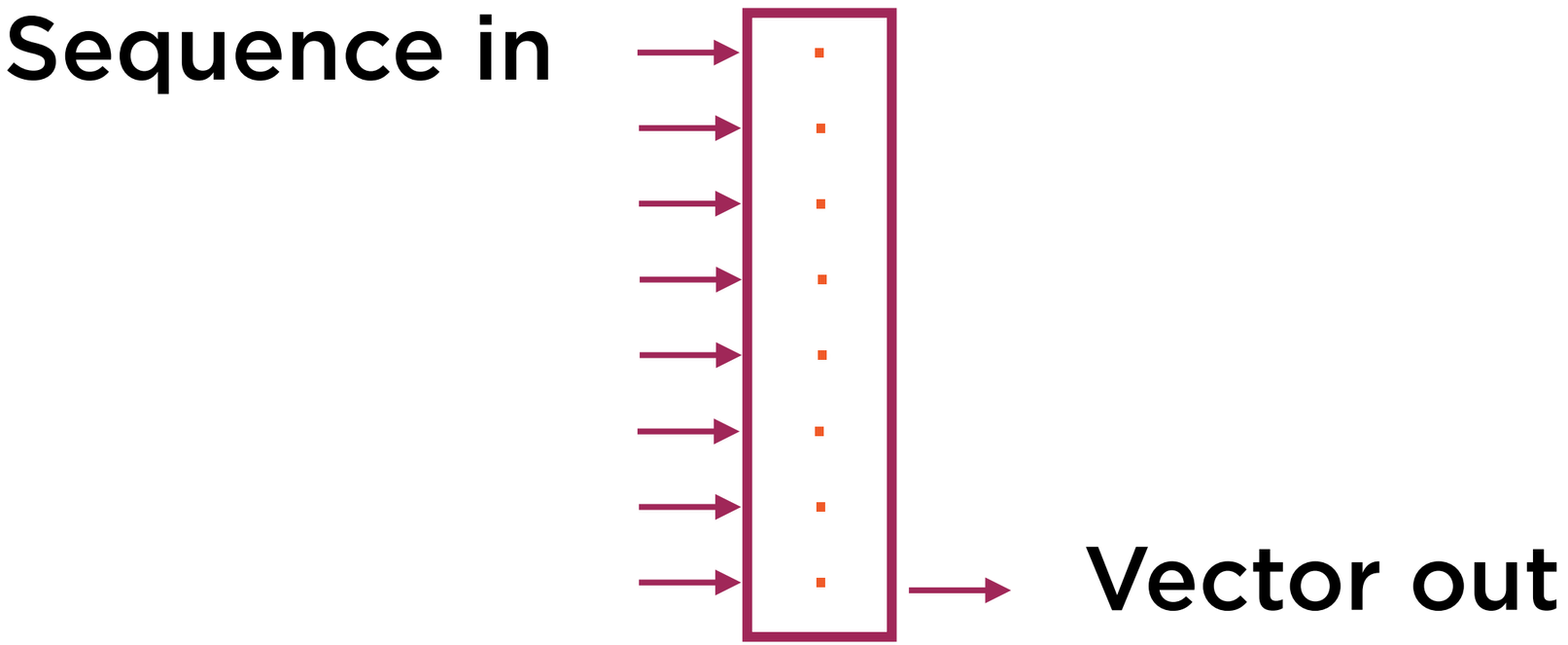
Only Y_t

$Y_1, Y_2, \ldots Y_t$

Language Translation Using RNNs

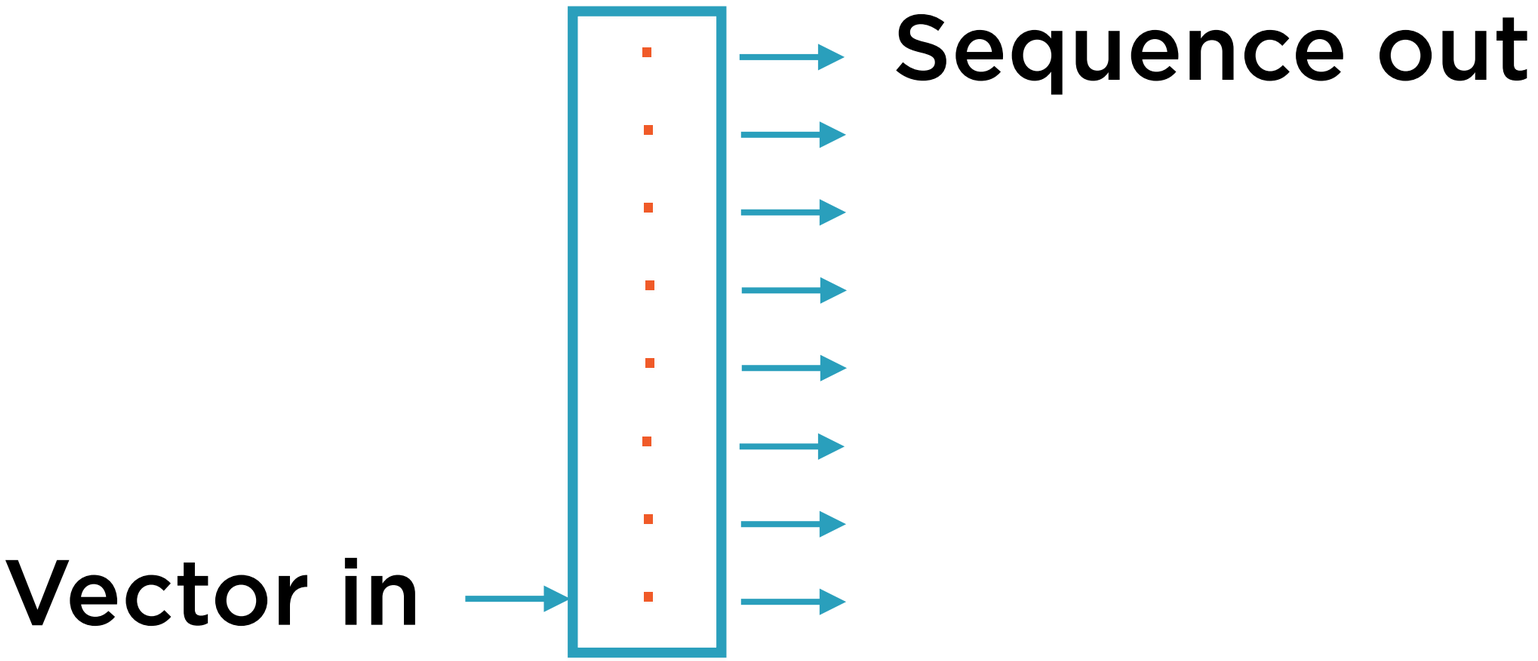
Encoders and Decoders

Encoder = Sequence in, Vector out



Encoders and Decoders

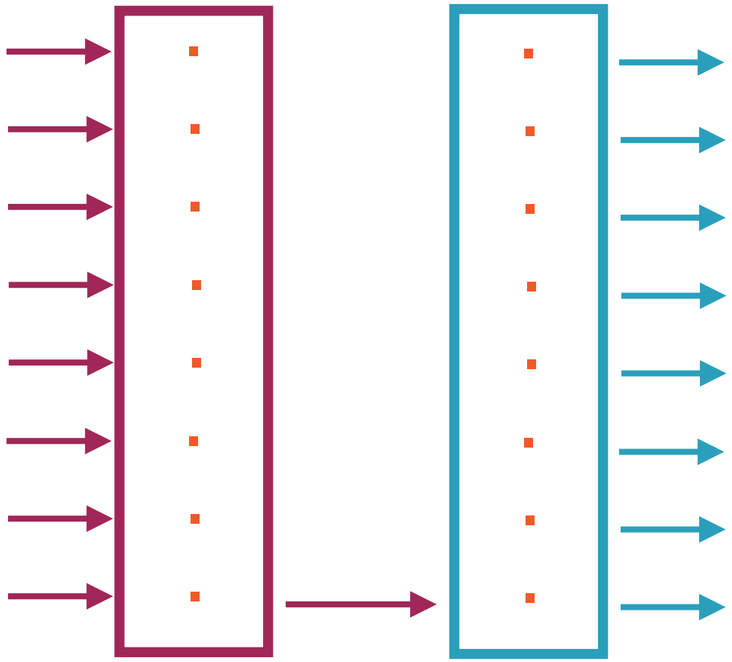
Decoder = Vector in, Sequence out



Encoders and Decoders

Encoder-Decoder = **Sequence in, Sequence out**

Sequence in

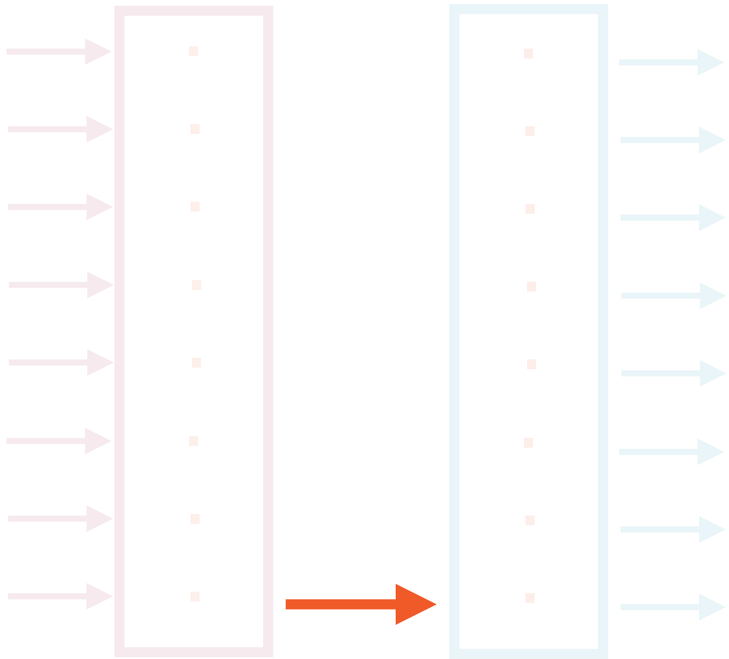


Sequence out

Encoders and Decoders

The final hidden state of the encoder

Sequence in

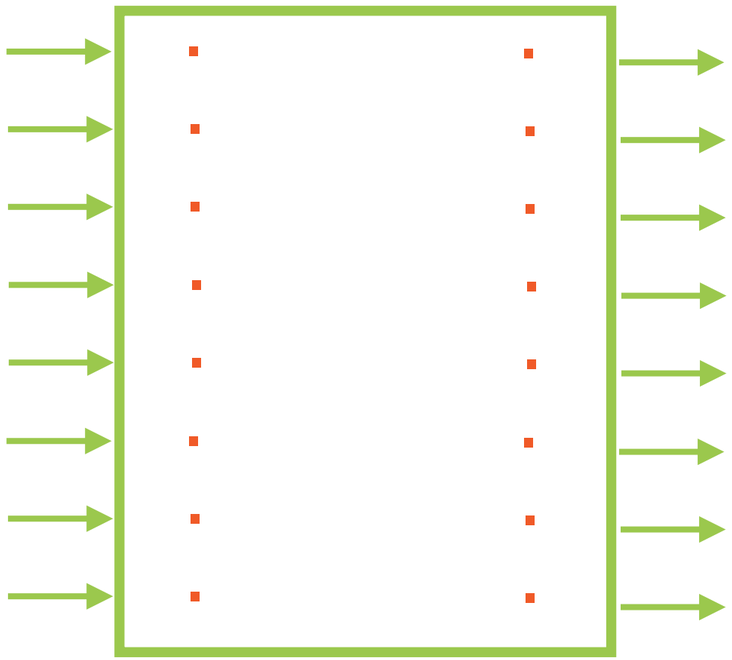


Sequence out

Encoders and Decoders

Encoder-Decoder = **Sequence in, Sequence out**

Sequence in

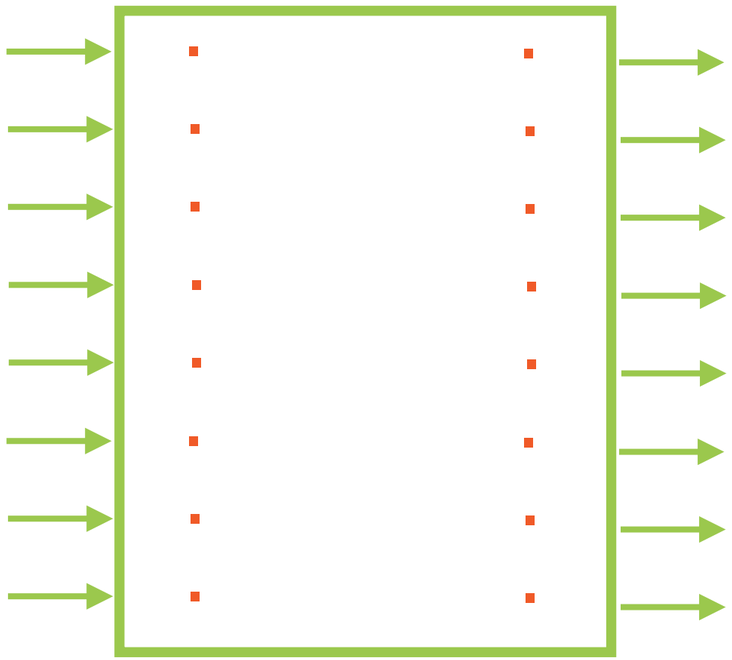


Sequence out

English to French

Sequence-to-sequence RNN with GRU cells

“The cat sat on the mat”

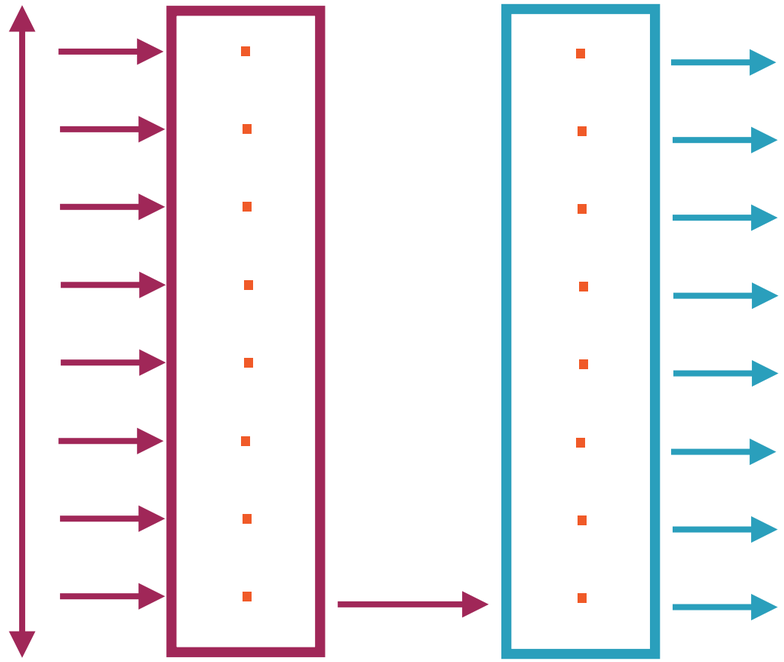


“le chat etait assis sur le tapis”

English to French

Sequence-to-sequence RNN with GRU cells

“The cat sat on the mat”



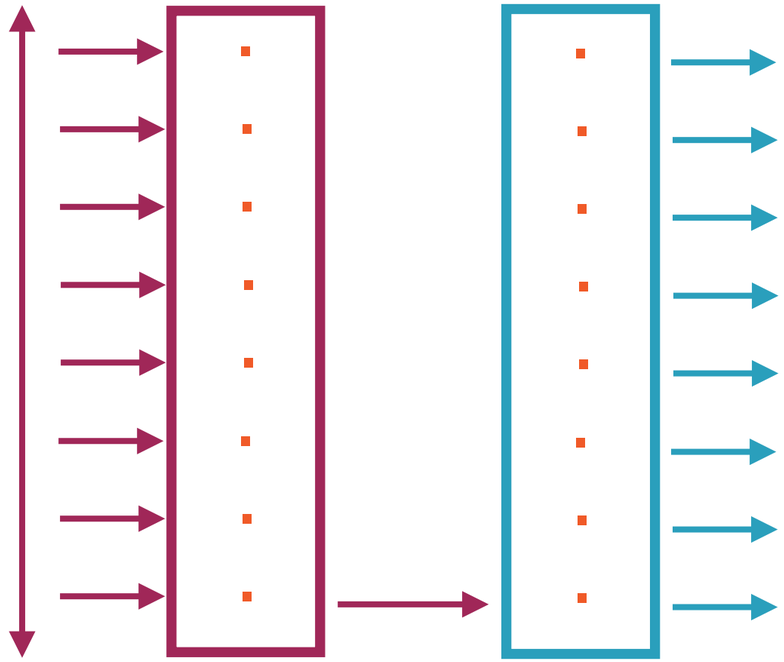
“le chat etait assis sur le tapis”

Representing Input and Target Sentences

English to French

Pad inputs to be as long as longest possible English sentence

“The cat sat on the mat”

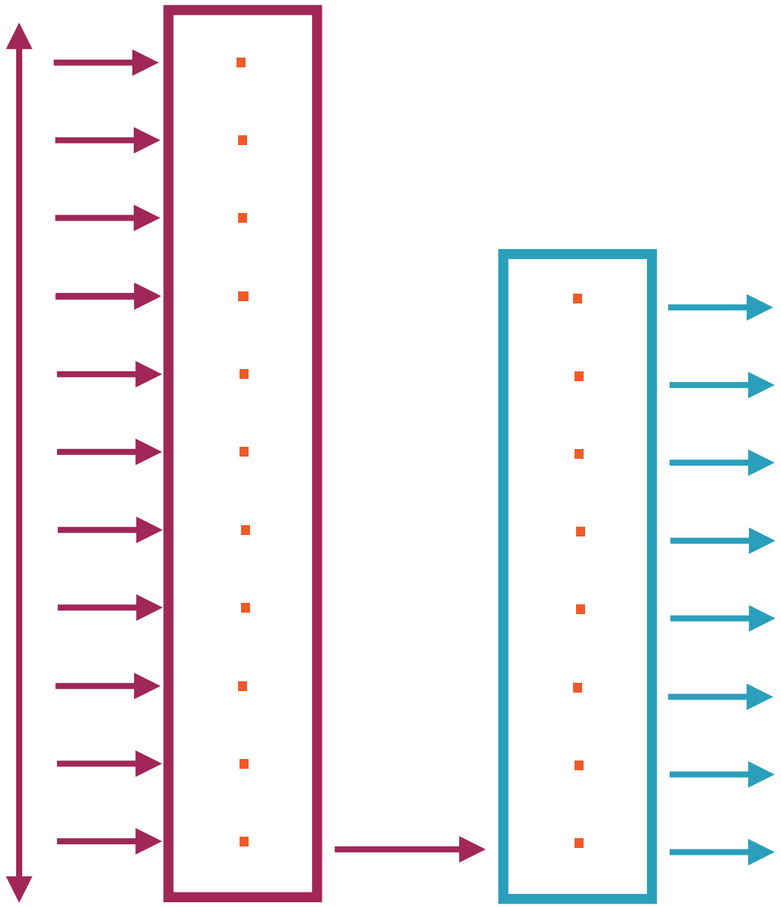


“le chat etait assis sur le tapis”

English to French

Pad inputs to be as long as longest possible English sentence

“The cat sat on the mat”

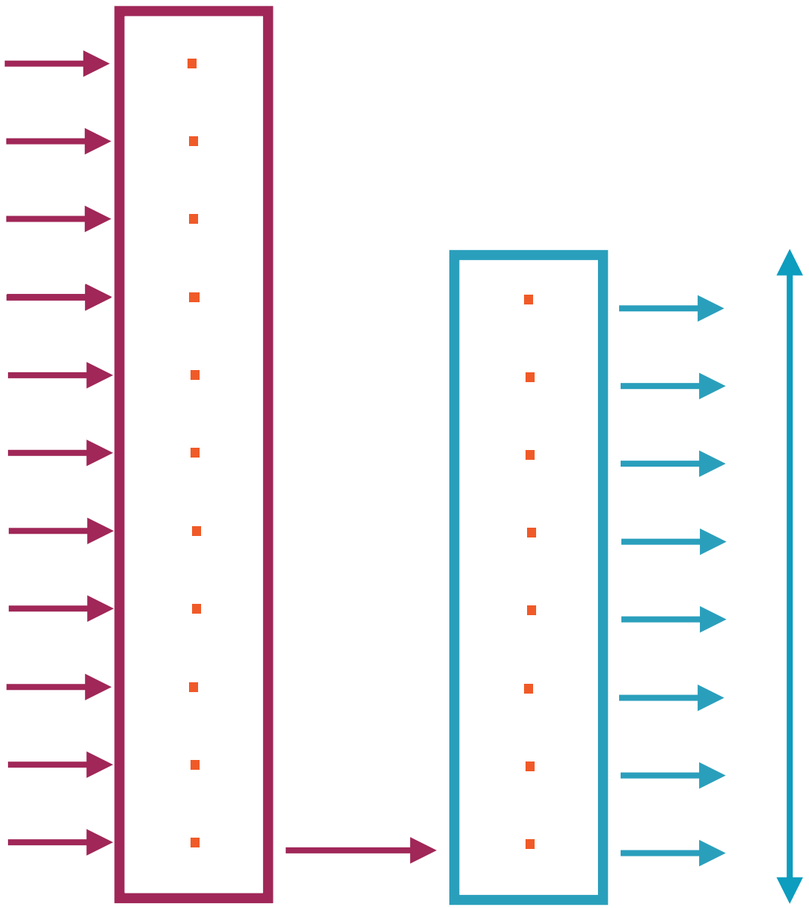


“le chat etait assis sur le tapis”

English to French

Pad outputs to be as long as longest possible French sentence

“The cat sat on the mat”

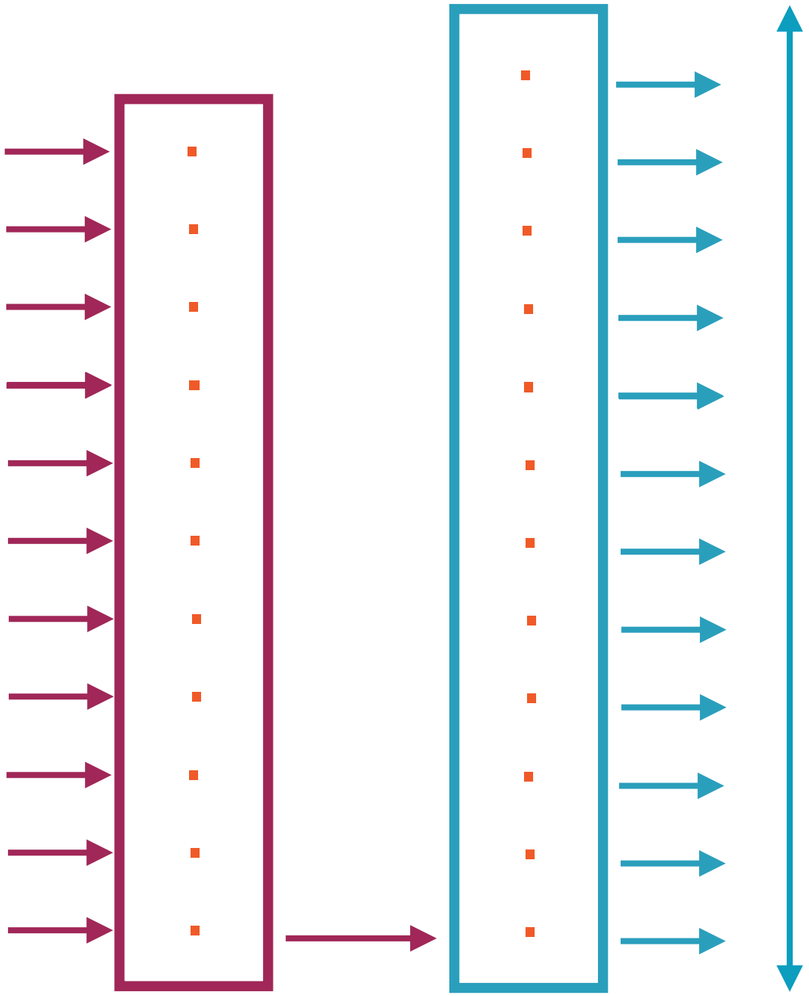


“le chat etait assis sur le tapis”

English to French

Pad outputs to be as long as longest possible French sentence

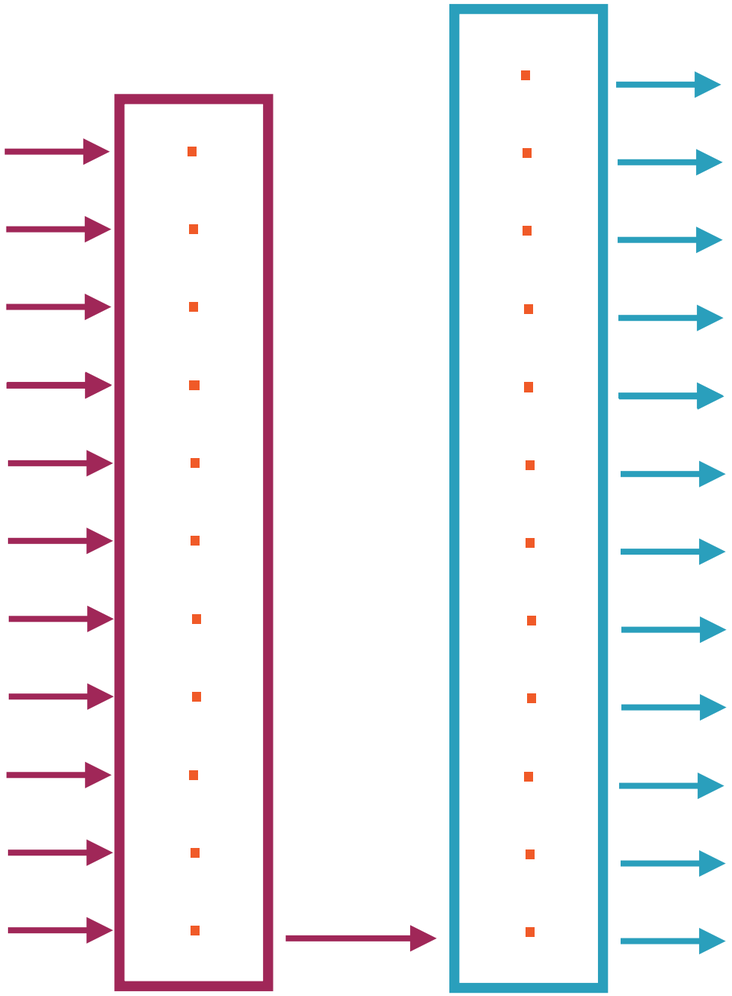
“The cat sat on the mat”



“le chat etait assis sur le tapis”

Each input corresponds to an English word (represented as an embedding)

English to French

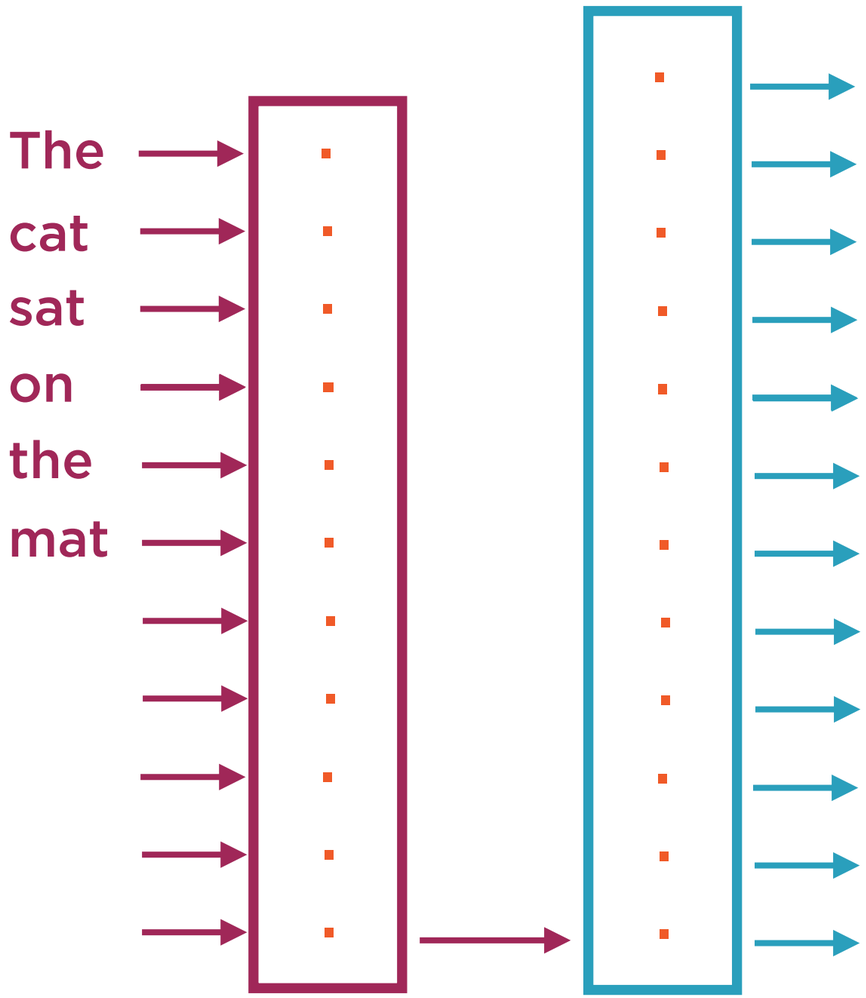


“The cat sat on the mat”

“le chat etait assis sur le tapis”

Each input corresponds to an English word (represented as an embedding)

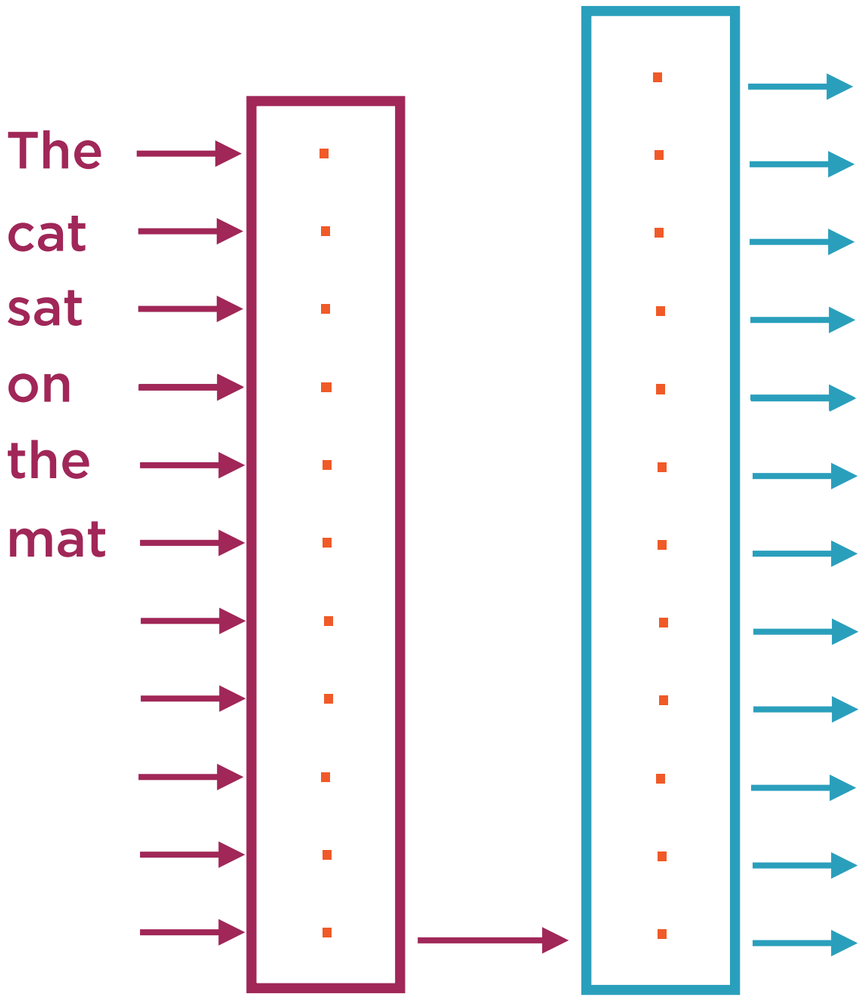
English to French



“le chat etait assis sur le tapis”

Each input corresponds to an English word (represented as an embedding)

English to French

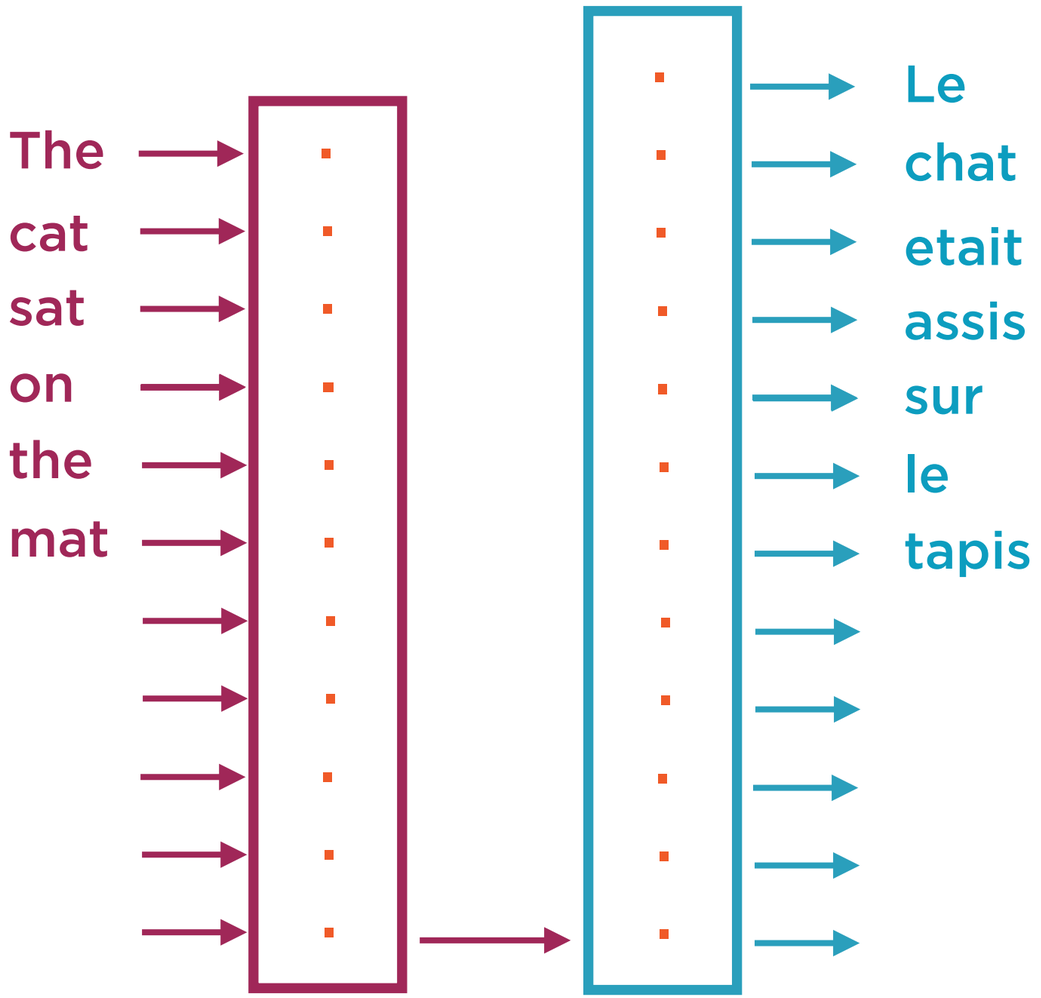


Each output corresponds to an French word (represented as an embedding)

“le chat etait assis sur le tapis”

Each input corresponds to an English word (represented as an embedding)

English to French



Each output corresponds to an French word (represented as an embedding)

Language Translation

Decoder Input

Training

Feed previous predicted French word as input to decoder

Prediction

Feed previous predicted French word as input to decoder

During training we have the correct translation available

Teacher Forcing

Training

Feed correct previous French word into the decoder

Prediction

Feed predicted French word as input to decoder

Feed the real word from the translated language a fraction of the time

Teacher Forcing

Training

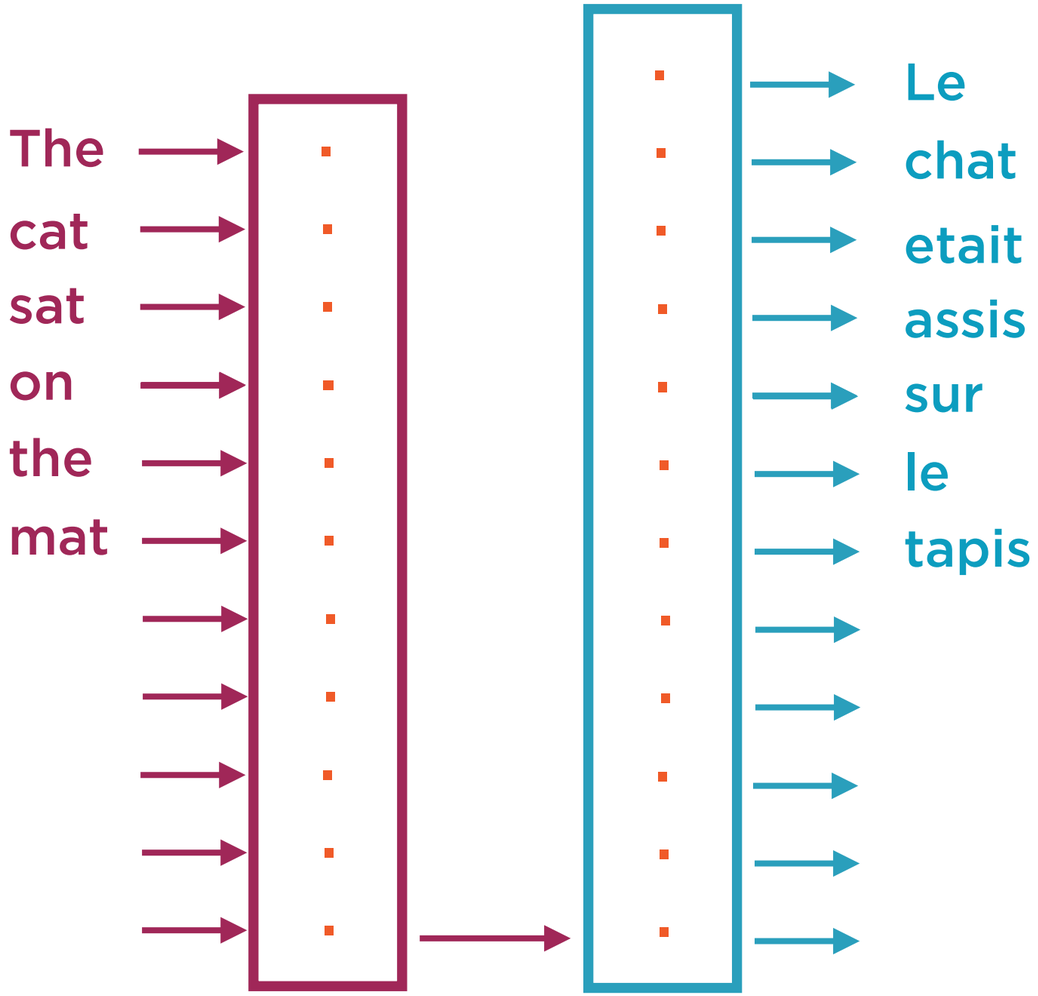
Feed correct previous French word into the decoder

Prediction

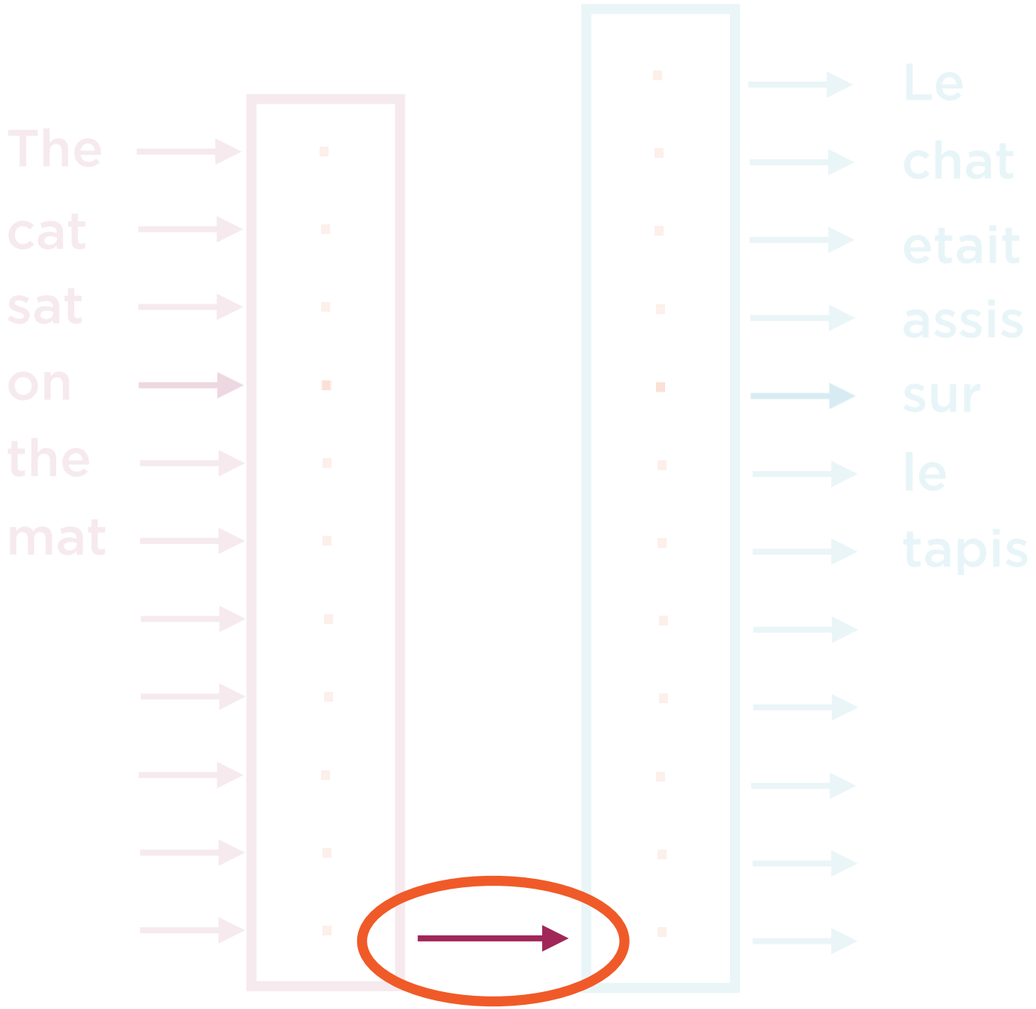
Feed predicted French word as input to decoder

The model converges faster but may be unstable when used in the real world

Decoder During Training

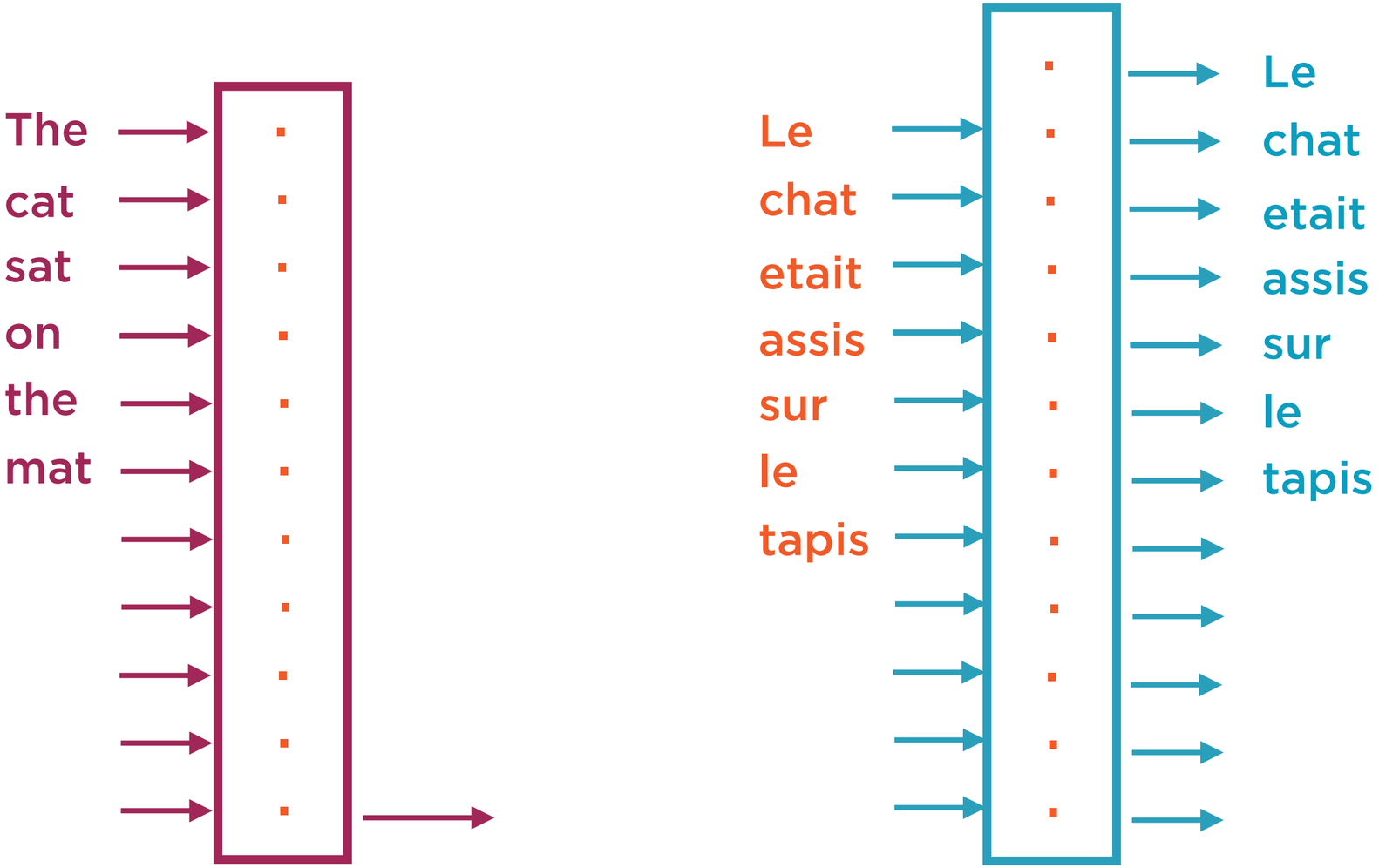


Decoder During Training

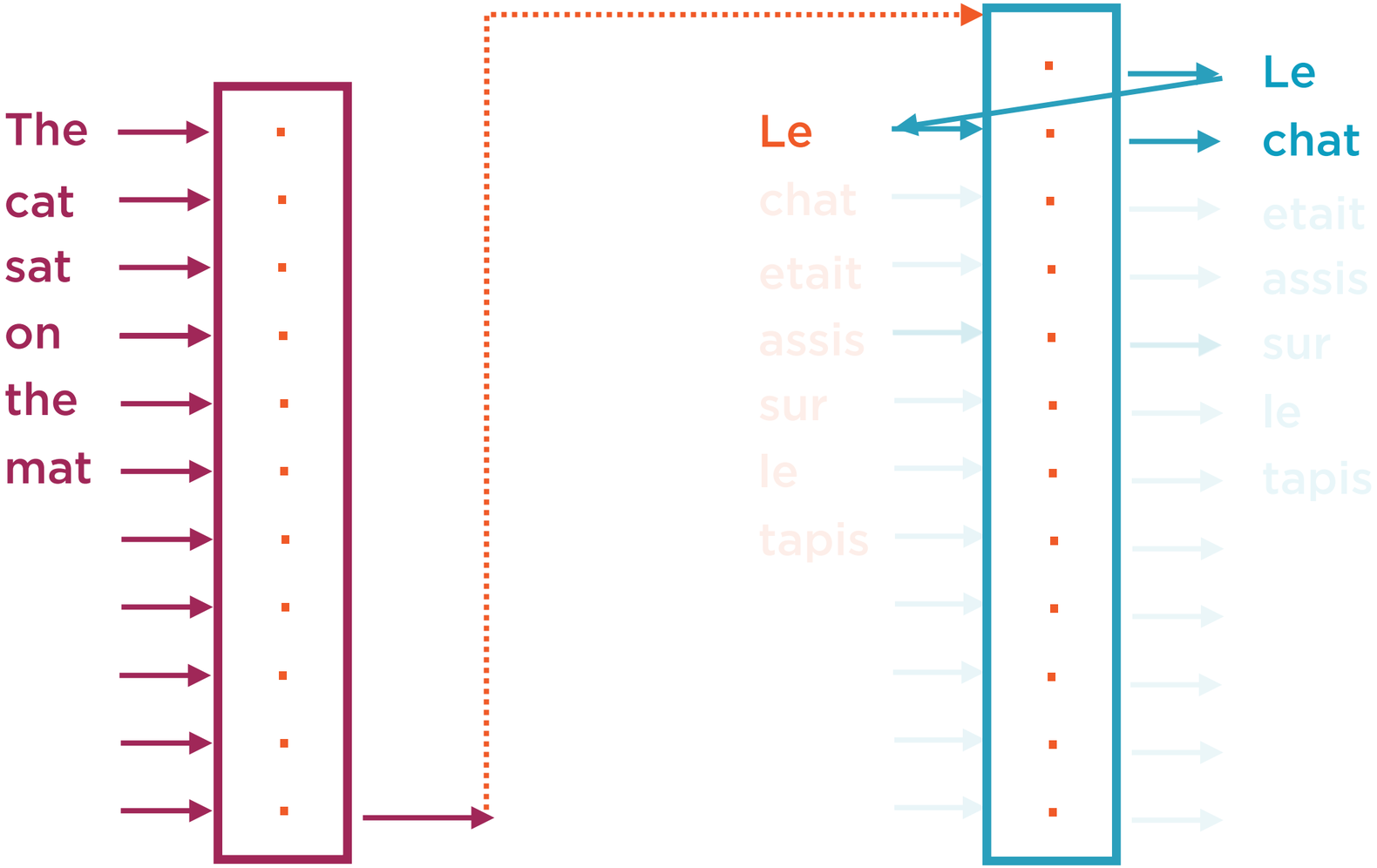


The encoder output is fed into the decoder to initialize states

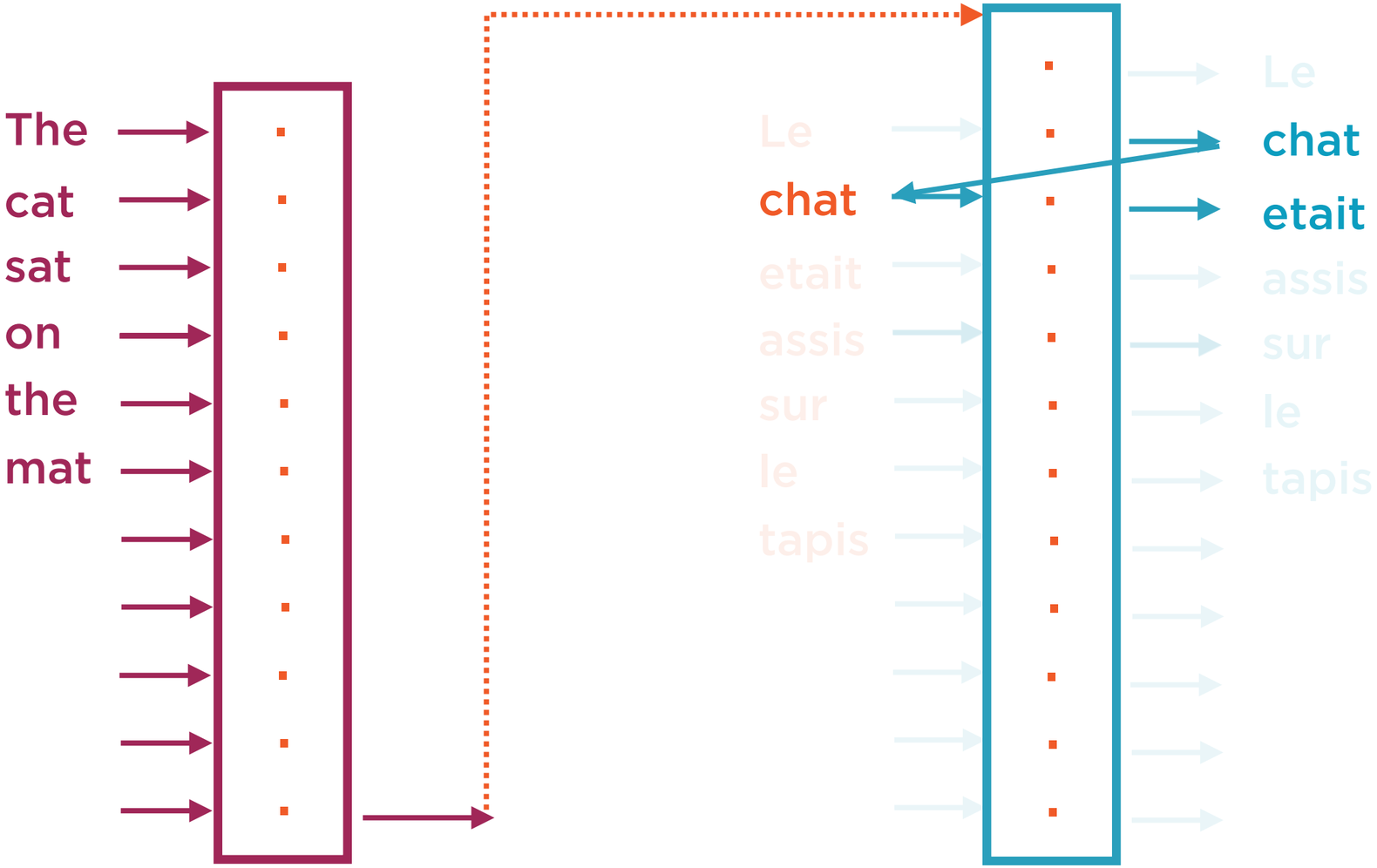
Decoder During Training



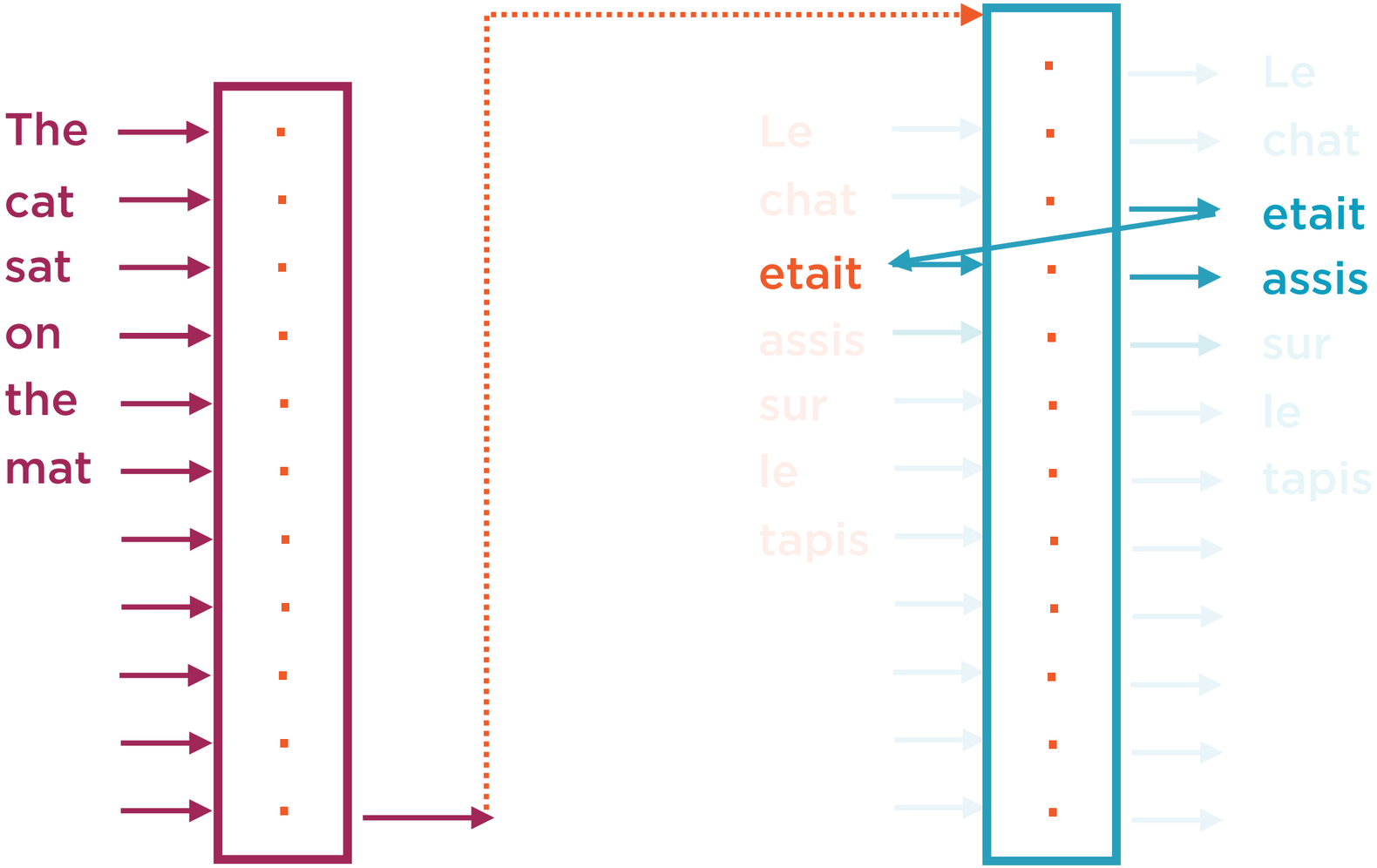
Decoder During Training



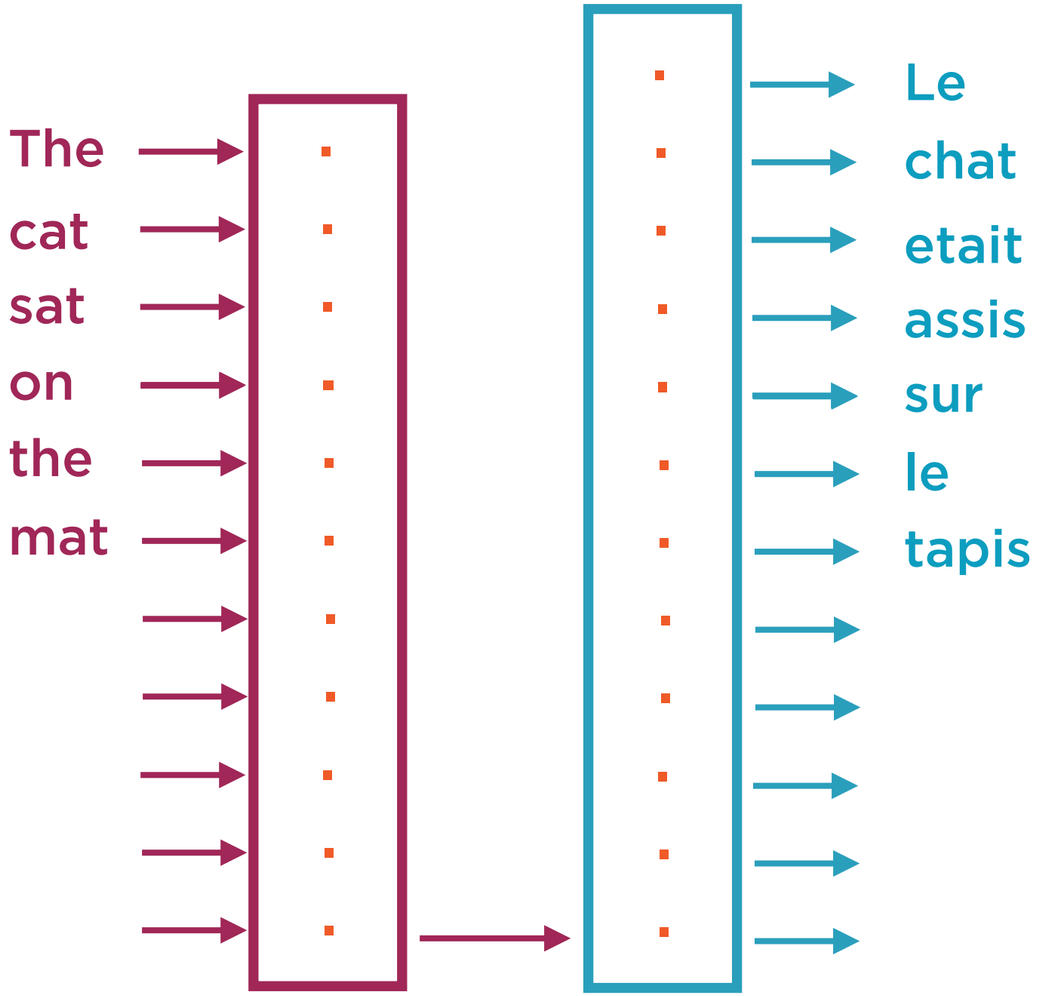
Decoder During Training



Decoder During Training

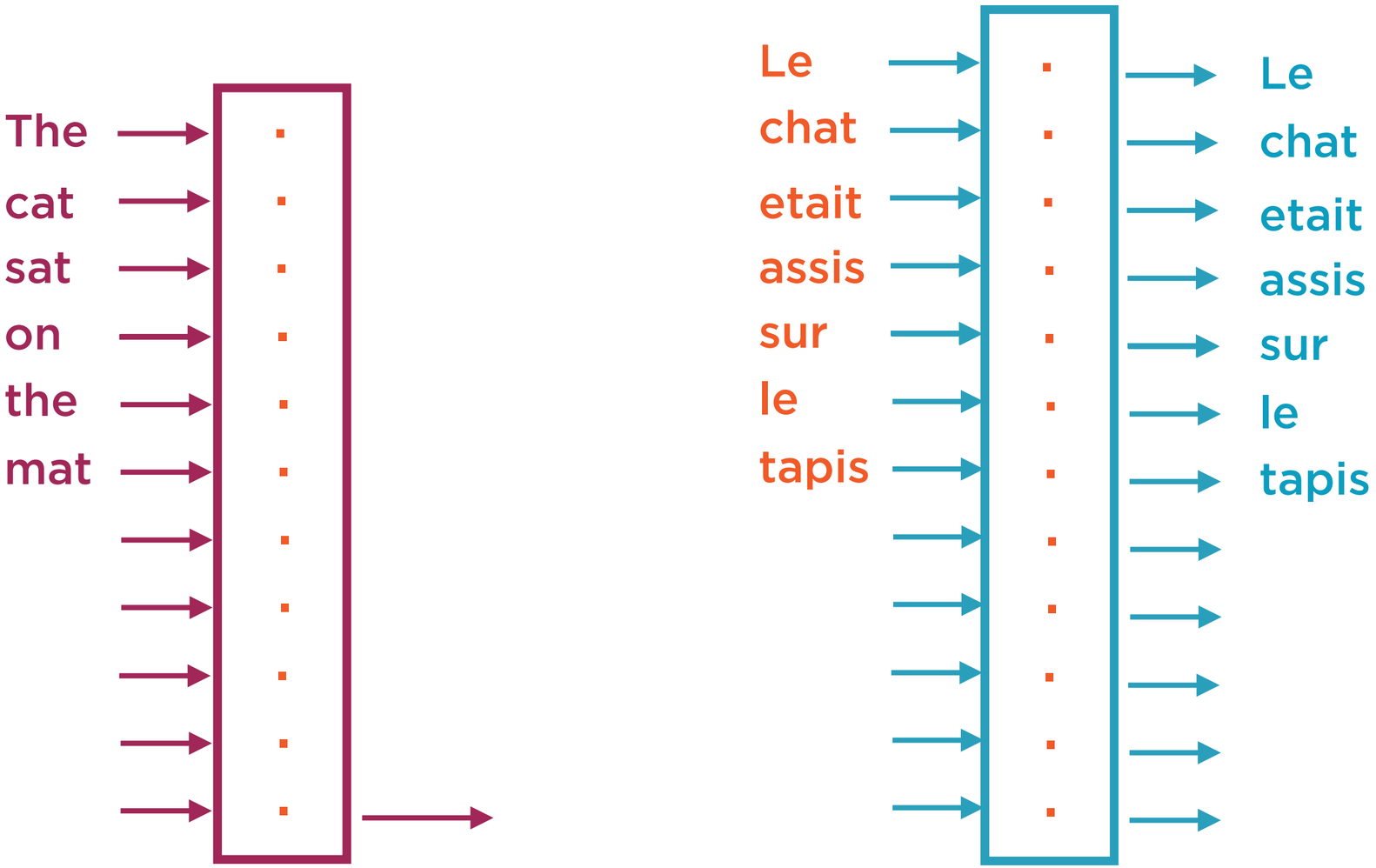


Teacher Forcing

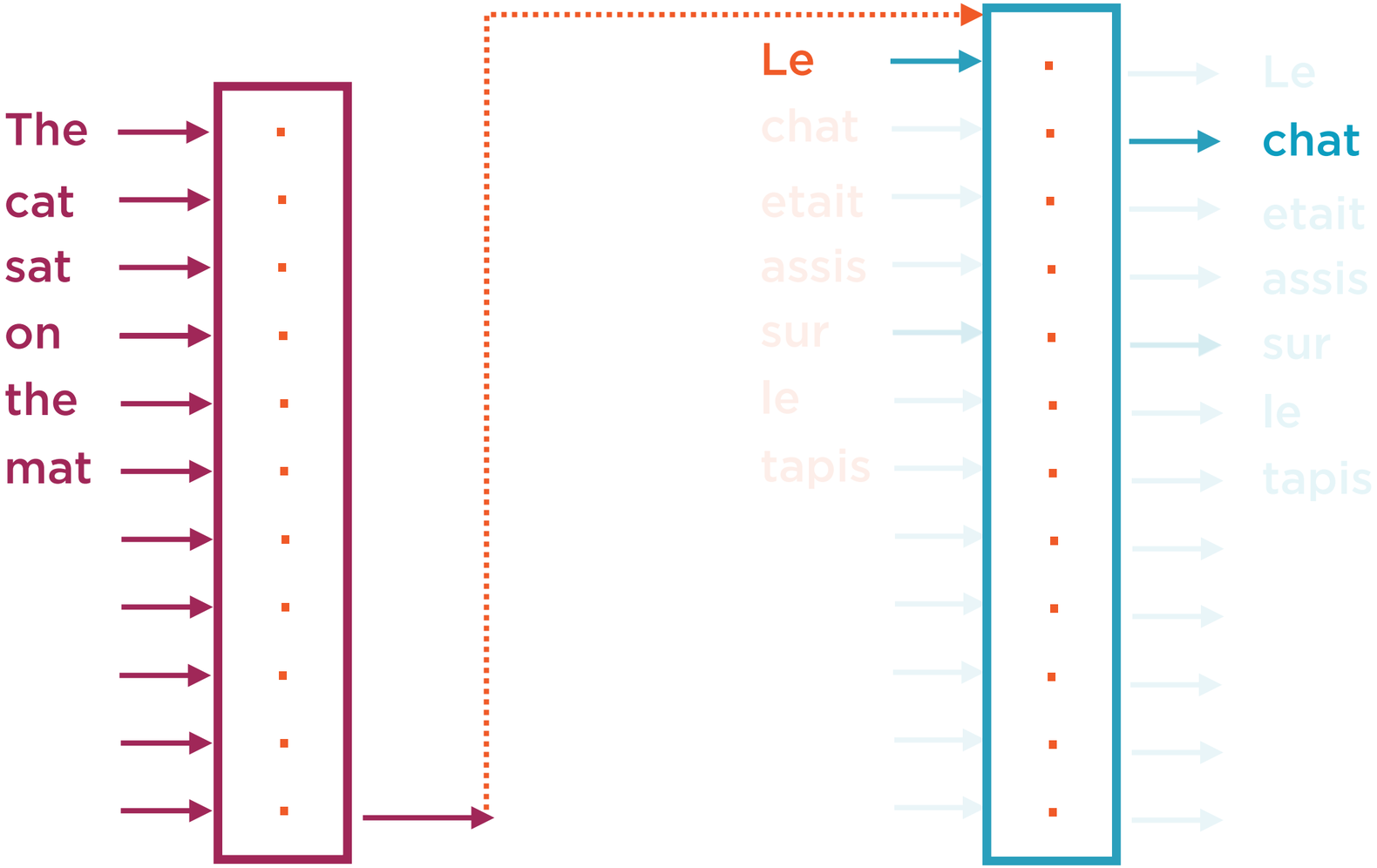


During training, we have the correct French translation available - use it a fraction of the time

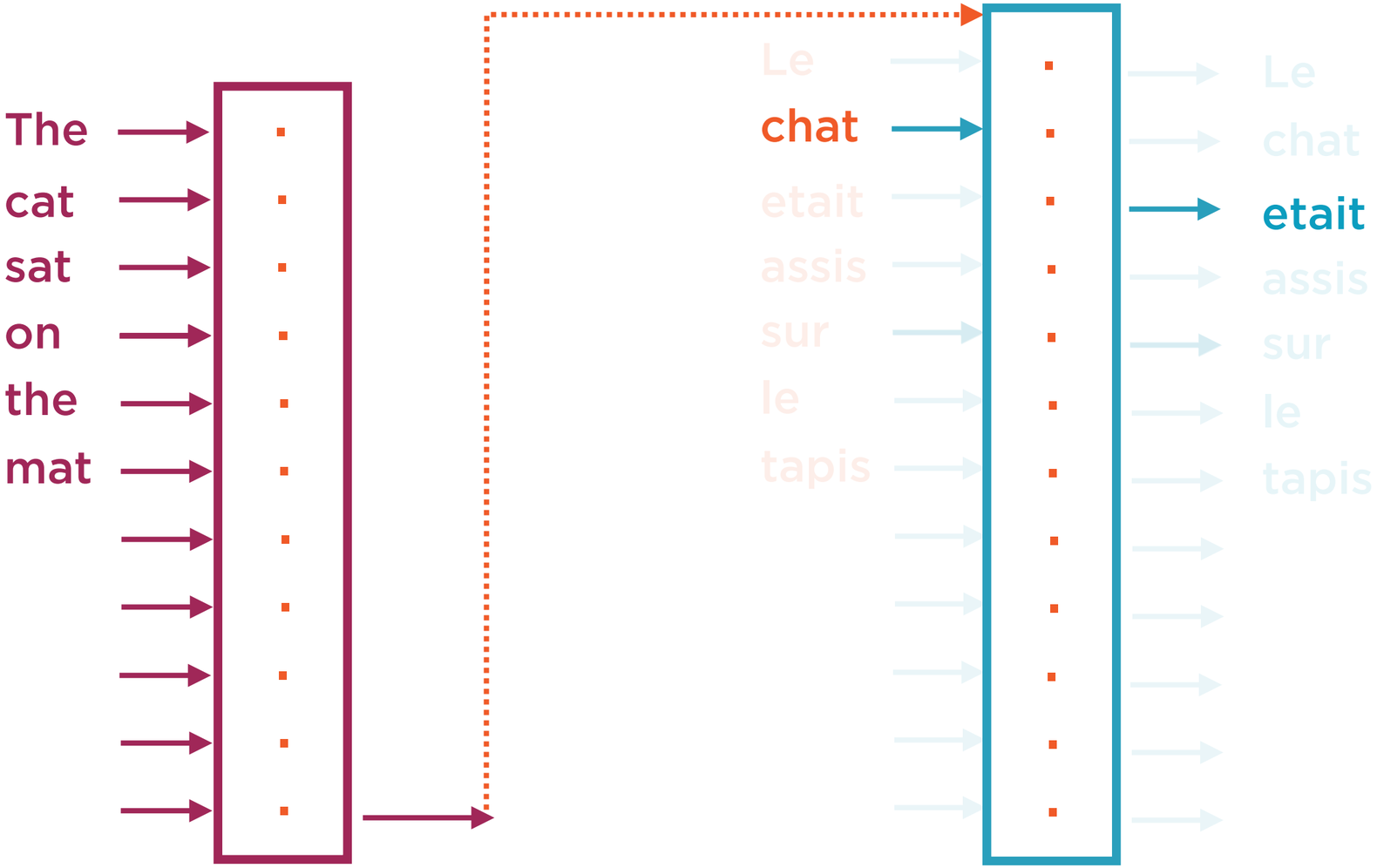
Teacher Forcing



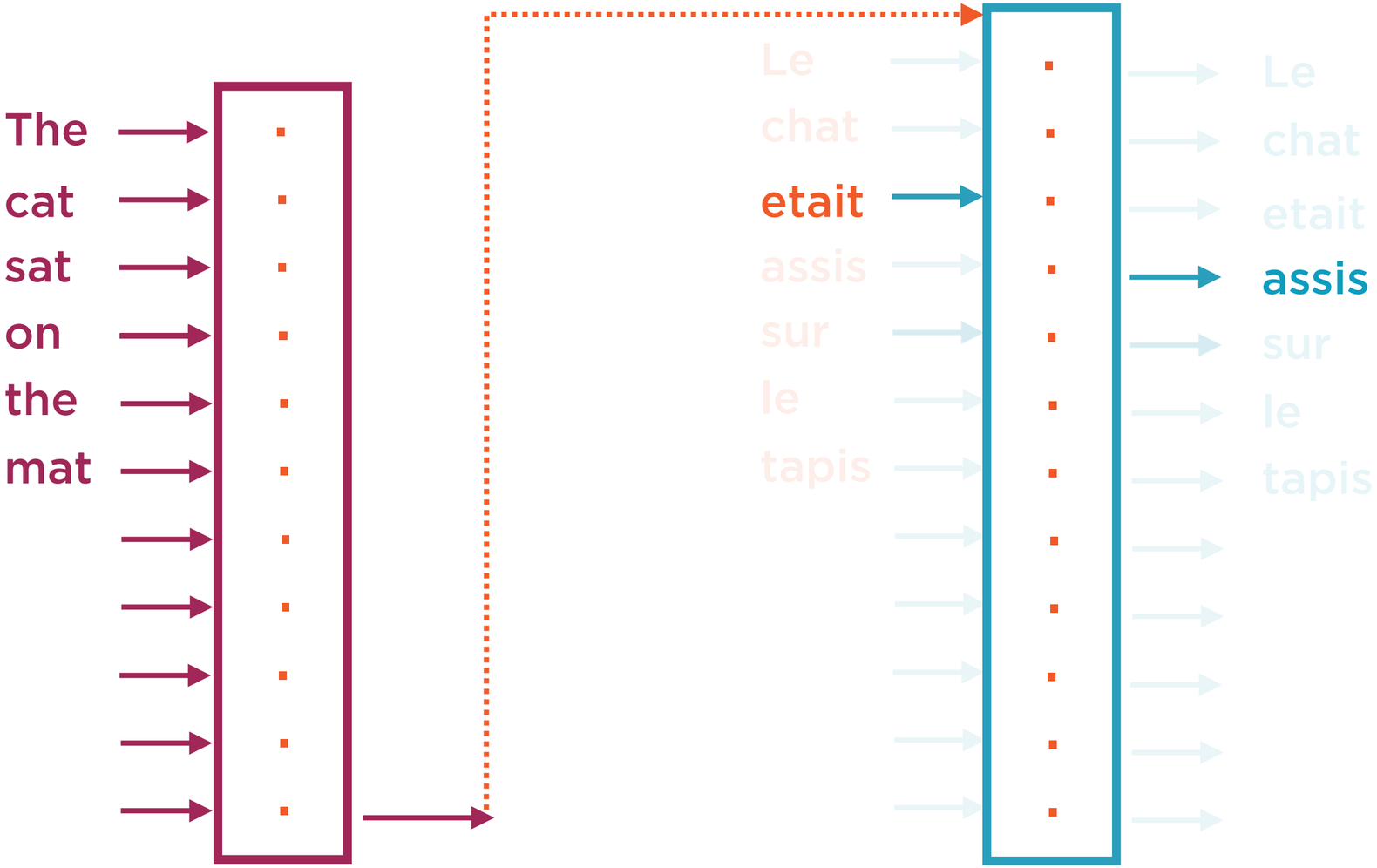
Teacher Forcing



Teacher Forcing



Teacher Forcing



Demo

Sequence-to-sequence model for language translation

Summary

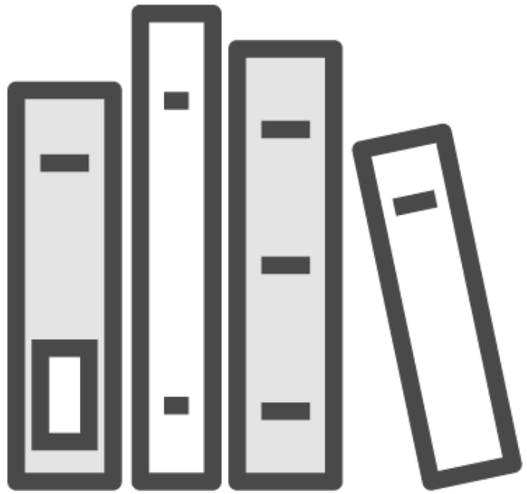
Exploring the different kinds of models built using neural networks

Contrasting seq2seq and seq2vec models

Understanding encoder-decoder architectures for language translation

Implementing language translation using RNNs

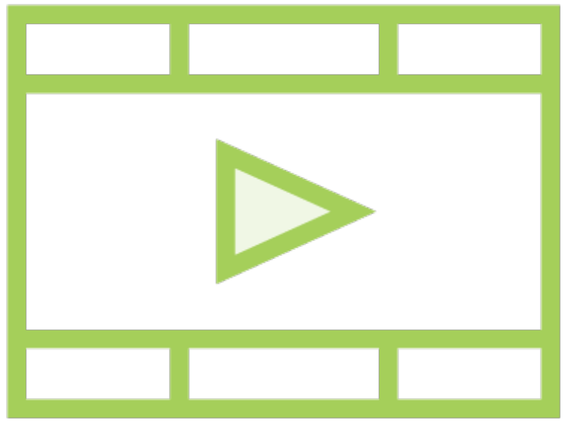
Books



Hands-On Machine Learning with Scikit-Learn and TensorFlow

Aurélien Géron

Related Courses



Building Features from Text Data

Neural Style Transfer with PyTorch

Deploying PyTorch Models in Production: PyTorch Playbook