DL Dev Course: Week 07
Seq2Seq Intro

RED DRAGON AI

Seq2Seq Model Uses

- Machine Translation
- Auto Reply
- Dialogue Systems
- Speech Recognition
- Time Series
- Chatbots
- Audio
- Image Captioning
- Q&A
- many more

Why Seq2seq

- Sequences preserve the order of the inputs
- They allow us to process information that has a time or order element to it
- They allow us to preserve information that couldn’t be done done via normal neural networks

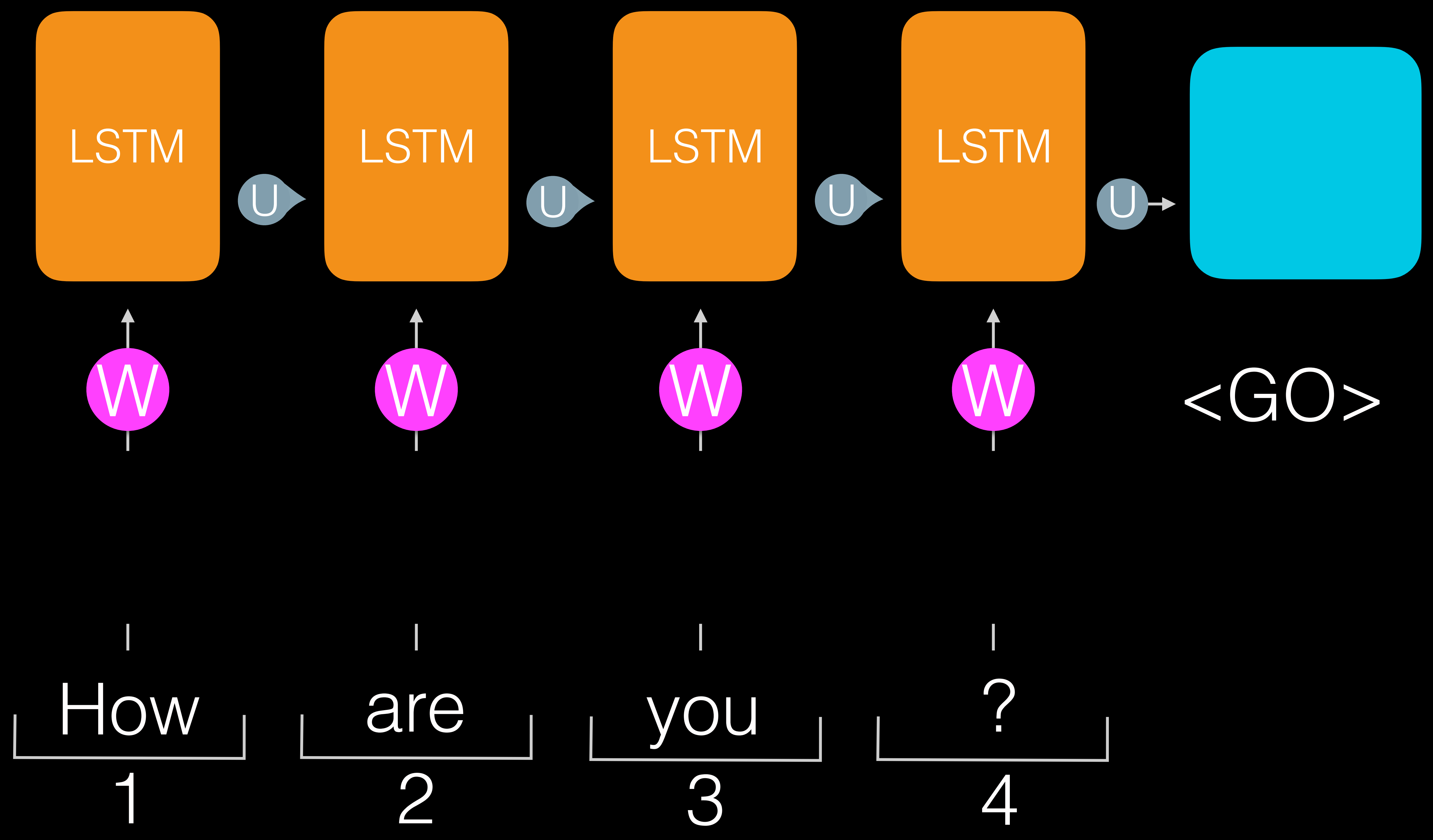
Seq2Seq Key Components

- Encoder - takes info in, in time steps then creates a hidden state to be passed to the decoder
- Decoder - takes the hidden state/states and uses that to predict the correct next step in the sequence
- <u>Lots</u> of data

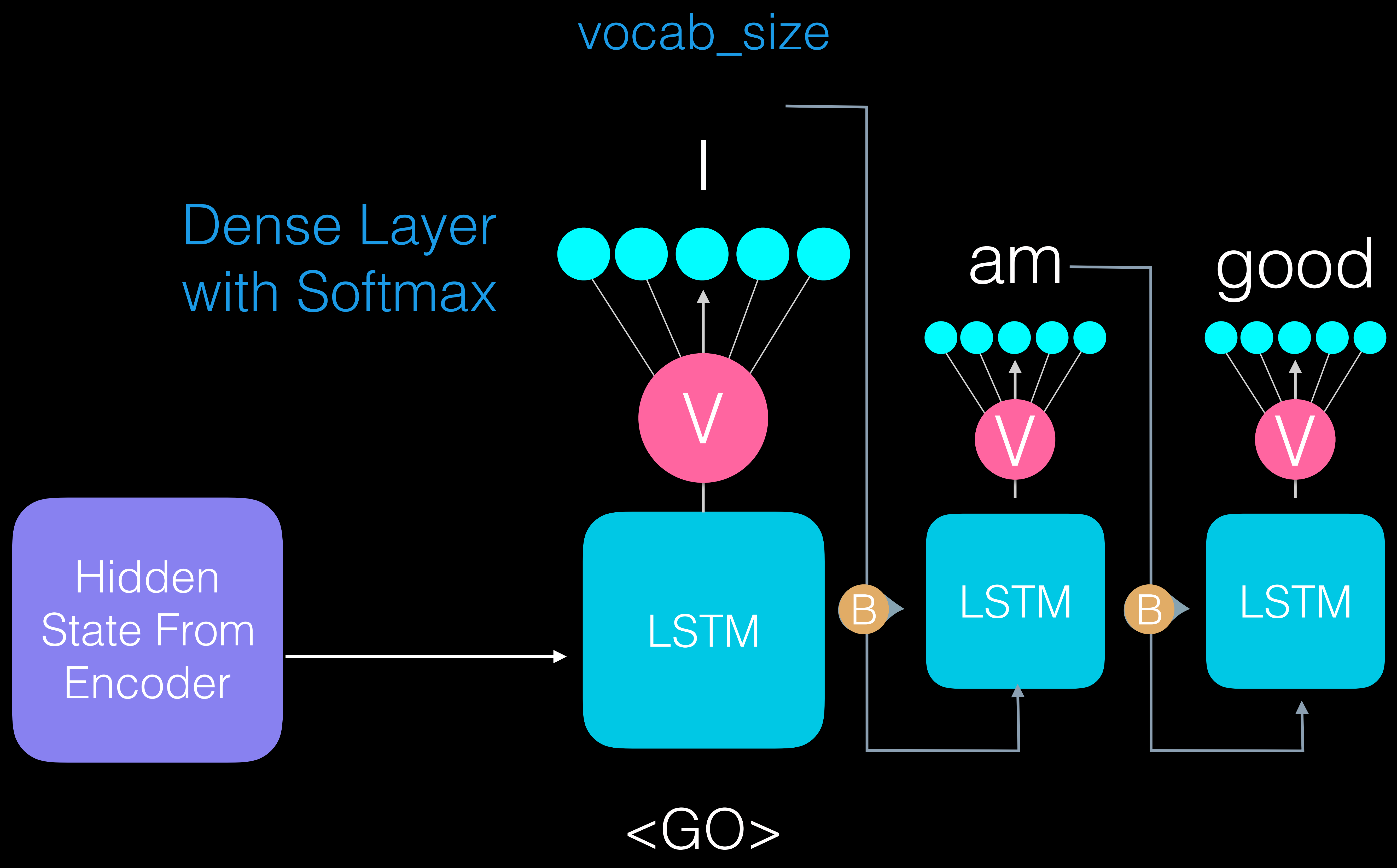
Seq2Seq Key idea

The aim is to convert a sequence into a fixed size feature vector that encodes only the important information in the sequence while losing the unnecessary information.

Encoder

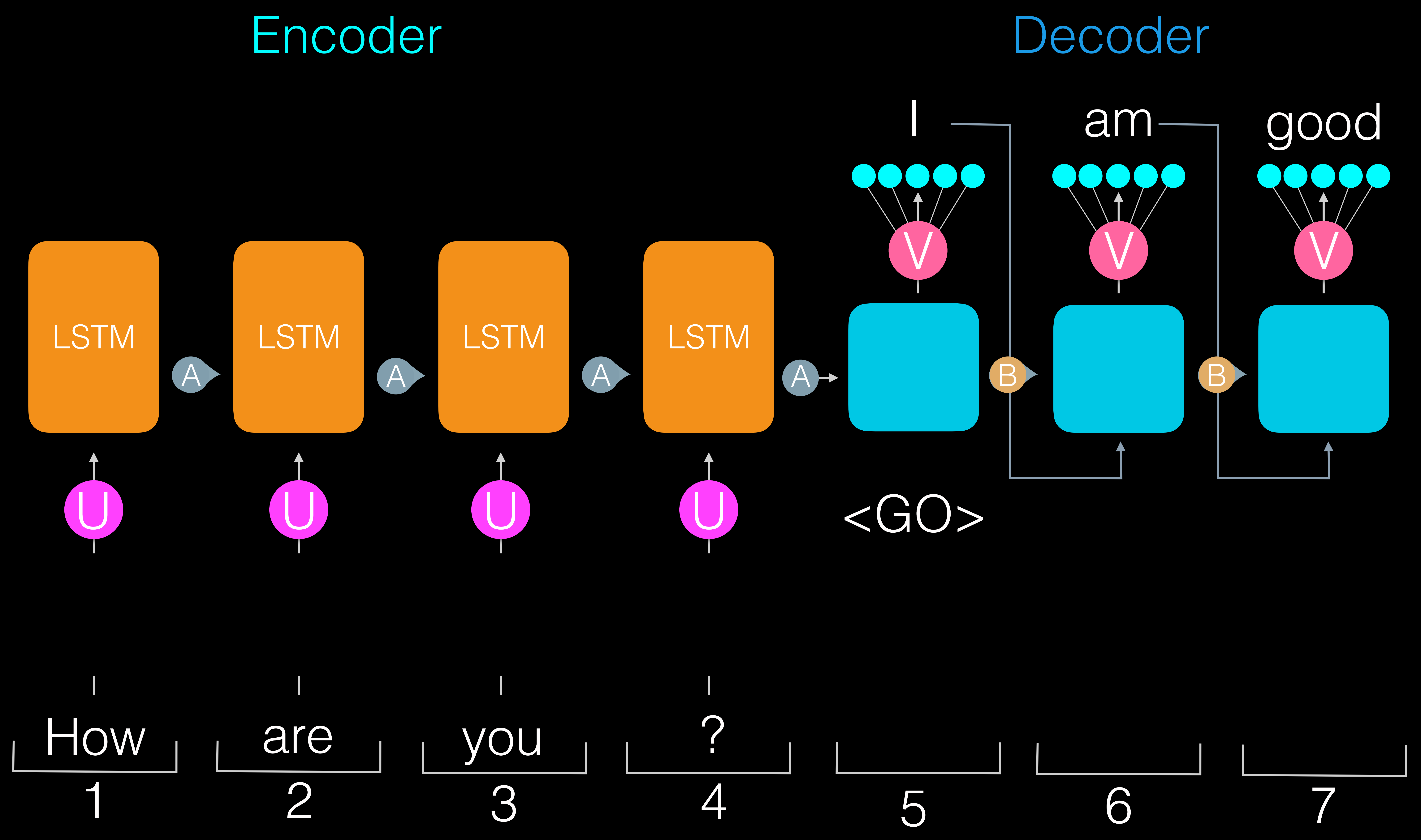


Decoder



vocab_size
I
Dense Layer
with Softmax
am
good
V
V
V
Hidden
State From
Encoder
LSTM
B
LSTM
B
LSTM
<GO>

Seq2Seq



Often La

What is padding

No padding

Hello how are you today ?

I am fine

Padded Length 8

‘Hello’ ‘how’ ‘are’ ‘you’ ‘?’ ‘<pad>’ ‘<pad>’ ‘<pad>’

‘I’ ‘am’ ‘fine’‘<pad>’ ‘<pad>’ ‘<pad>’ ‘<pad>’ ‘<pad>’

All input sequences must be same length as each other.

All output sequences must be same length as each other.

Special Tokens

<PAD> padded zero input

<EOS> end of sentence

<GO> telling the decoder to start

<OOV> out of vocabulary

<UNK> unknown

<ES2> language to translate to

Lookup tables

- We can’t pass words directly to the network
- We have to assign each word to an index number

‘Hello’, ‘how’, ‘are’, ‘you’, ‘?’, ‘<pad>’, ‘<pad>’, ‘<pad>’

23,4,13,14,8,0,0,0

‘Hello’, ‘you’, ‘!’,’<pad>’, ‘<pad>’, ‘<pad>’, ‘<pad>’, ‘<pad>’

23,14,23,0,0,0,0,0

Lookup tables

- We often discard words that are used very little and replace them with unknown <UNK>
- The bigger the vocabulary we have the harder it will be at prediction time, so we want the words that are used the most.

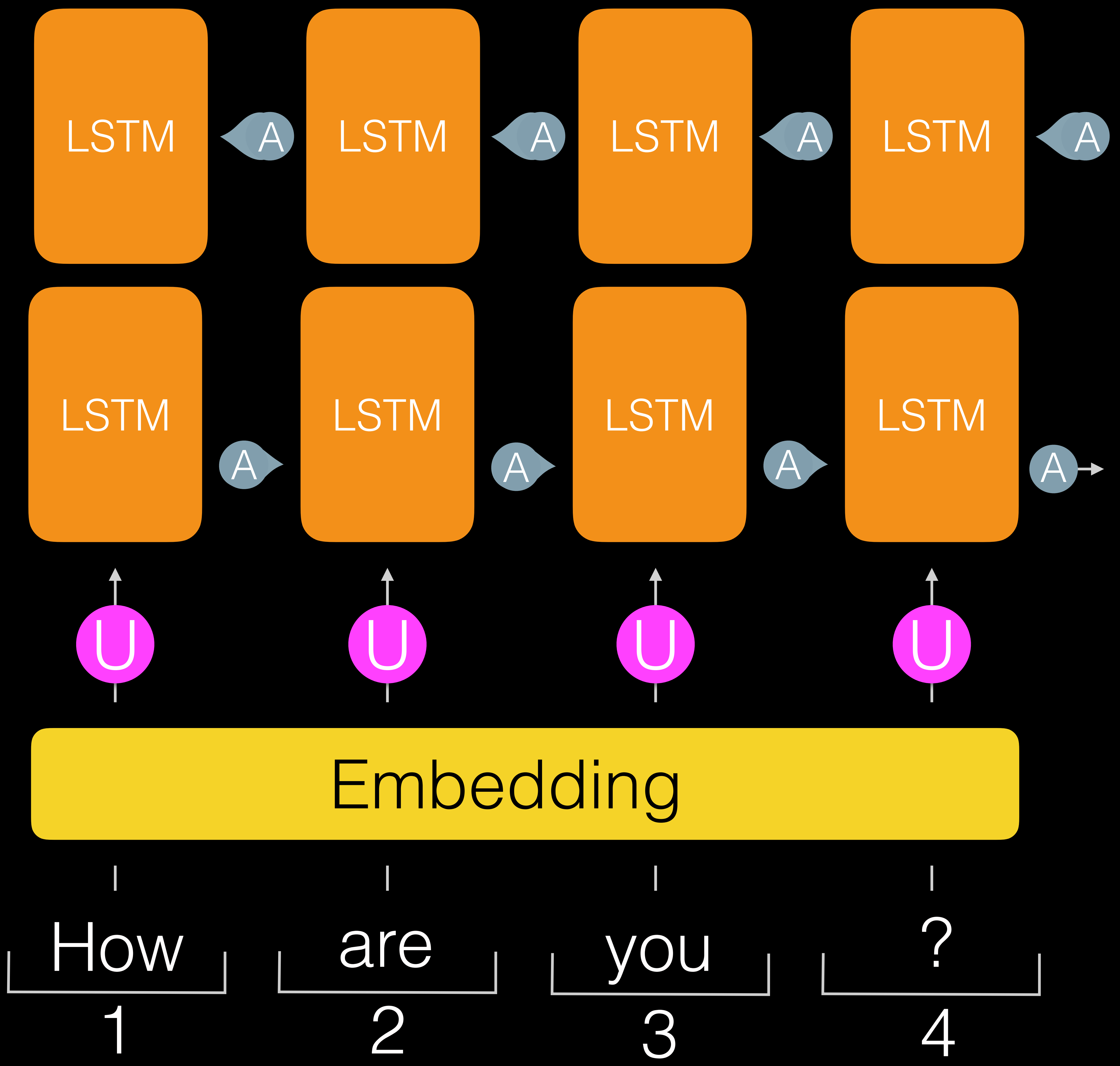
Code time

Embeddings

- Embeddings allow us to extract more semantic meaning from the words
- Embeddings like Word2Vec have been trained on billions of words and have abstracted a lot of the meaning from those words.

Seq2Seq

Bidirectional Encoder



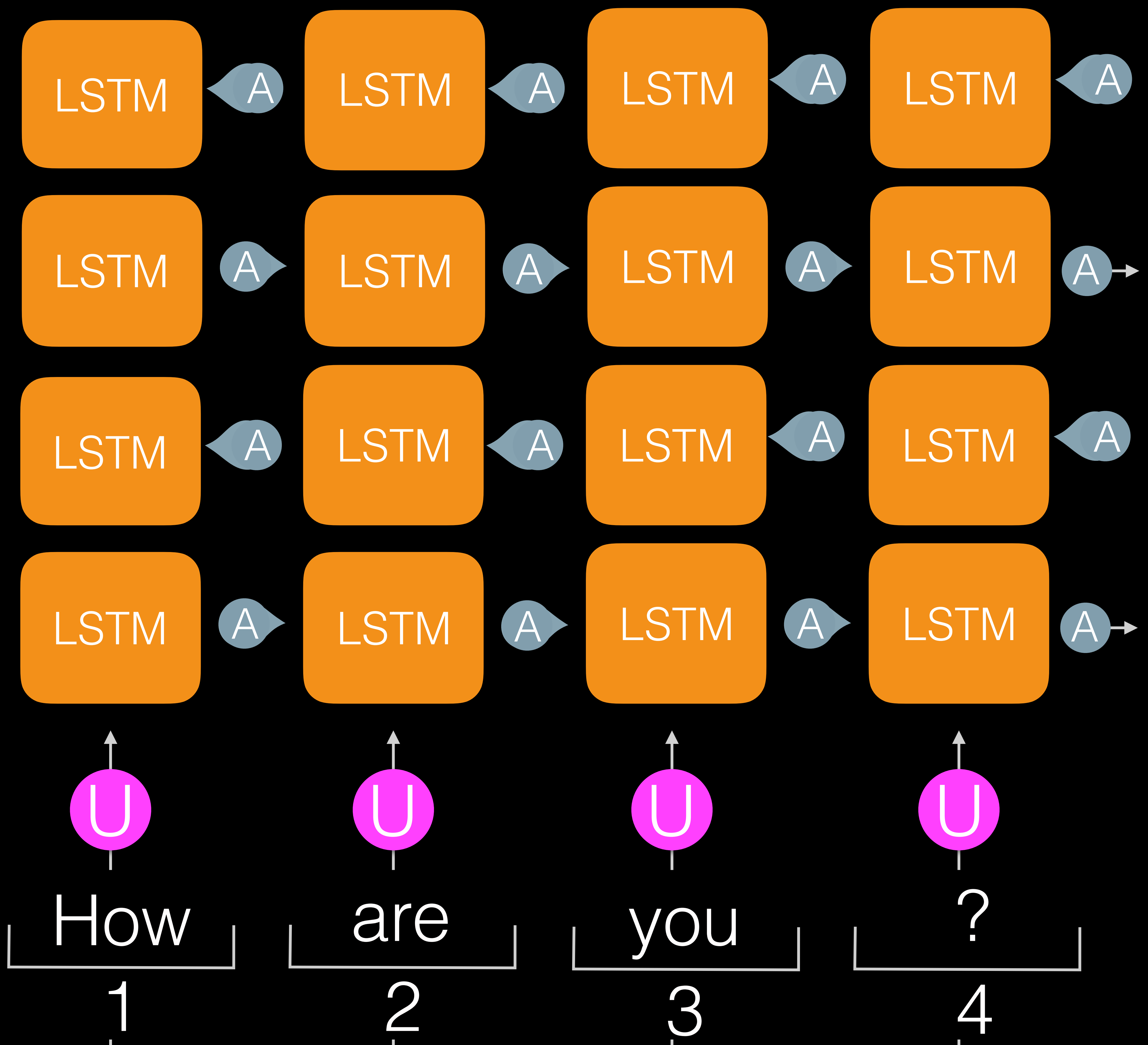
Often La

The BiLSTM Secret

- Bidirectional LSTMs generally work better than anything else for almost every NLP task
- Often the more BiLSTMs the better
- State of the art is usually BiLSTMs with Attention
- State of the art is still often lacking

Stacked

Bidirectional Encoder



Often La

Inputs

- Glove embedding = 100 dim
- Max sequence length = 30
- Batch size = 64
- Tensor input shape = (64,30,100)

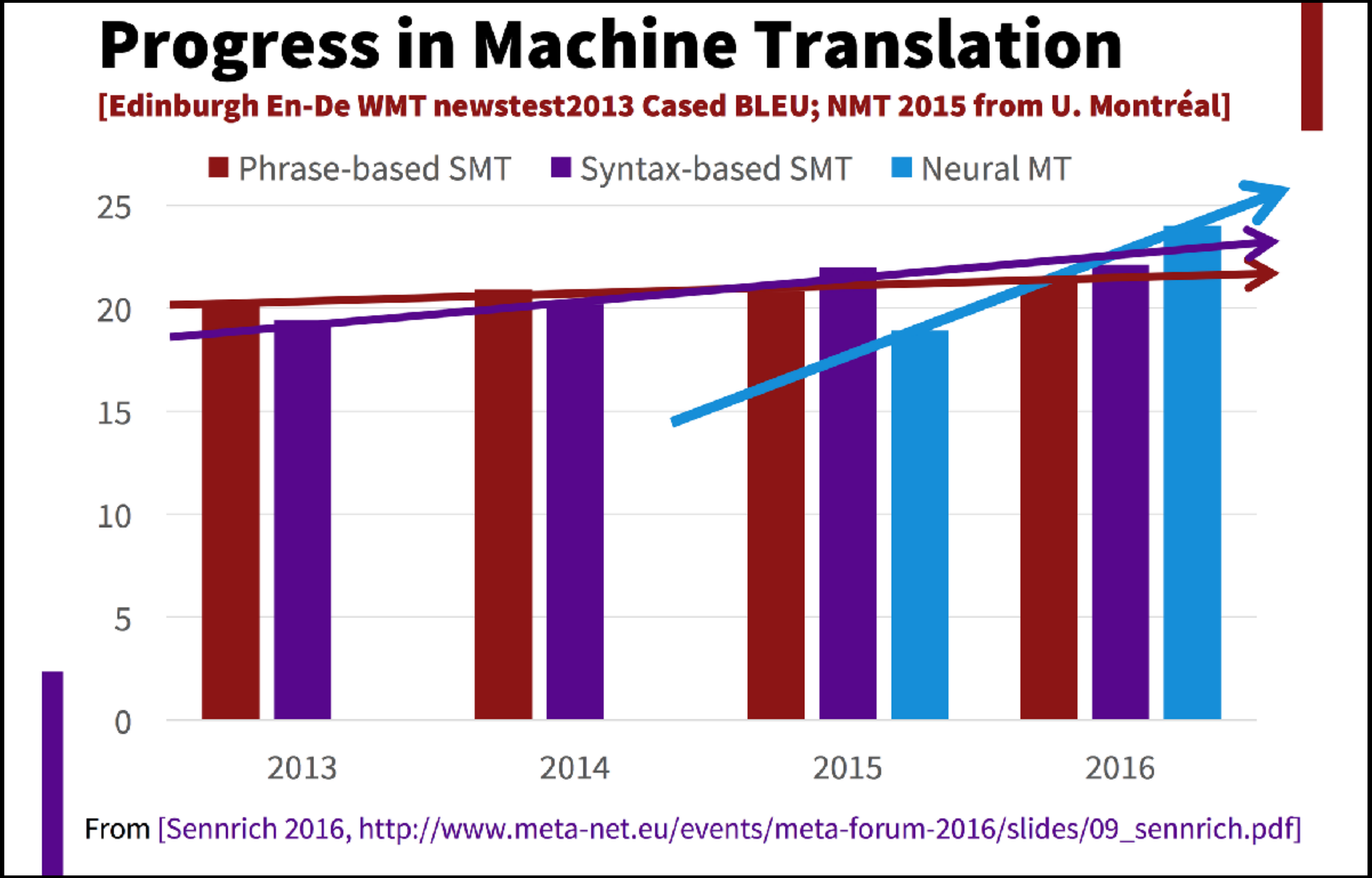
Language Translation

- USD$40 Billion a year
- Google translates over a 100 billion words a day
- Facebook has been working on it’s own systems
- Ecommerce etc etc

Why is translation hard?

- The correct word to use depends on other words in the sentence
- Order of words can change in different languages
- Rules don’t work, need to use statistical approaches
- Traditional SMT was really complicated

Why NMT



Chris Manning - Stanford Course

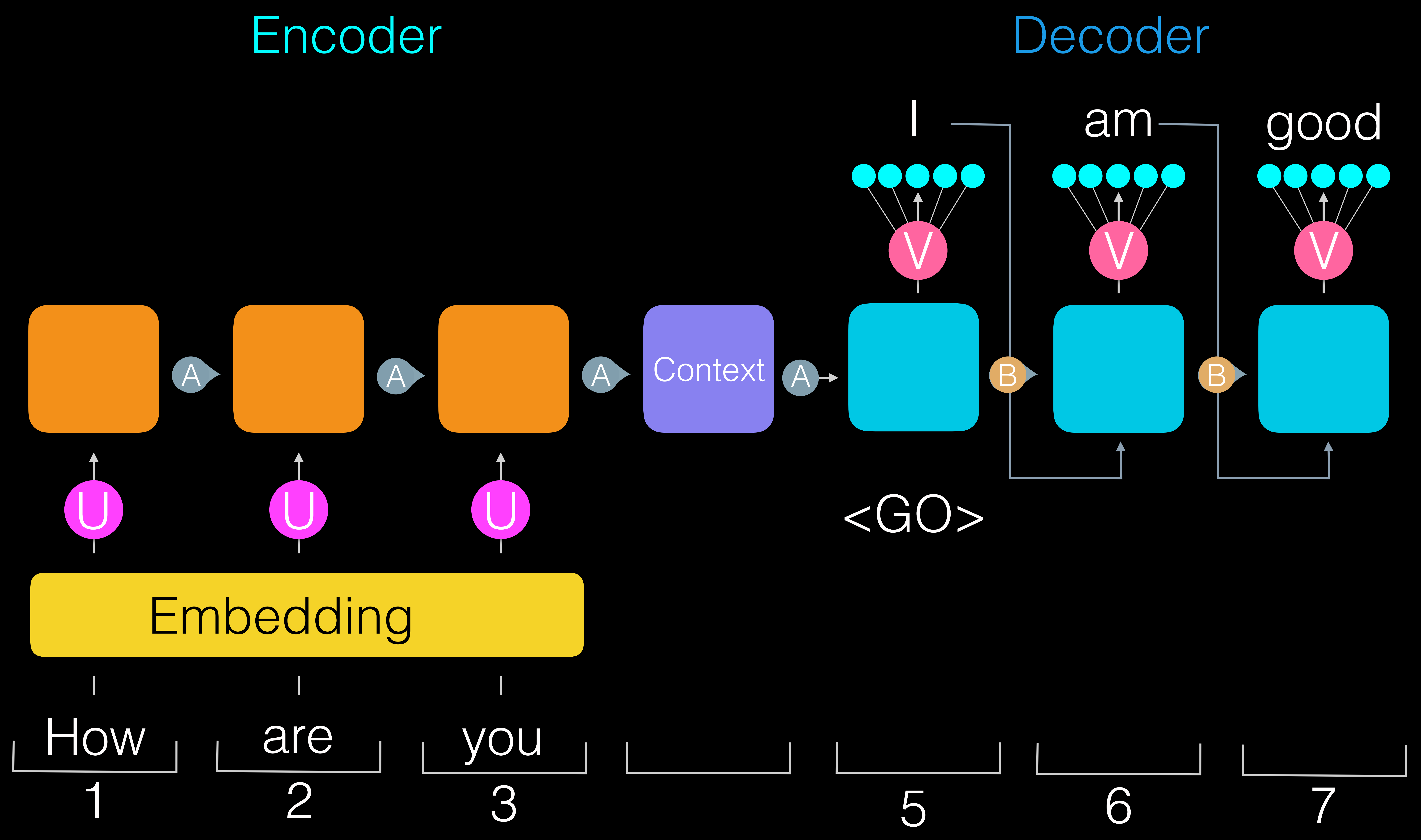
Any to Any in the past

- Google ~ 80 languages
- 6400 MT systems Bilingual systems
- Interlingua 80 Encoders 80 Decoders

<ES2> translate to Spanish

<IT2> translate to Italian

NMT


Encoder
Decoder
I
am
good
V
V
V
A
A
A
Context
A
B
B
U
U
U
<GO>
Embedding
How
are
you
1
2
3
5
6
7

Often La

Vanilla Seq2Seq Problems

- Works well on short sentences but not long ones
- LSTMs can remember out to about 30 steps
- Drops off very quickly after 30

Advanced Seq2Seq

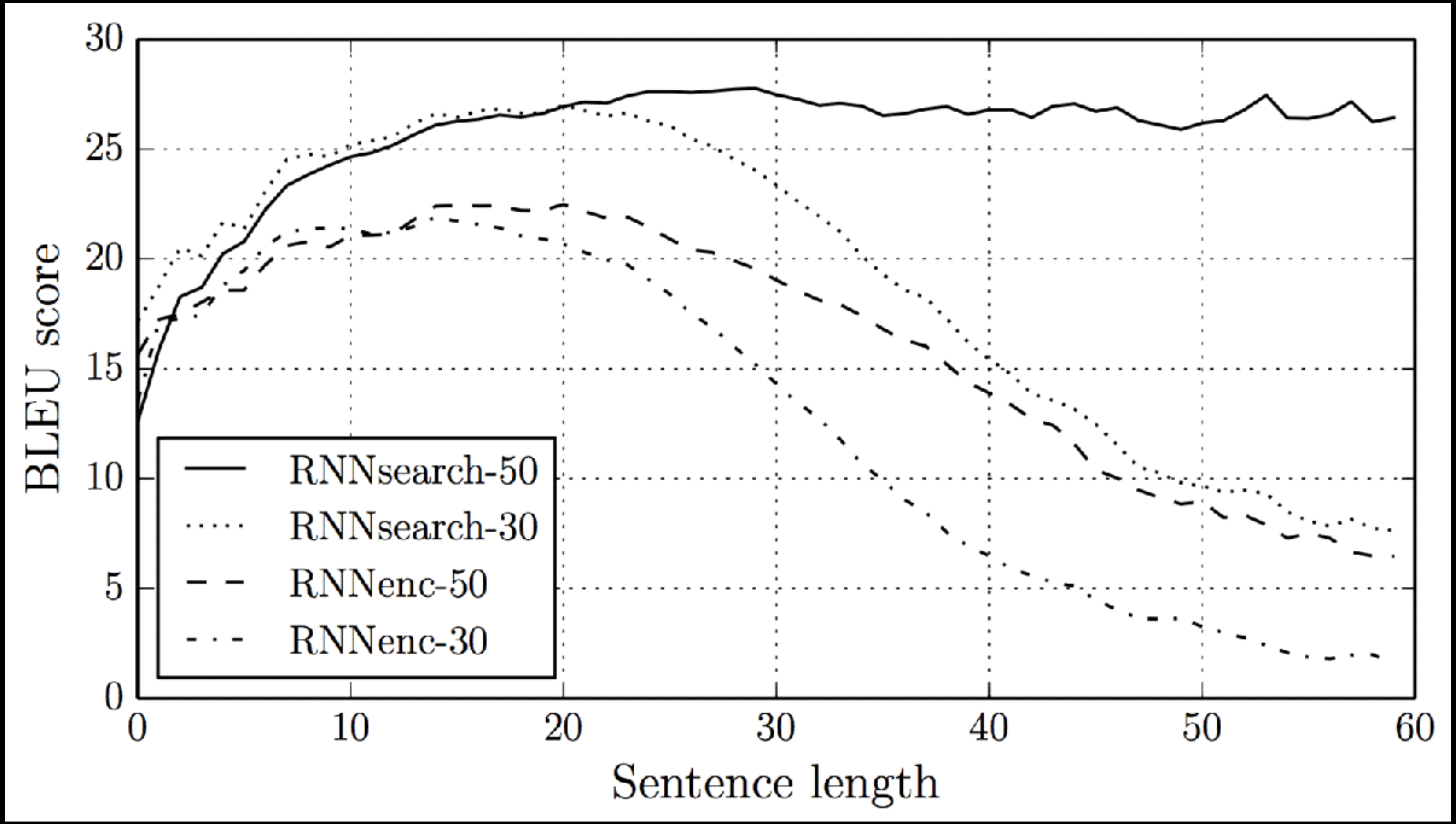
Attention

Teacher Forcing

Peeking

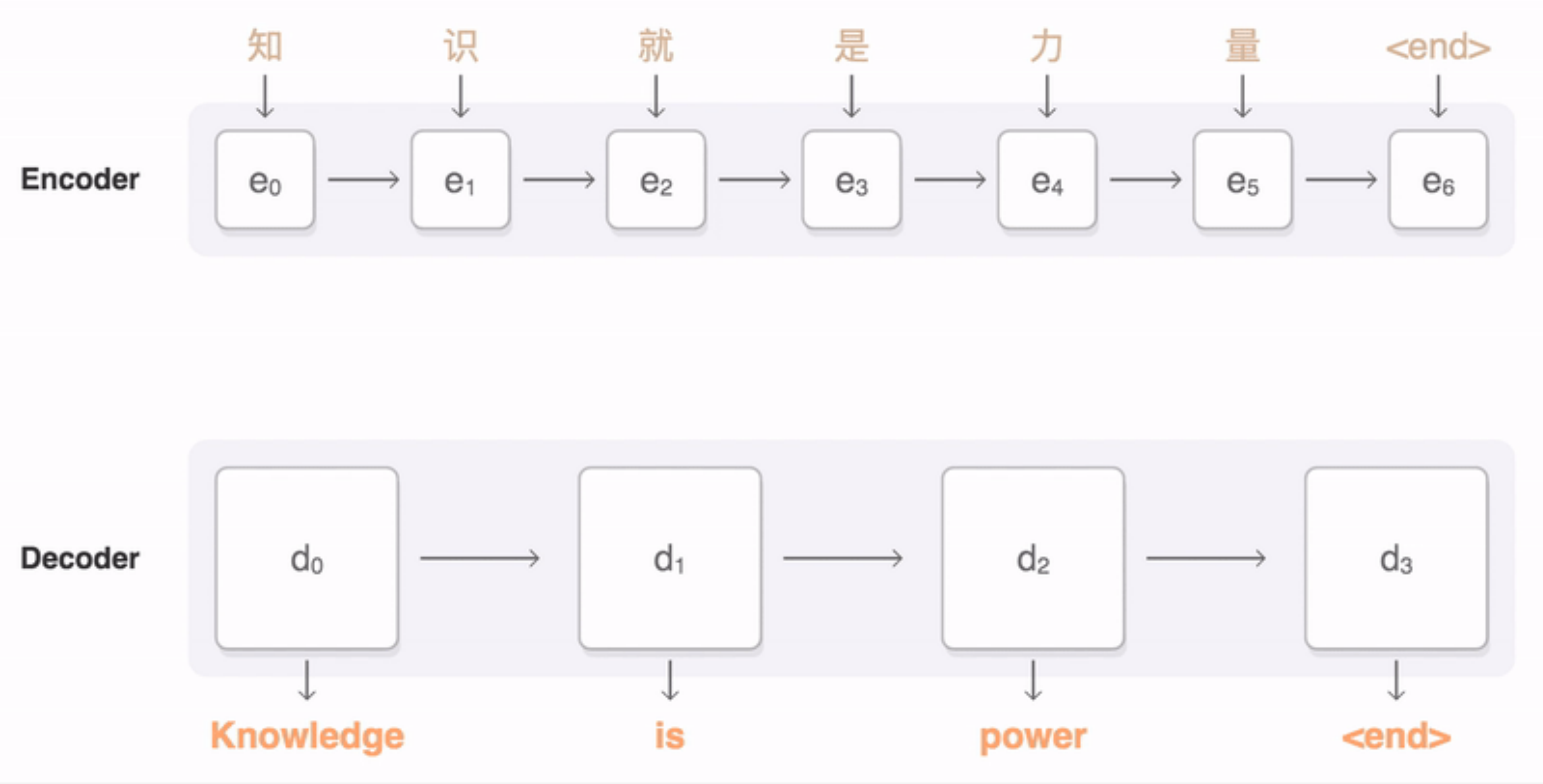
Beam Search

Attention



30
25
20
15
10
5
0
BLEU score
0
10
20
30
40
50
60
Sentence length
RNNsearch-50
RNNsearch-30
RNNenc-50
RNNenc-30

Attention



知
识
就
是
力
量
<end>
Encoder
e_0
e_1
e_2
e_3
e_4
e_5
e_6
Decoder
d_0
d_1
d_2
d_3
Knowledge
is
power
<end>

Attention

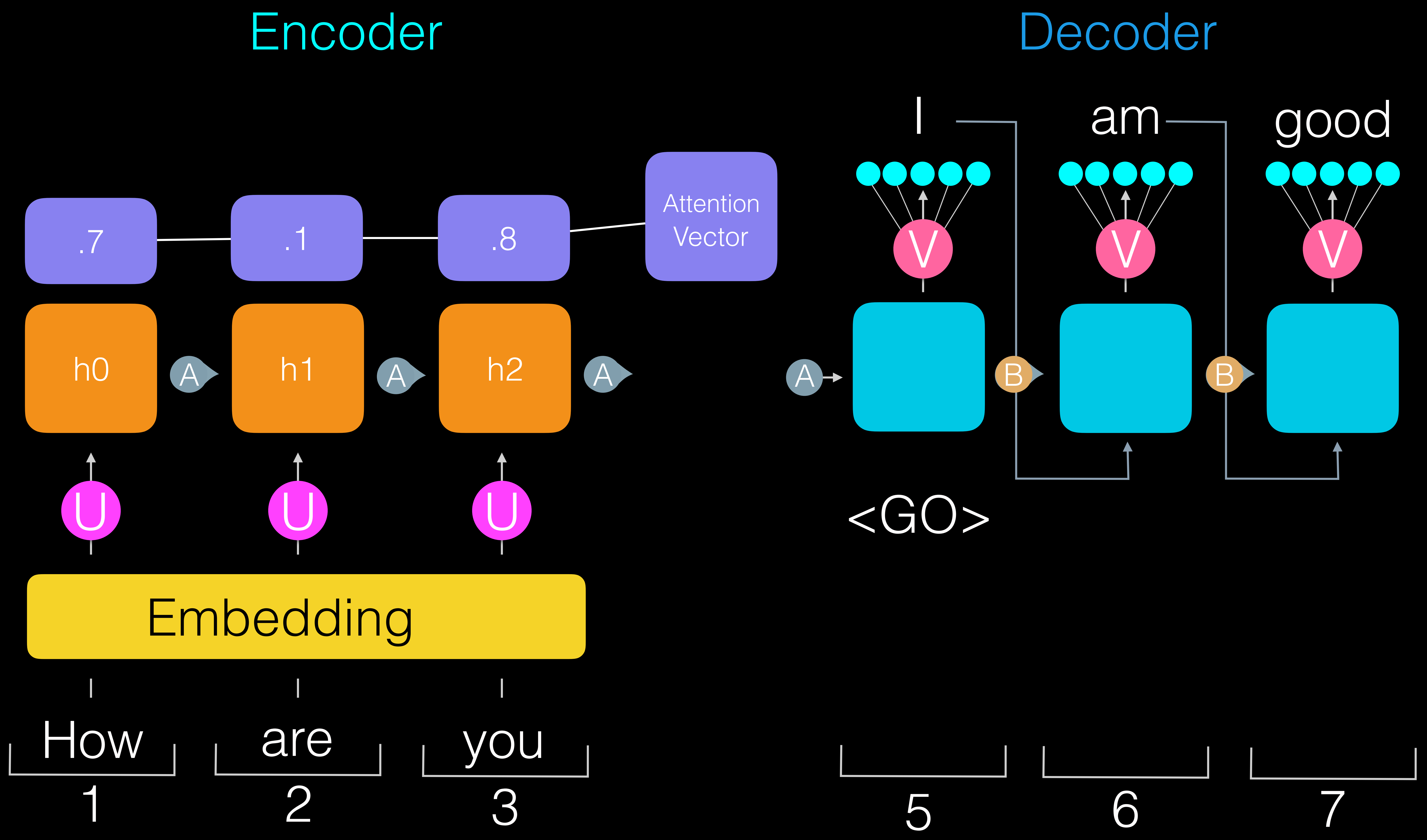
- It even does better for short sentence length
- NMT without attention often generate sentences with good grammar but gets the name wrong or repeats itself
- Attention gives us like a fixed vector of RAM to score the words

What words are important?

Last Friday David’s team went out but the others stayed in

Last Friday David’s team went out but the others stayed in

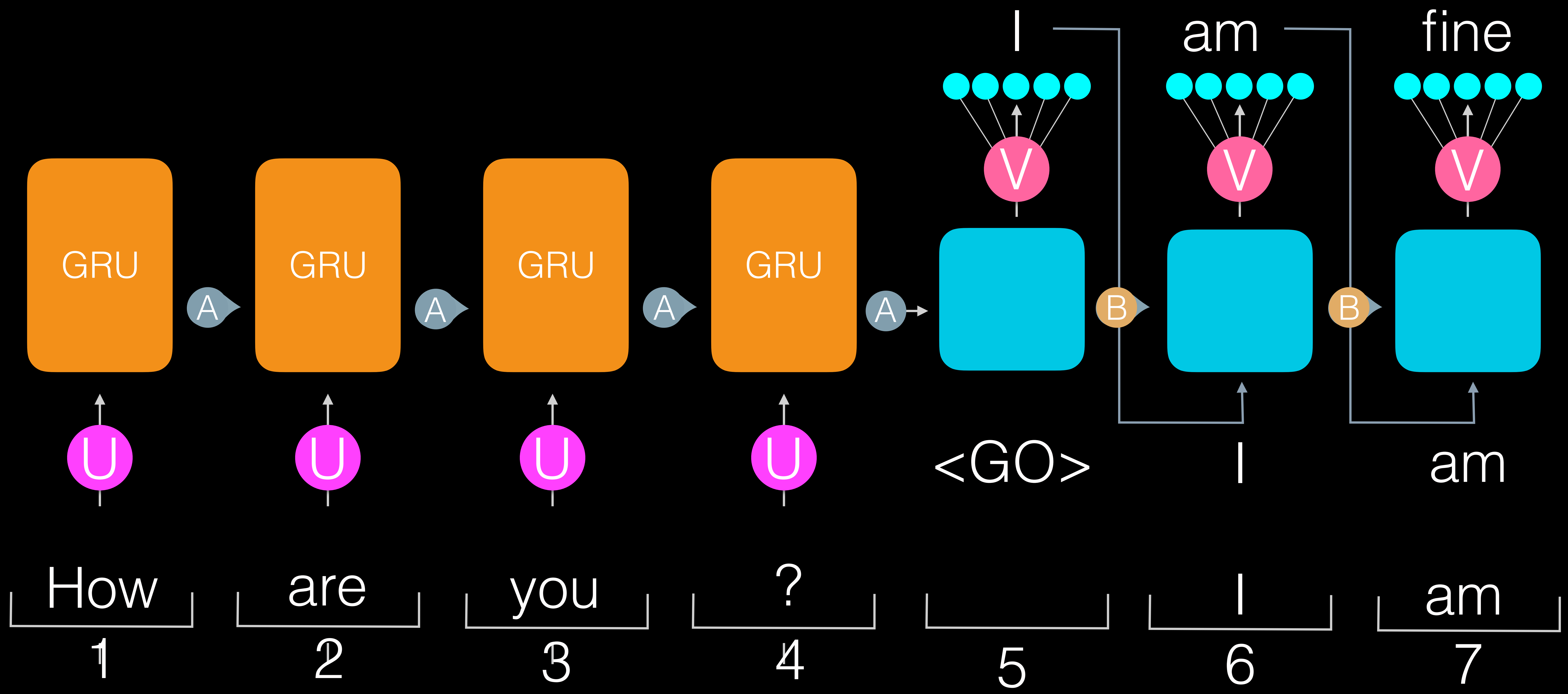
Attention Scoring



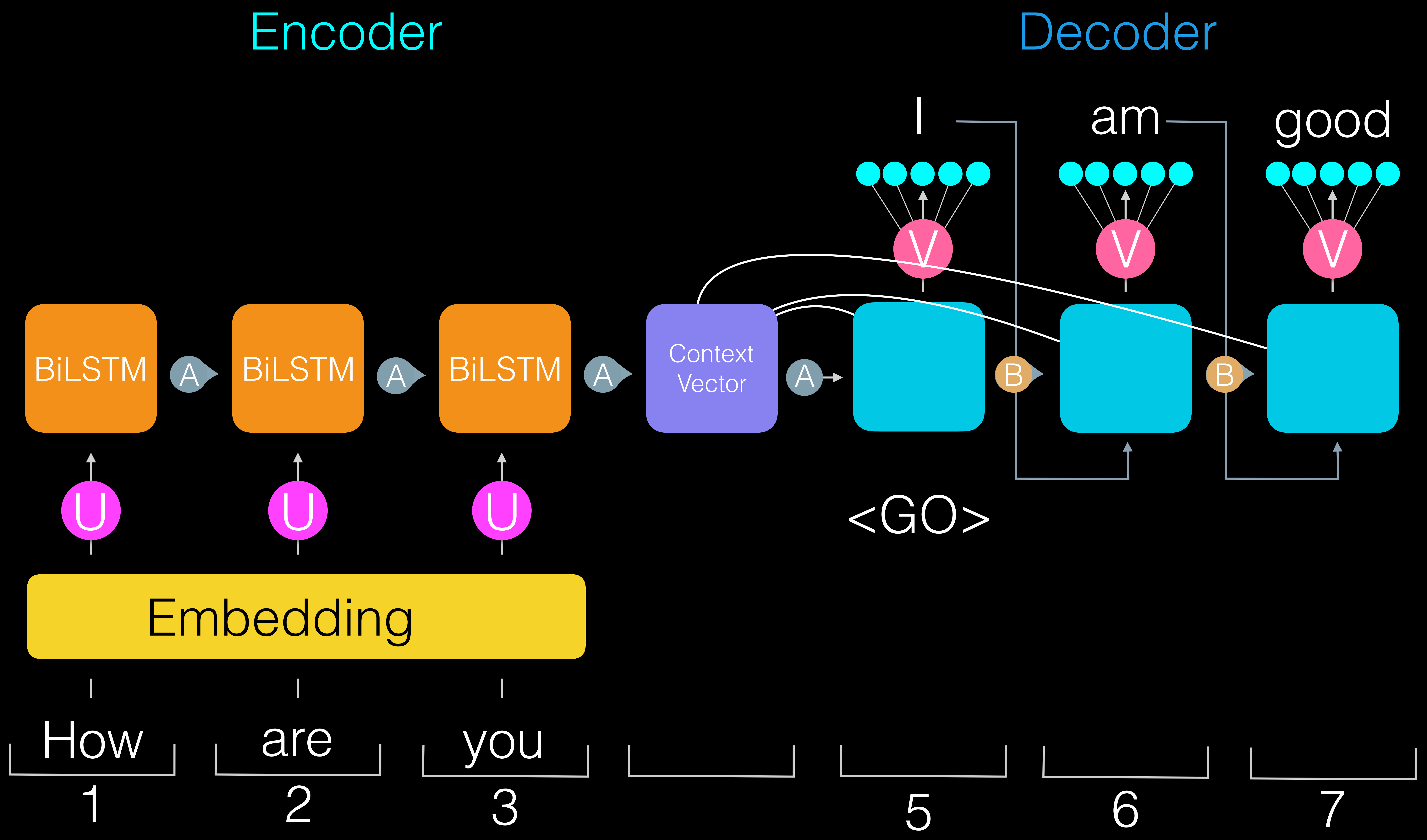
Often La

Teacher Forcing

When training the network instead of letting the decoder pass its predictions to the next layer, pass the correct word/state

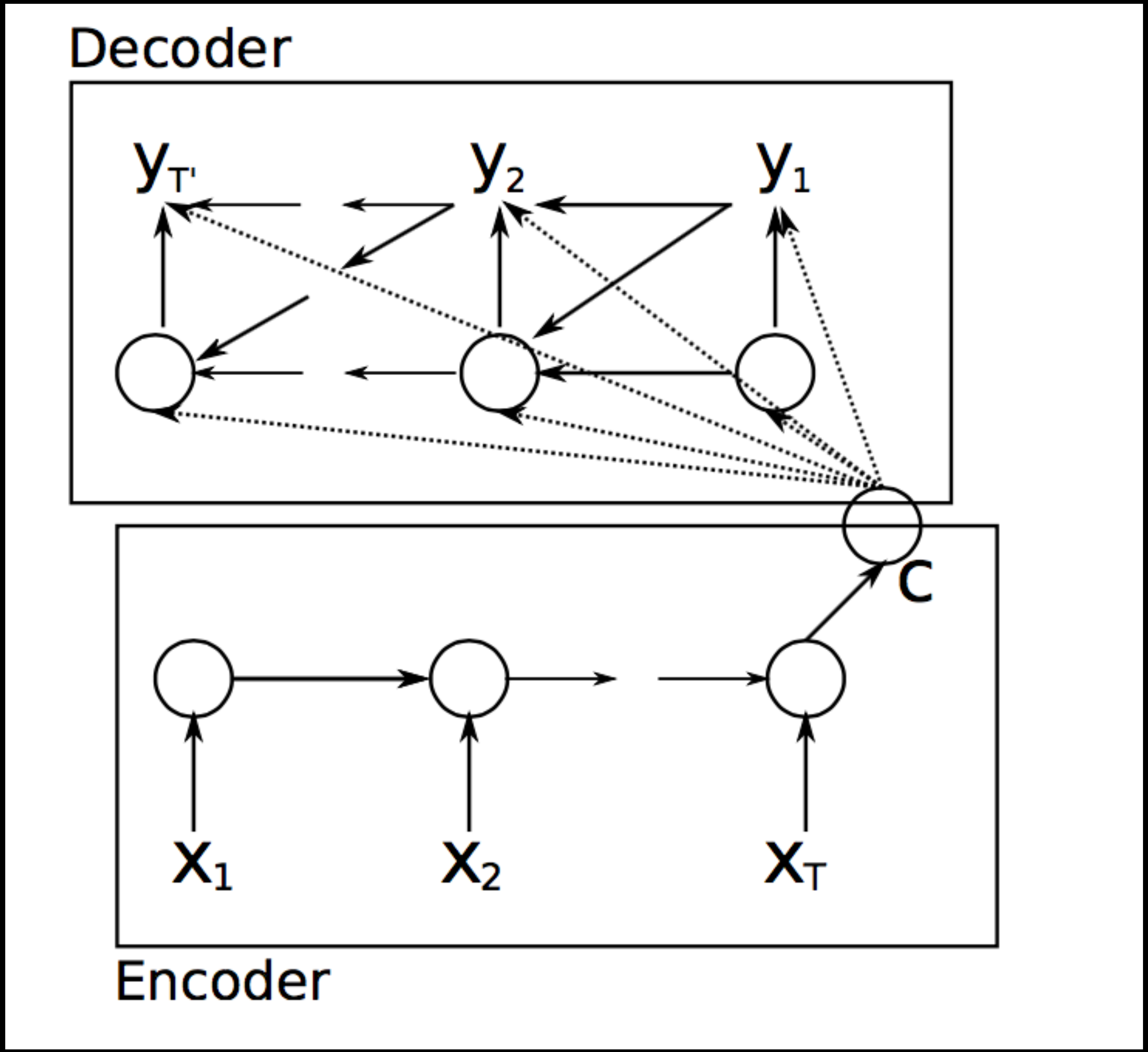


Peeking



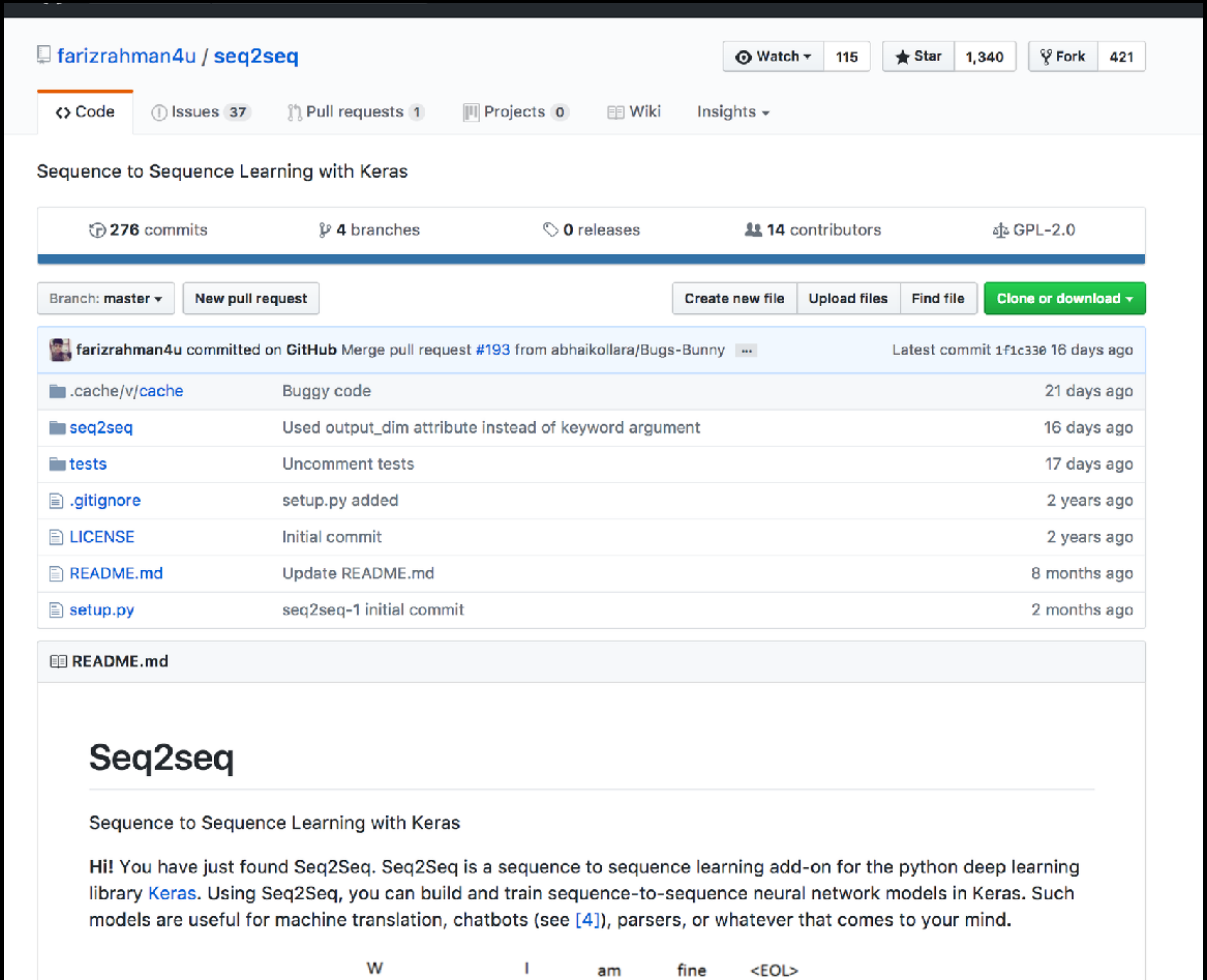
Often La

Peeking



Often La

Keras Resources



farizrahman4u / seq2seq

Watch 115 | Star 1,340 | Fork 421

Code | Issues 37 | Pull requests 1 | Projects 0 | Wiki | Insights

Sequence to Sequence Learning with Keras

276 commits | 4 branches | 0 releases | 14 contributors | GPL-2.0

Branch: master | New pull request | Create new file | Upload files | Find file | Clone or download

farizrahman4u committed on GitHub Merge pull request #193 from abhaikollara/Bugs-Bunny ... Latest commit 1f1c330 16 days ago

.cache/v/cache	Buggy code	21 days ago
seq2seq	Used output_dim attribute instead of keyword argument	16 days ago
tests	Uncomment tests	17 days ago
.gitignore	setup.py added	2 years ago
LICENSE	Initial commit	2 years ago
README.md	Update README.md	8 months ago
setup.py	seq2seq-1 initial commit	2 months ago

README.md

Seq2seq

Sequence to Sequence Learning with Keras

Hi! You have just found Seq2Seq. Seq2Seq is a sequence to sequence learning add-on for the python deep learning library Keras. Using Seq2Seq, you can build and train sequence-to-sequence neural network models in Keras. Such models are useful for machine translation, chatbots (see [4]), parsers, or whatever that comes to your mind.

W I am fine <EOL>

Resources

- https://www.nytimes.com/2016/12/14/magazine/the-great-ai-awakening.html
- http://distill.pub/2016/augmented-rnns/
- Chris Manning - NLP with Deep Learning Stanford
- Quoc Le - Seq2Seq Deep Learning: https://www.youtube.com/watch?v=G5RY_SUJih4

Papers

- Grammar as a foreign language -Neubig
- Google’s Neural Machine Translation System : Bridging the gap - Wu et al
- Google’s multilingual MNT System: enabling zero-shot translation
- NMT and sequence to sequence models: a tutorial-Neubig