Machine translation

Content

- What is NLP
- Machine Translation Introduction
- Machine Translation Evaluation
- Moses
- Word Embeddings
- RNN
- LSTM
- TensorFlow Machine Translation Tutorial

What is NLP

- NOT neuro-linguistic programming
- Natural Language Processing

What is NLP

- POS-tagging
- Sentence parsing
- Sentiment analysis
- Spam filtering
- Topic categorization
- Machine translation
- Text summarization
- Named entity recognition
- Natural language understanding, text-to-speech, speech recognition, question answering......

Machine Translation

- It can be done - we are doing it
- We don’t know how we do it
- We need a lot less examples to learn a language than a neural network

Stuffed blackboard erasers

Stuffed blackboard erasers - Пълнени гъби

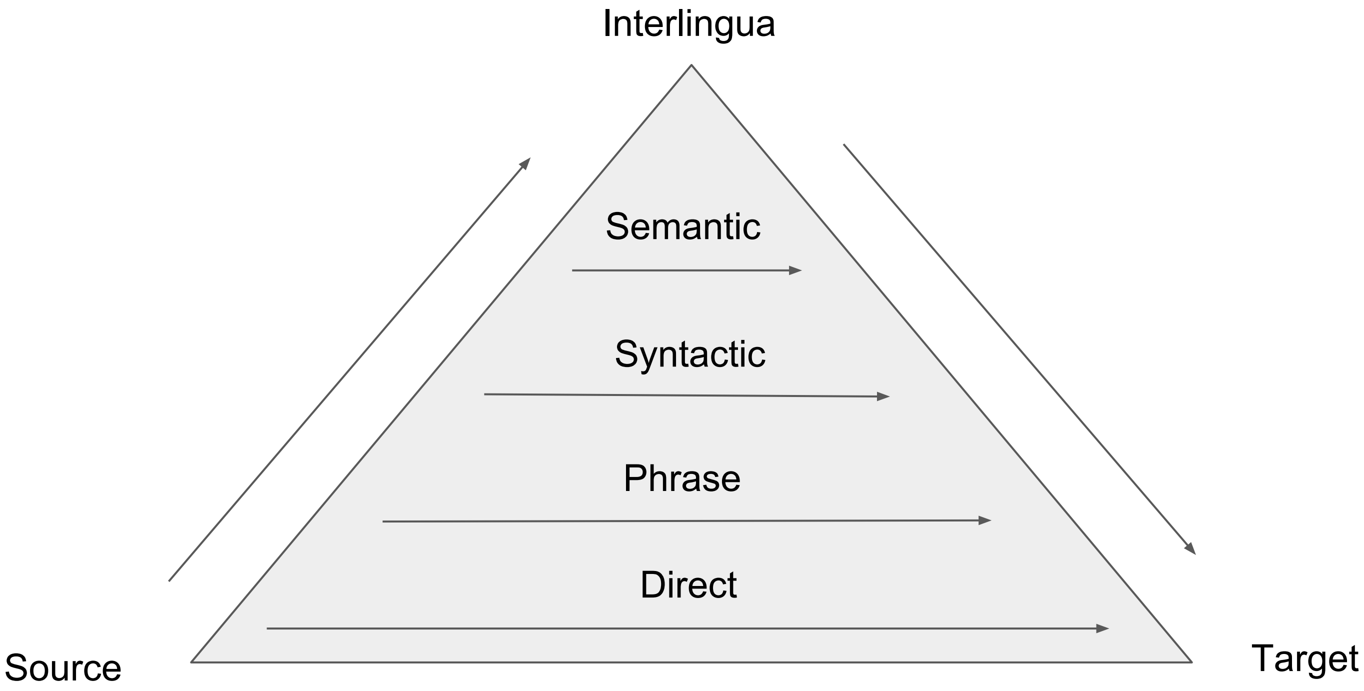
Out of sight, out of mind

Out of sight, out of mind - Blind idiot

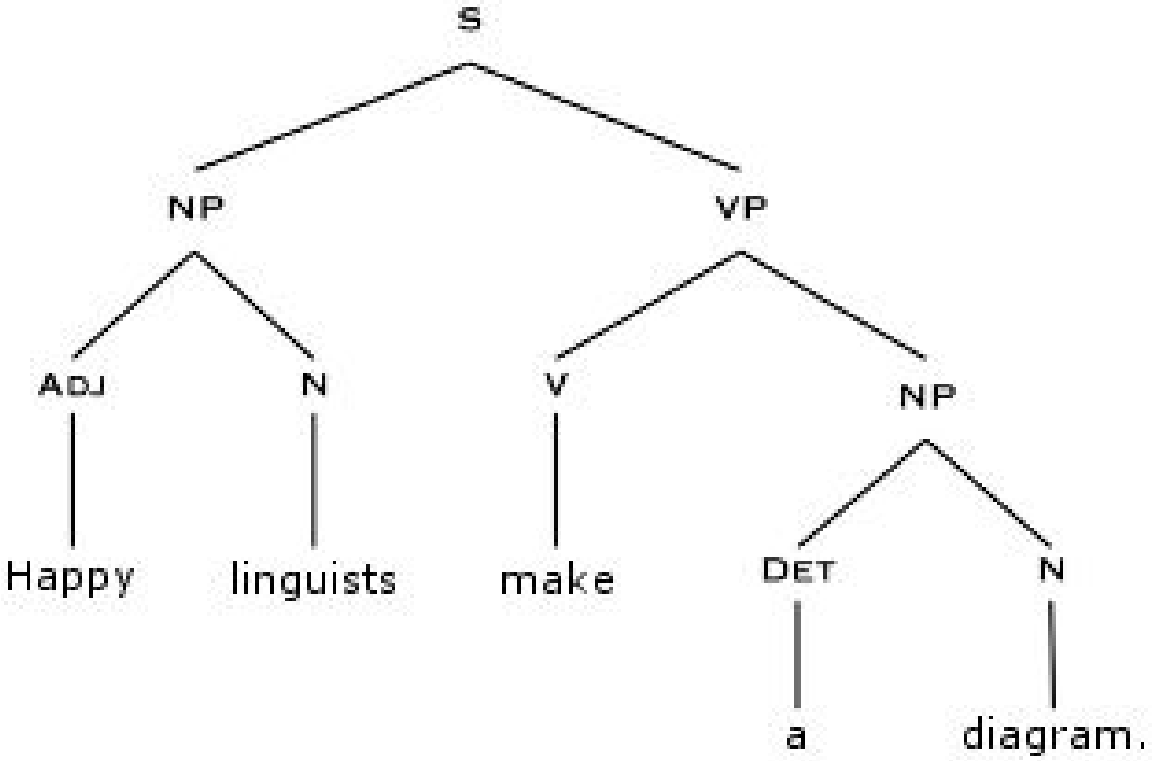
Machine translation types

- Rule-based
- Statistical
- Neural

Machine translation



Syntactic tree



By domain

- Legal documents - easy to be translated
- Poetry - not so much

Evaluation

- Original: I like watching football
- Google Translate: Харесва ми да гледам футбол

Evaluation - BLEU

- Original: I like watching football
- Google Translate: Харесва ми да гледам футбол
- Reference 1: Обичам да гледам футбол
- Reference 2: Приятно ми е да гледам футбол

BLEU - calculation

- BiLingual Evaluation Understudy
- Uses multiple sentences as reference
- Precision - combined modified n-gram precision
  - Uses the number of times 1-gram, 2-gram, 3-gram and 4-gram are met both in in the Candidate and Reference texts.
- Recall - brevity penalty
  - Punishes for length of sentence

BLEU - limitations

- It depends on the number of references that are given
- It can only be used in comparison - never absolutely
- Sometimes perfect human translations score lower than machine translations
- Low n-gram score is not necessarily indicative of a poor translation, although a high n-gram score is probably indicative of a good translation.
- n-gram metrics are really document similarity measures rather than true translation quality measures

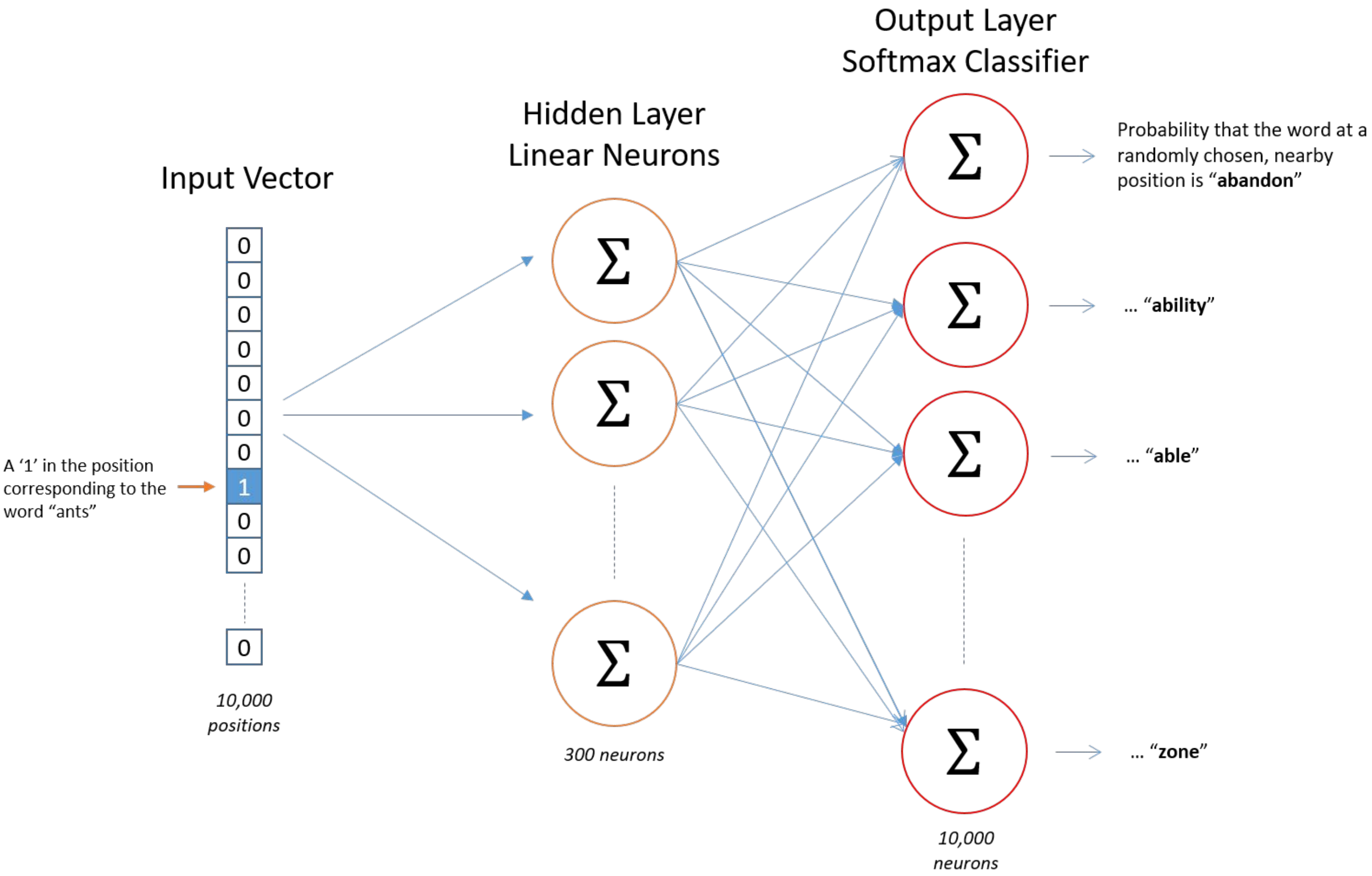
Moses

Open Source Statistical Machine translation system

Word embeddings

Word embeddings

Ракия + Жена - Мъж =
Боза



Input Vector

Hidden Layer
Linear Neurons

Output Layer
Softmax Classifier

A ‘1’ in the position corresponding to the word “ants”

10,000 positions

300 neurons

10,000 neurons

Probability that the word at a randomly chosen, nearby position is “**abandon**”

… “**ability**”

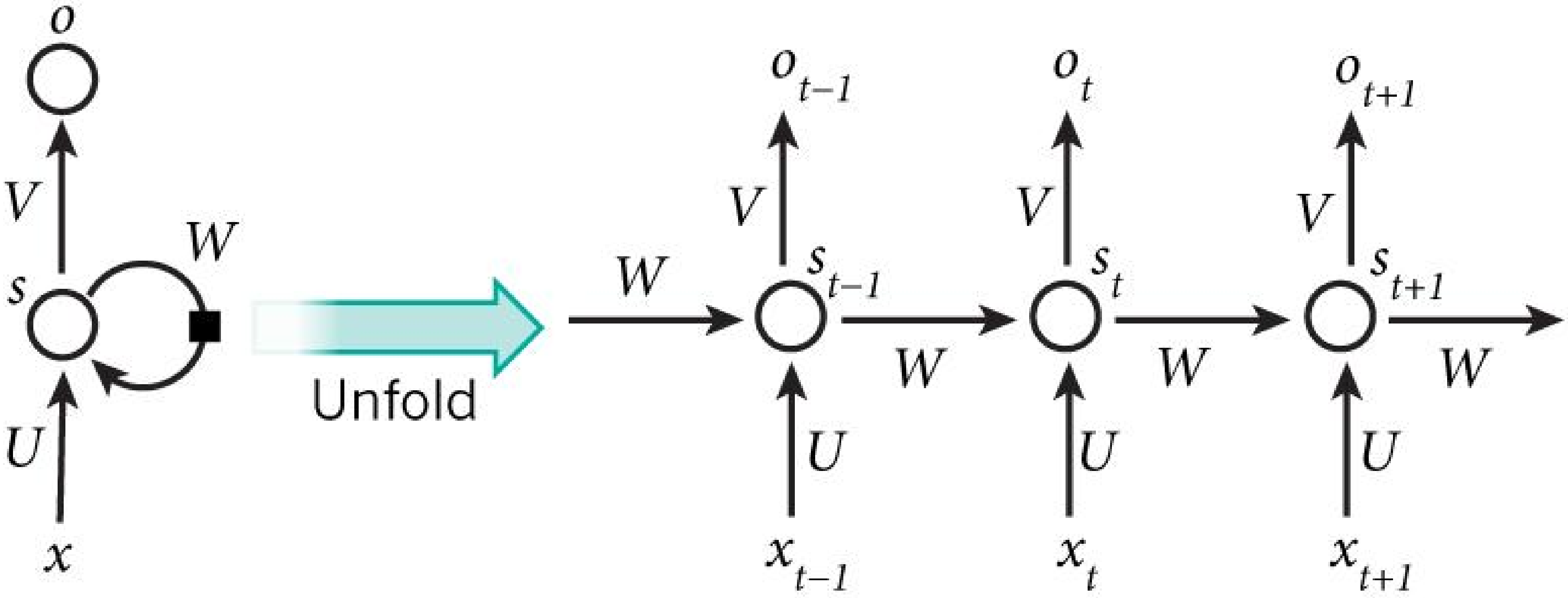
… “**able**”

… “**zone**”

Word embeddings

$$\begin{bmatrix} 0 & 0 & 0 & 1 & 0 \end{bmatrix} \times \begin{bmatrix} 17 & 24 & 1 \\ 23 & 5 & 7 \\ 4 & 6 & 13 \\ 10 & 12 & 19 \\ 11 & 18 & 25 \end{bmatrix} = \begin{bmatrix} 10 & 12 & 19 \end{bmatrix}$$

Recurrent Neural Network (RNN)



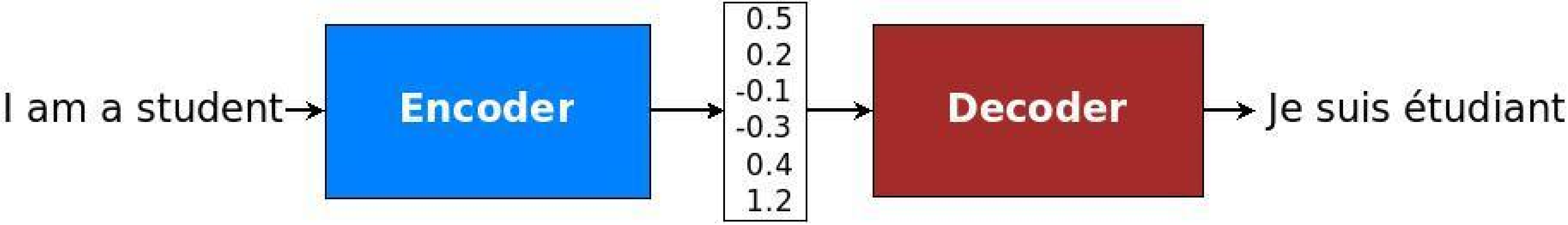
Long short-term memory (LSTM)

- RNN
- Can remember context from way back

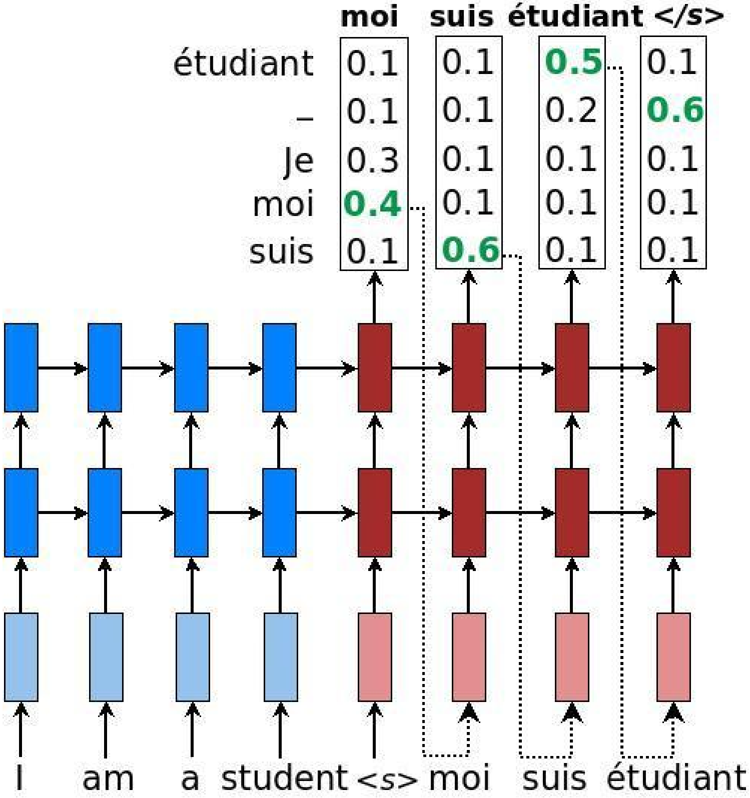
TensorFlow Machine Translation Tutorial

https://github.com/tensorflow/nmt

TensorFlow Machine Translation Tutorial

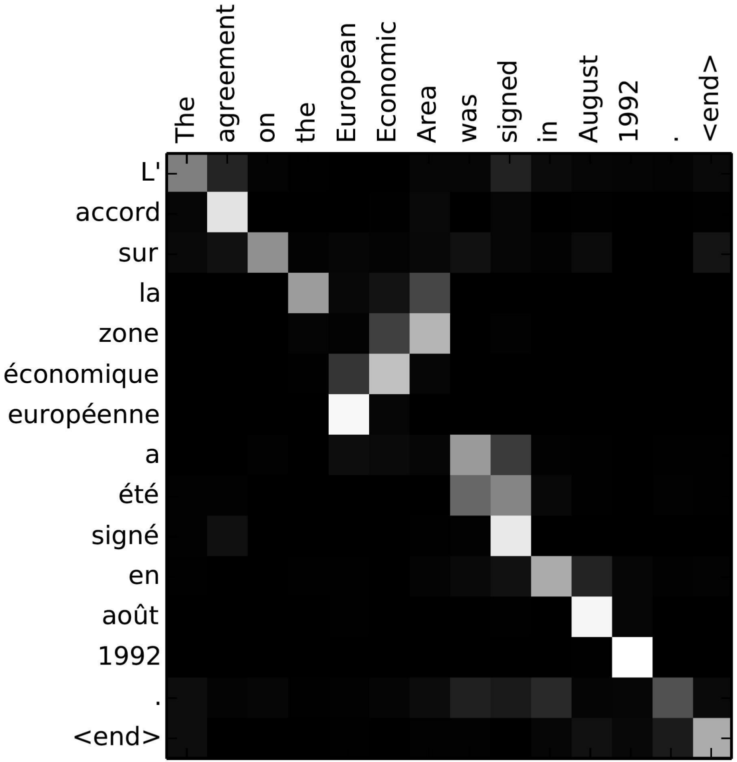


Sequence to sequence

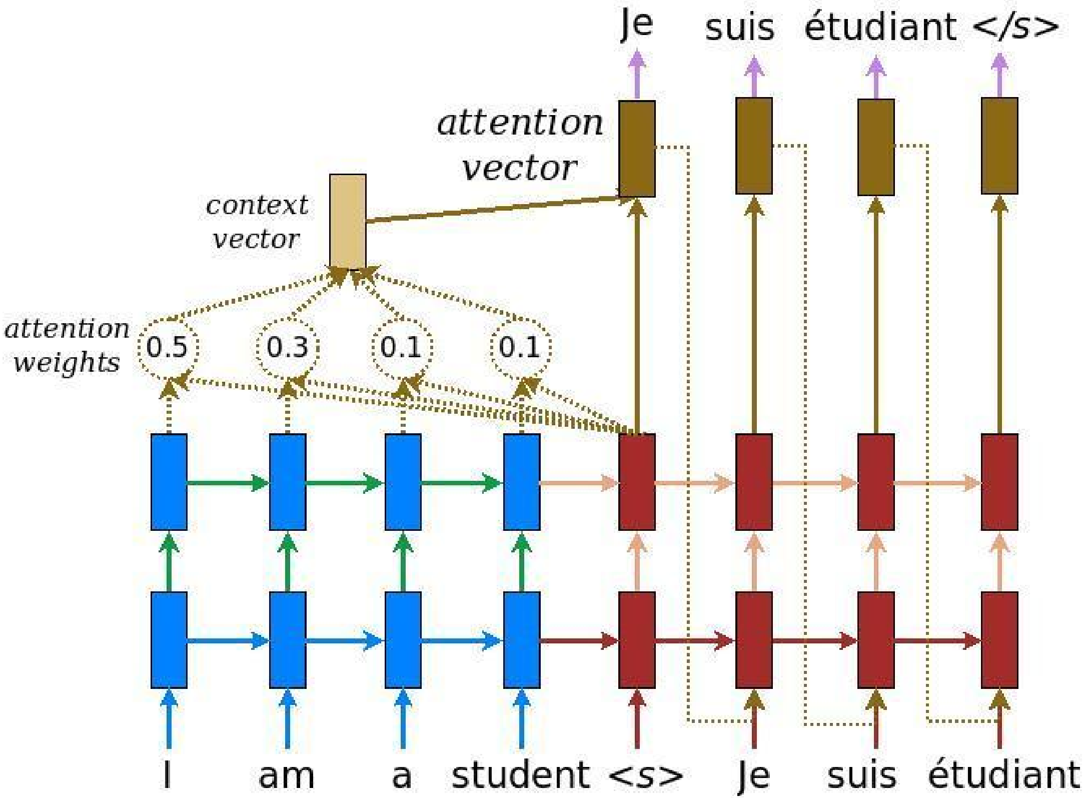


	moi	suis	étudiant	</s>
étudiant	0.1	0.1	**0.5**	0.1
_	0.1	0.1	0.2	**0.6**
Je	0.3	0.1	0.1	0.1
moi	**0.4**	0.1	0.1	0.1
suis	0.1	**0.6**	0.1	0.1

Attention Mechanism



Attention Mechanism



Attention Mechanism

$$\alpha_{ts} = \frac{\exp\left(\mathrm{score}(\boldsymbol{h}_t, \bar{\boldsymbol{h}}_s)\right)}{\sum_{s'=1}^{S} \exp\left(\mathrm{score}(\boldsymbol{h}_t, \bar{\boldsymbol{h}}_{s'})\right)} \qquad \text{[Attention weights]} \qquad (1)$$

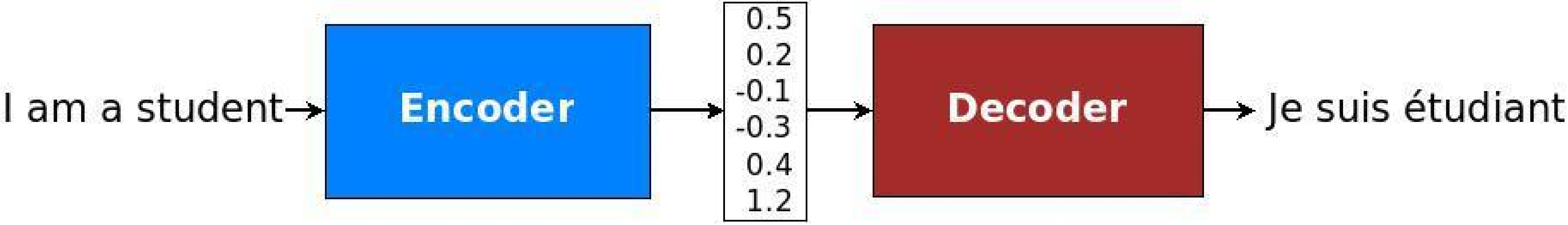
$$\boldsymbol{c}_t = \sum_s \alpha_{ts} \bar{\boldsymbol{h}}_s \qquad \text{[Context vector]} \qquad (2)$$

$$\boldsymbol{a}_t = f(\boldsymbol{c}_t, \boldsymbol{h}_t) = \tanh(\boldsymbol{W_c}[\boldsymbol{c}_t; \boldsymbol{h}_t]) \qquad \text{[Attention vector]} \qquad (3)$$

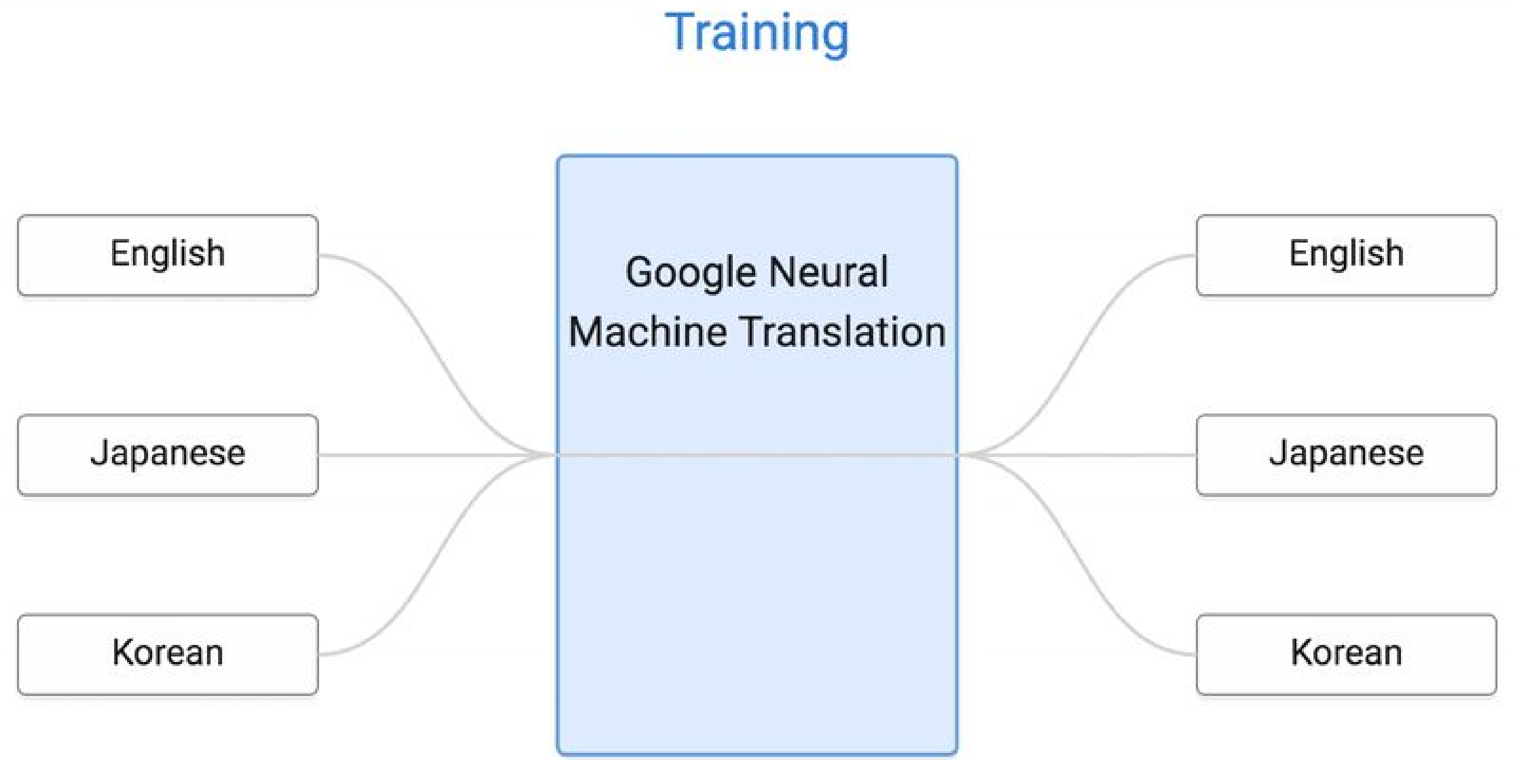
TensorFlow Machine Translation Tutorial Recap

- Sequence to sequence consisting of coder and decoder
- Input word embeddings, Output word embeddings
- Encoder and Decoder are different LSTM RNN
- Context vector computed from the encoder
- Decoder using the context vector plus translated sentences so far
- Attention mechanism used to dynamically change the context vector for each target word
- The most probable output can be chosen by beam search

TensorFlow Machine Translation Tutorial



Zero-shot translation



References

- http://mccormickml.com/2016/04/19/word2vec-tutorial-the-skip-gram-model/
- https://github.com/tensorflow/nmt
- http://www.wildml.com/2015/09/recurrent-neural-networks-tutorial-part-1-introduction-to-rnns/
- https://research.googleblog.com/2016/11/zero-shot-translation-with-googles.html