Deep Learning

Big Data & Machine Learning Bootcamp - Keep Coding

Outline

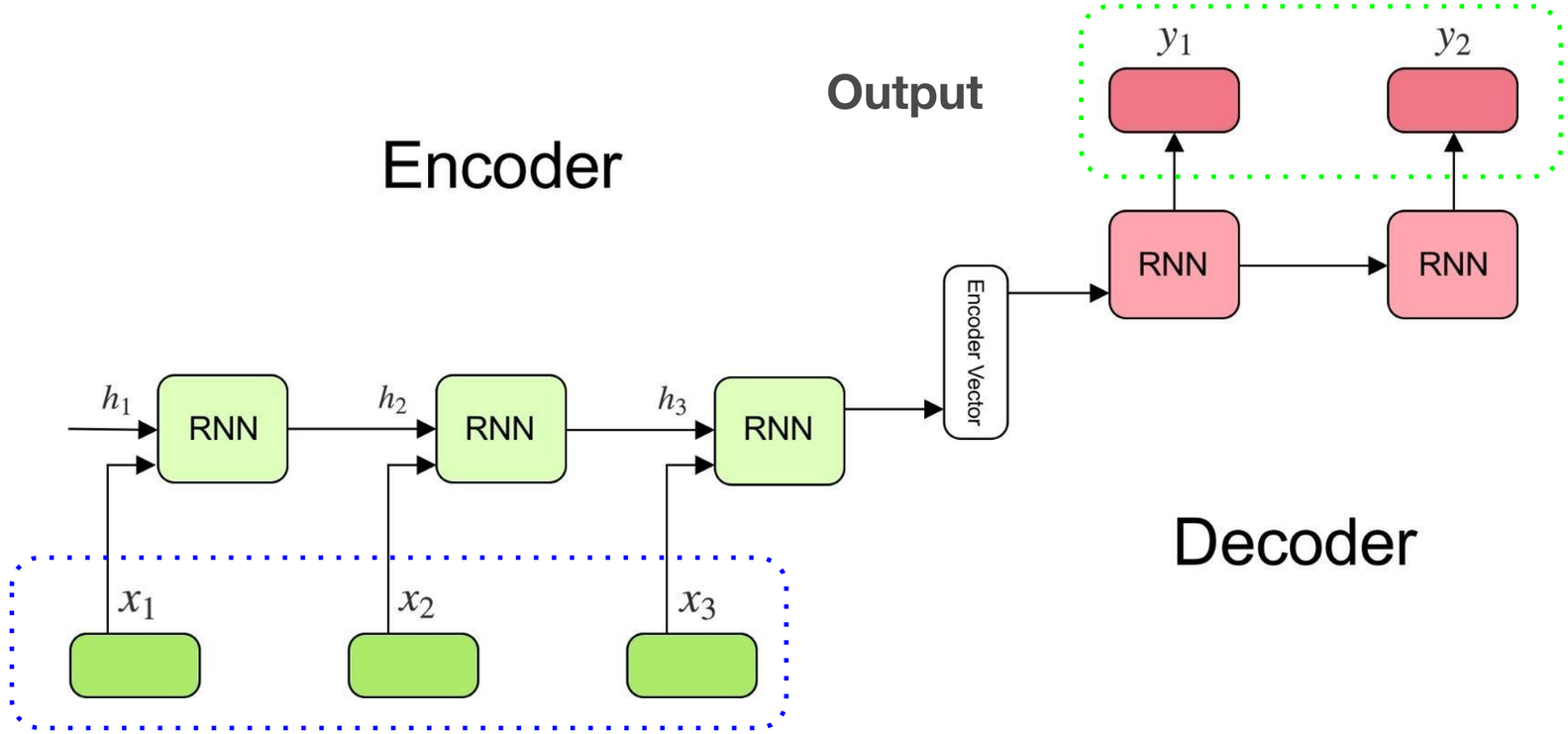
1. Various sequence to sequence architectures
2. Picking the most likely sequence
3. Beam search
4. Bleu score
5. Attention model
6. Speech recognition
7. Trigger word detection

1. Various sequence to sequence architectures

Sequence to sequence models are useful to machine translation and speech recognition

For instance, you want to translate from:

Jane visite l'Afrique en septembre to
Jane is visiting Africa in September

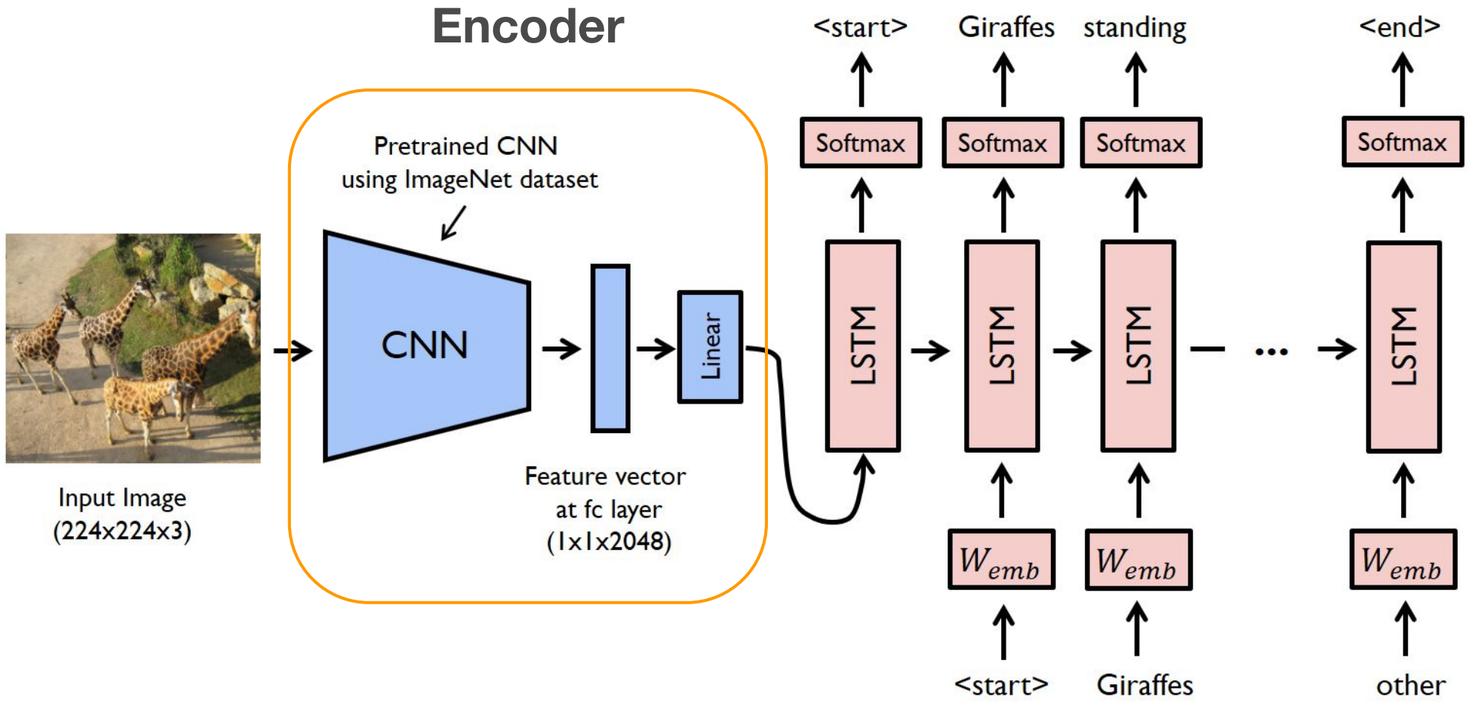


Sutskever, Ilya, Oriol Vinyals, and Quoc V. Le. "Sequence to sequence learning with neural networks." *Advances in neural information processing systems* 27 (2014): 3104-3112.

Sources:
- Coursera
- https://towardsdatascience.com/understanding-encoder-decoder-sequence-to-sequence-model-679e04af4346

1. Various sequence to sequence architectures

Encoder-Decoder architecture for Image Captioning



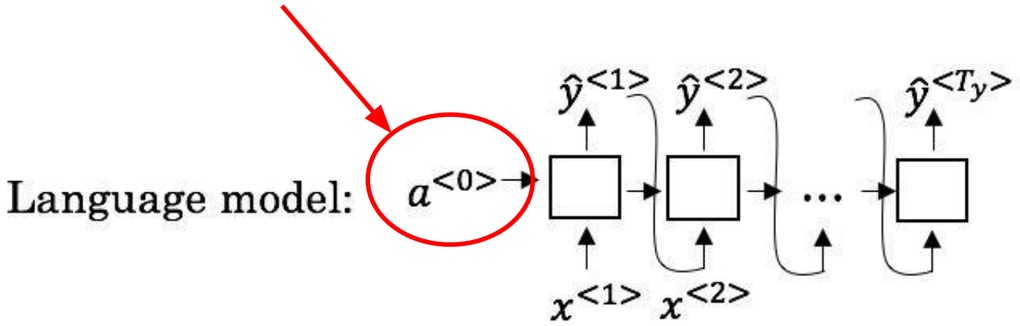
1- Sutskever, Ilya, Oriol Vinyals, and Quoc V. Le. "Sequence to sequence learning with neural networks." *Advances in neural information processing systems* 27 (2014): 3104-3112.
2- Mao, Junhua, Wei Xu, Yi Yang, Jiang Wang, Zhiheng Huang, and Alan Yuille. "Deep captioning with multimodal recurrent neural networks (m-rnn)." arXiv preprint arXiv:1412.6632 (2014).

Sources:
- Coursera
- https://www.analyticsvidhya.com/blog/2018/04/solving-an-image-captioning-task-using-deep-learning/

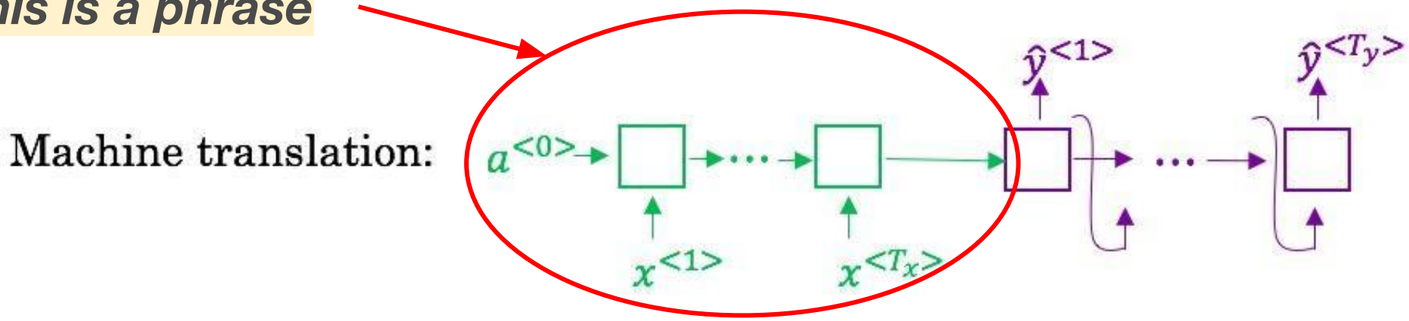
2. Picking the most likely sequence

Machine translation can be seen as building a conditional language model

For a language model, this is zero



For machine translation, this is a phrase



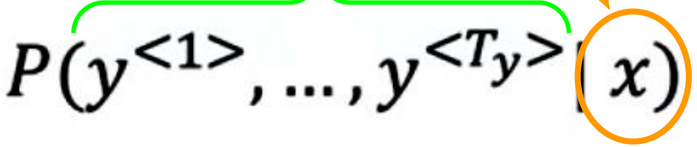
1- Sutskever, Ilya, Oriol Vinyals, and Quoc V. Le. "Sequence to sequence learning with neural networks." *Advances in neural information processing systems* 27 (2014): 3104-3112.

Sources:
- Coursera

2. Picking the most likely sequence

In machine translation and speech recognition we don't want any translation, we actually want the most likely sentence

English translation **French input**



French Input Jane visite l'Afrique en septembre.

Good translation

→ Jane is visiting Africa in September.

→ Jane is going to be visiting Africa in September.

→ In September, Jane will visit Africa.

Very poor translation

→ Her African friend welcomed Jane in September.

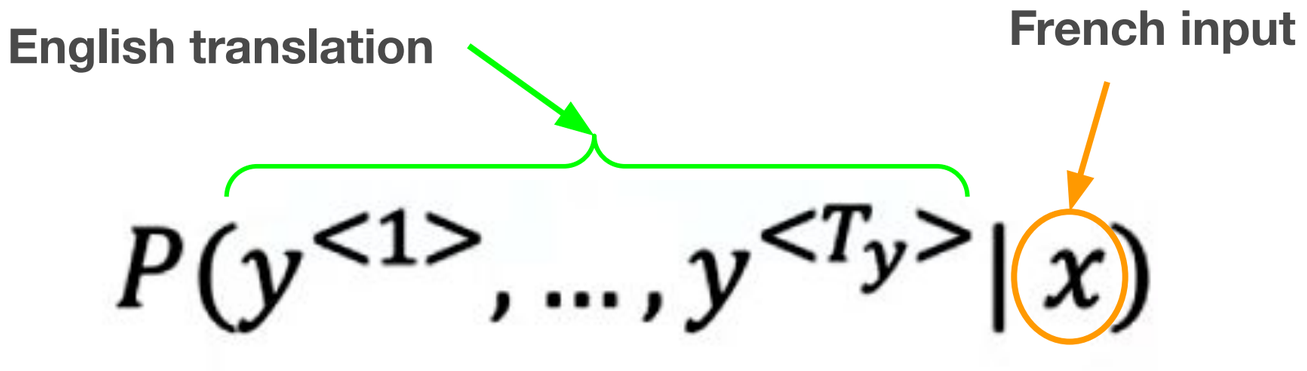
1- Sutskever, Ilya, Oriol Vinyals, and Quoc V. Le. "Sequence to sequence learning with neural networks." *Advances in neural information processing systems* 27 (2014): 3104-3112.

Sources:
- Coursera

2. Picking the most likely sequence

In machine translation and speech recognition we don't want any translation, we actually want the most likely sentence.

This means, we want to **maximize the probability of having the english translation** given the French sentence:



1- Sutskever, Ilya, Oriol Vinyals, and Quoc V. Le. "Sequence to sequence learning with neural networks." *Advances in neural information processing systems* 27 (2014): 3104-3112.

Sources:
- Coursera

3. Beam search

Beam search is the most widely algorithm to find the most likely sentence.

It has a parameters called B, which is the number of words that the system will consider to compute the most likely words. ***It has mainly three steps****:*

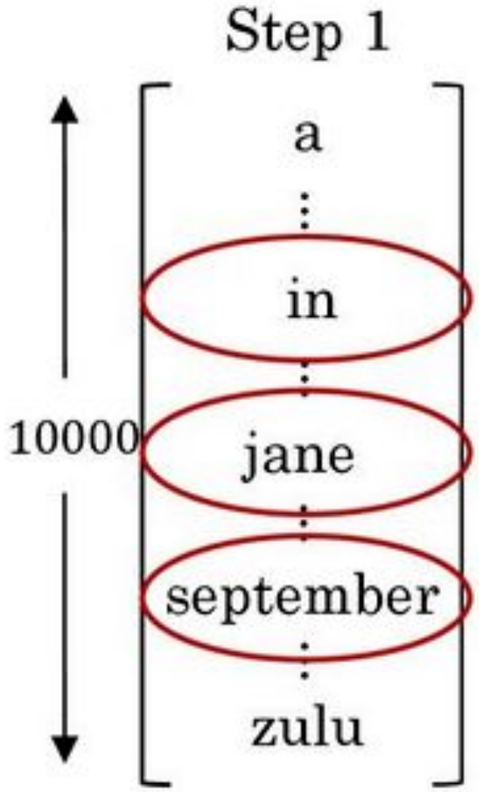
1. *From all the words in the vocabulary, pick the B (3, 5 or 7) more likely.*
2. *From the B more likely, evaluate which word in the vocabulary is the most likely.*
   This means, repeat step 1 with the B selected words.
3. *Fix the B-2 previous words and compute the B most likely word*

Sources:
- Coursera

3. Beam search

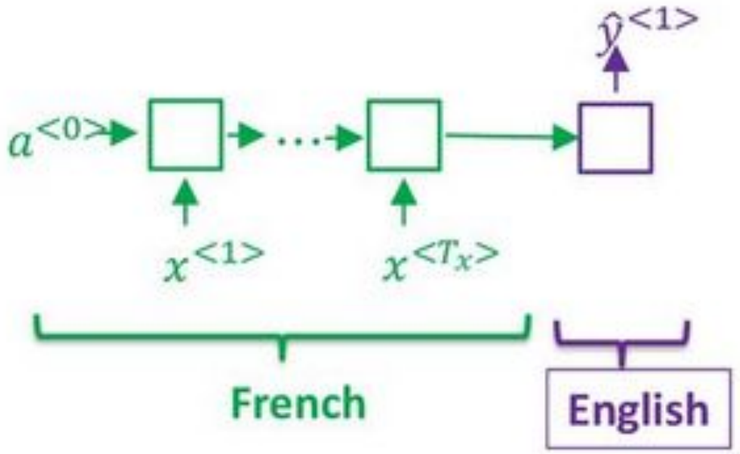
Step 1: *Jane visite l'Afrique en septembre*

Dictionary



B = 3 (Beam width)

$P(y^{<1>} \mid x)$



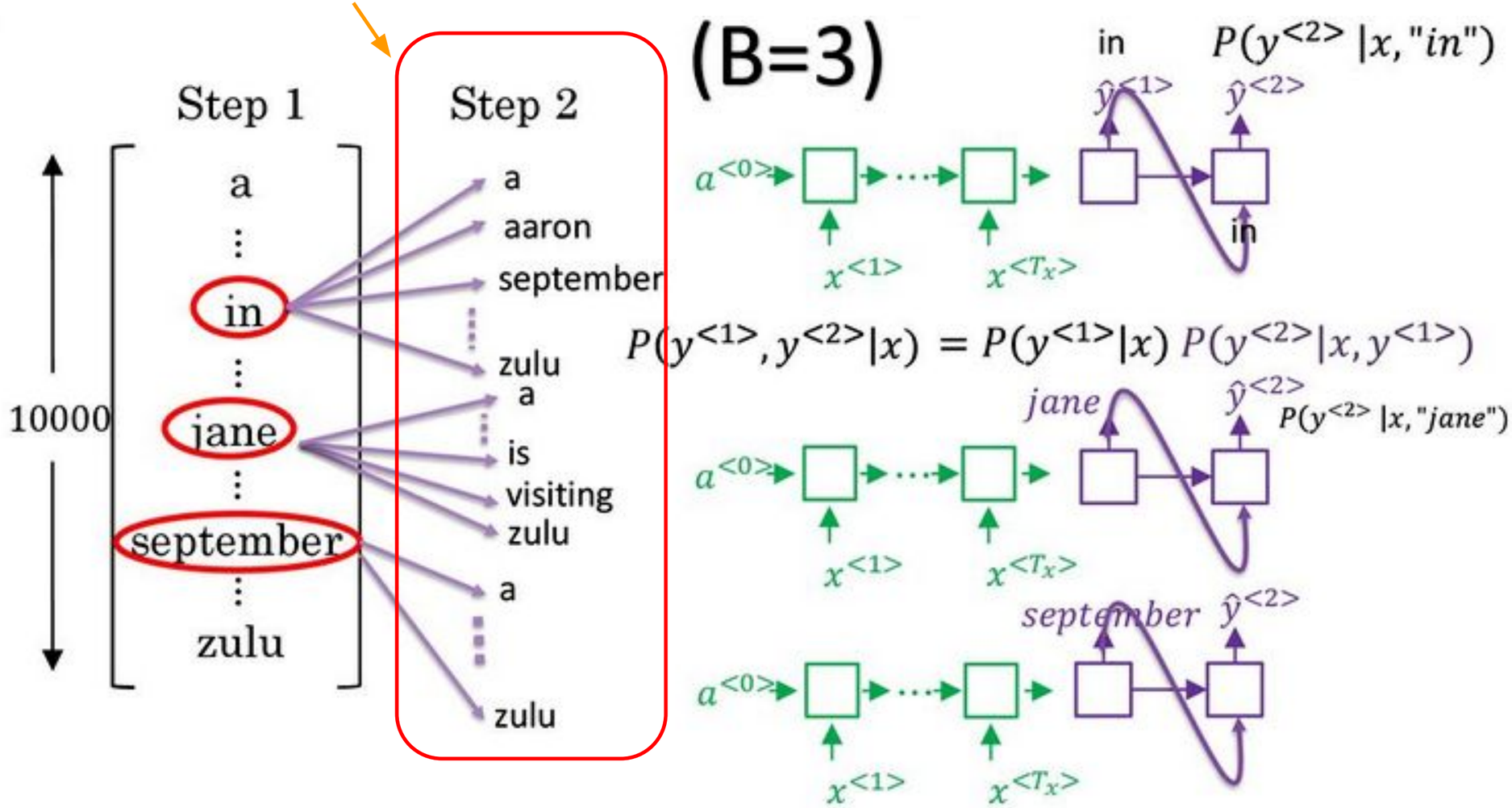
Sources:
- Coursera
- https://slideplayer.com/slide/16450517/

3. Beam search

Step 2

Compute the second most likely word based on the first ones

Dictionary

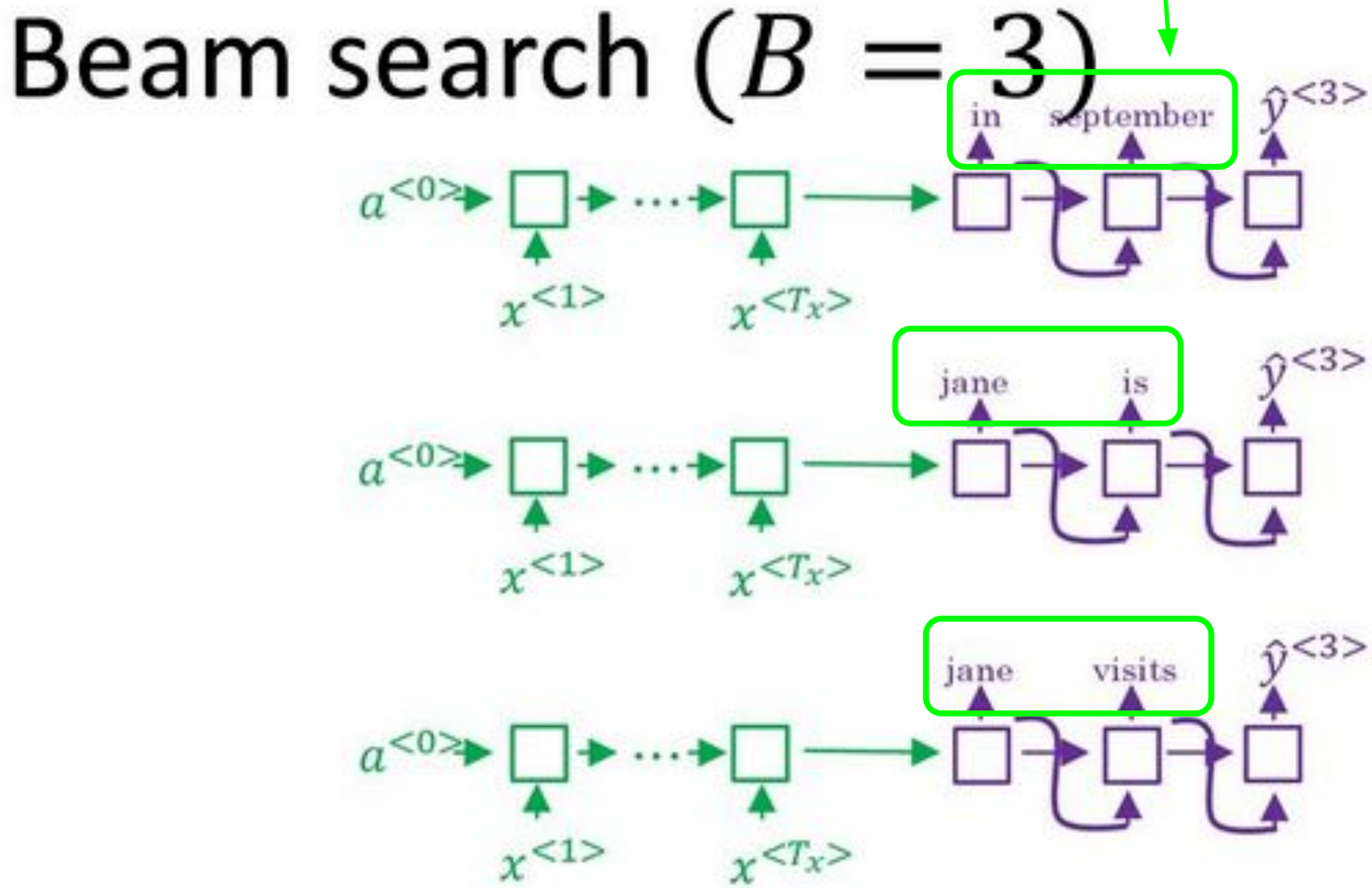


Sources:
- Coursera
- https://slideplayer.com/slide/16450517/

3. Beam search

Step 3: *Jane visite l'Afrique en septembre*

Fix the second previous words (B=3) and compute the third most likely

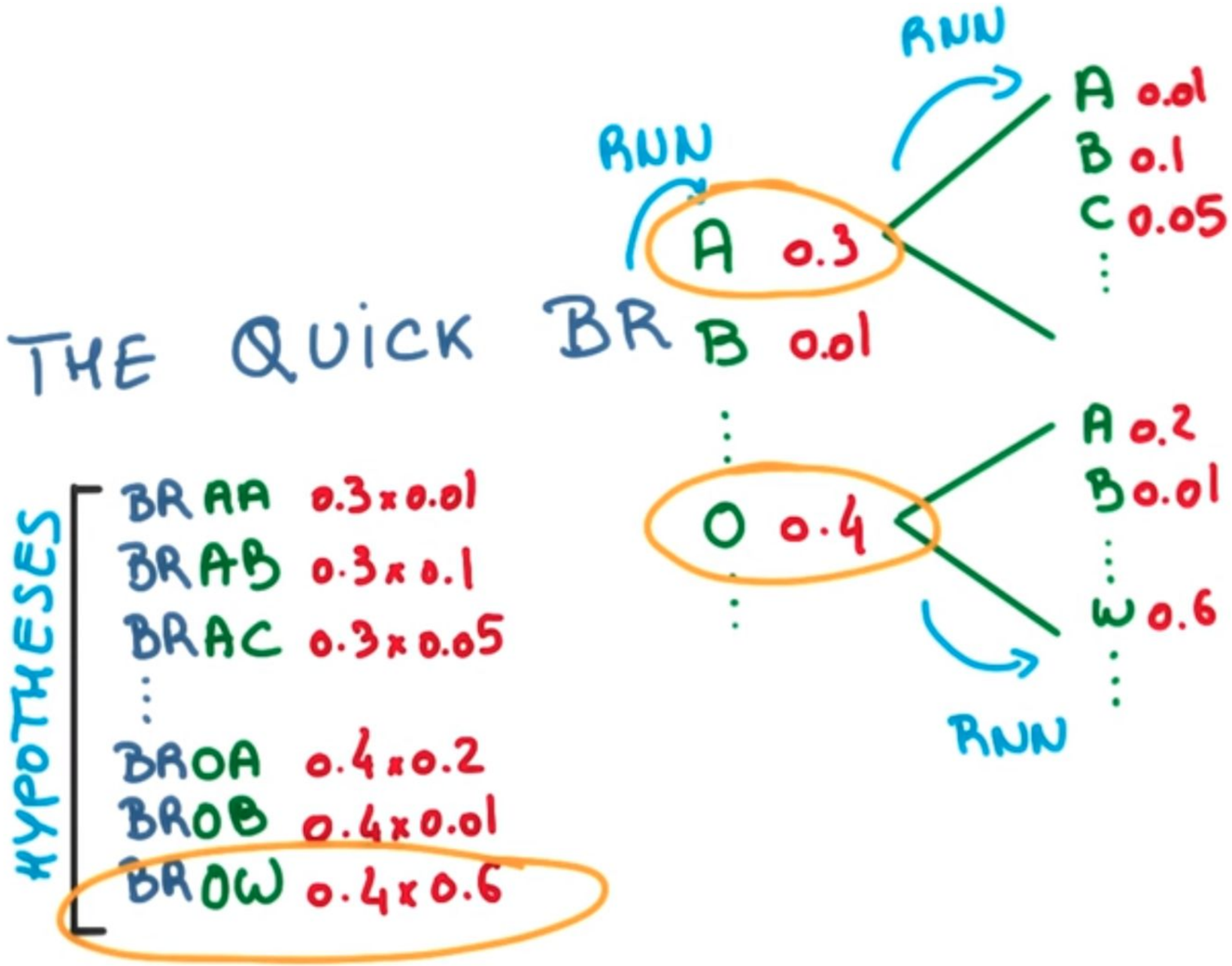


The larger the B*, the better the sentence is, but the more computationally expensive is the algorithm.*

Sources:
- Coursera
- https://slideplayer.com/slide/16450517/

3. Beam search

***The larger the B**, the better the sentence is, but the more computationally expensive is the algorithm.*



Sources:
- Udacity

4. Bleu score

How do you measure accuracy in machine translation? They can be multiple good translations of one sentence?

Bilingual evaluation understudy (BLEU) score *helps us with that!*

The idea: *The closer a machine translation is to a professional human translation, the better it is.*

- ***The BLEU score ranges from 0 to 1***
- ***1 is very rare: only for perfect match***
- ***The more the better***

Papineni, Kishore, Salim Roukos, Todd Ward, and Wei-Jing Zhu. "BLEU: a method for automatic evaluation of machine translation." In *Proceedings of the 40th annual meeting of the Association for Computational Linguistics*, pp. 311-318. 2002.

Sources:
- Coursera

4. Bleu score

BLEU score compares n-grams of the candidate with the n-grams of the reference translation and count the number of matches.

(Source original) Le professeur est arrivé en retard à cause de la circulation
(Reference translation) The teacher arrived late because of the traffic

Bigram is a pair of words that appears next to each other (The teacher, teacher arrived, arrived late, etc)

Papineni, Kishore, Salim Roukos, Todd Ward, and Wei-Jing Zhu. "BLEU: a method for automatic evaluation of machine translation." In *Proceedings of the 40th annual meeting of the Association for Computational Linguistics*, pp. 311-318. 2002.

Sources:
- Coursera

4. Bleu score

Example:

(Source original) Le professeur est arrivé en retard à cause de la circulation
(Reference translation) The teacher arrived late because of the traffic

(#1 Very low BLEU score) The professor was delayed due to the congestion
(#2 Slightly higher but low BLEU) Congestion was responsible for the teacher being late
(#3 Higher BLEU than #1 and #2) The teacher was late due to the traffic
(#4 Higher BLEU than #3) The professor arrived late because of circulation
(#5 The best BLEU score) The teacher arrived late because of the traffic

Green *= 4-gram match (Very good)*
Light blue *= bigrams match (Good)*
Red *= word not matched (bad)*

Papineni, Kishore, Salim Roukos, Todd Ward, and Wei-Jing Zhu. "BLEU: a method for automatic evaluation of machine translation." In *Proceedings of the 40th annual meeting of the Association for Computational Linguistics*, pp. 311-318. 2002.

Sources:
- Coursera

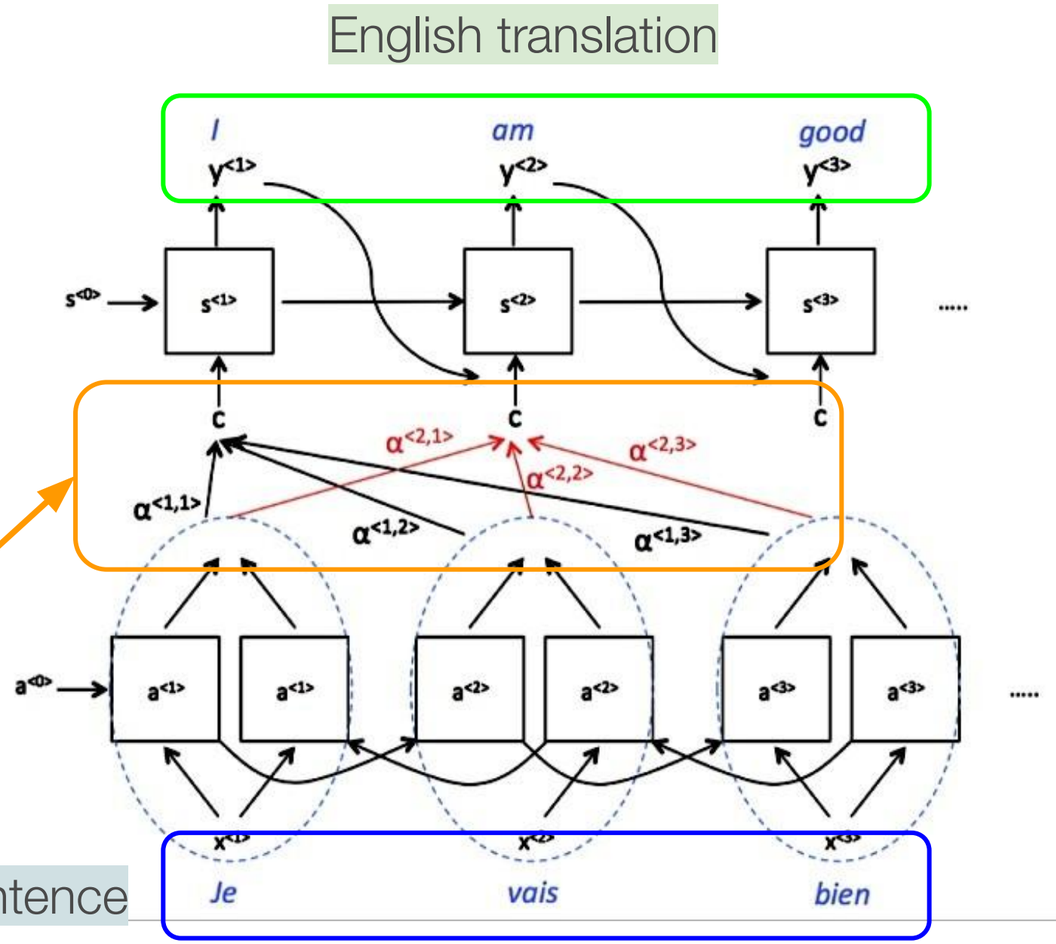
5. Attention model

So far we've worked with Encoder-Decoder architecture for machine translation

BUT there is something even more powerful called attention model. ***It solves the problem of translating long sequences.***

It tries to replicate human translation process by paying attention to a portion of the phrase to produce a translation.

These connections are ***small neural networks*** *that are learned during the training process*



Bahdanau, Dzmitry, Kyunghyun Cho, and Yoshua Bengio. "Neural machine translation by jointly learning to align and translate." arXiv preprint arXiv:1409.0473 (2014).

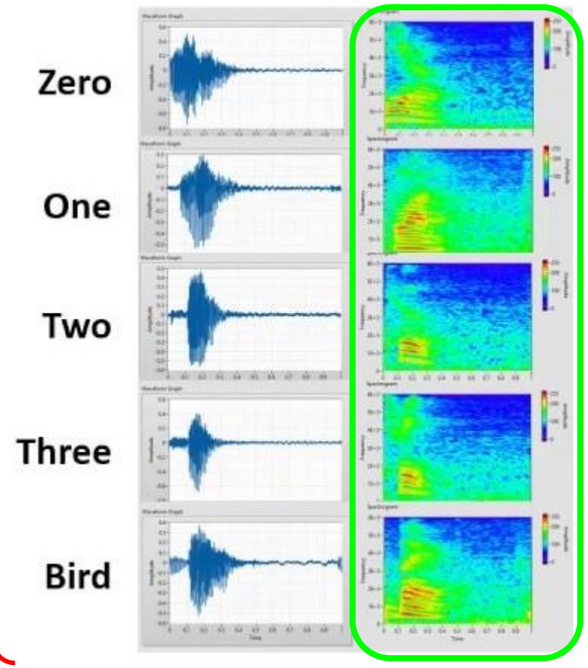
Sources:
- Coursera
- https://zhuanlan.zhihu.com/p/136316084

6. Speech recognition

One of the most exciting developments given by the sequence-to-sequence models is the rise of accurate speech recognition systems!

The speech recognition problem: *From an audio clip, try to obtain the transcript*

Transcripts



Fourier transform*: From time signal to frequency signal (spectrogram).*

X axis represents time,
Y axis represents frequencies and
the color indicates energy (how loud is the frequency at different time)

Sources:
- Coursera
- https://www.youtube.com/watch?v=s9gpKM7gqK4

6. Speech recognition

Problem! ***An audio input of 10 seconds is way bigger than the output (transcript).***

To solve that, we use the Connectionist temporal classification (CTC) cost

The basic rule for the CTC *is to collapse repeated characters not separated by “blank”*

Graves, Alex, Santiago Fernández, Faustino Gomez, and Jürgen Schmidhuber. "Connectionist temporal classification: labelling unsegmented sequence data with recurrent neural networks." In Proceedings of the 23rd international conference on Machine learning, pp. 369-376. 2006.

Sources:
- Coursera

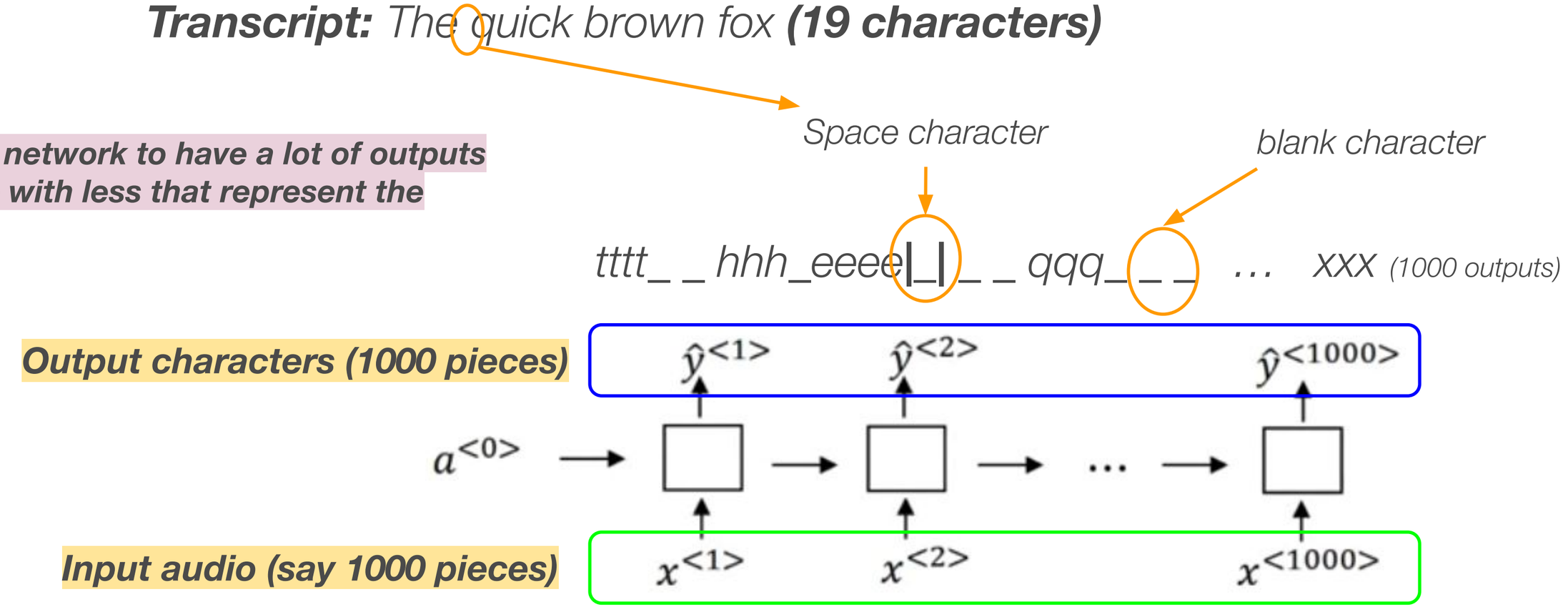
6. Speech recognition

Collapsing repeated characters by blank character - ***CTC cost***

From 1000 characters to 19 characters

Transcript: *The quick brown fox* ***(19 characters)***

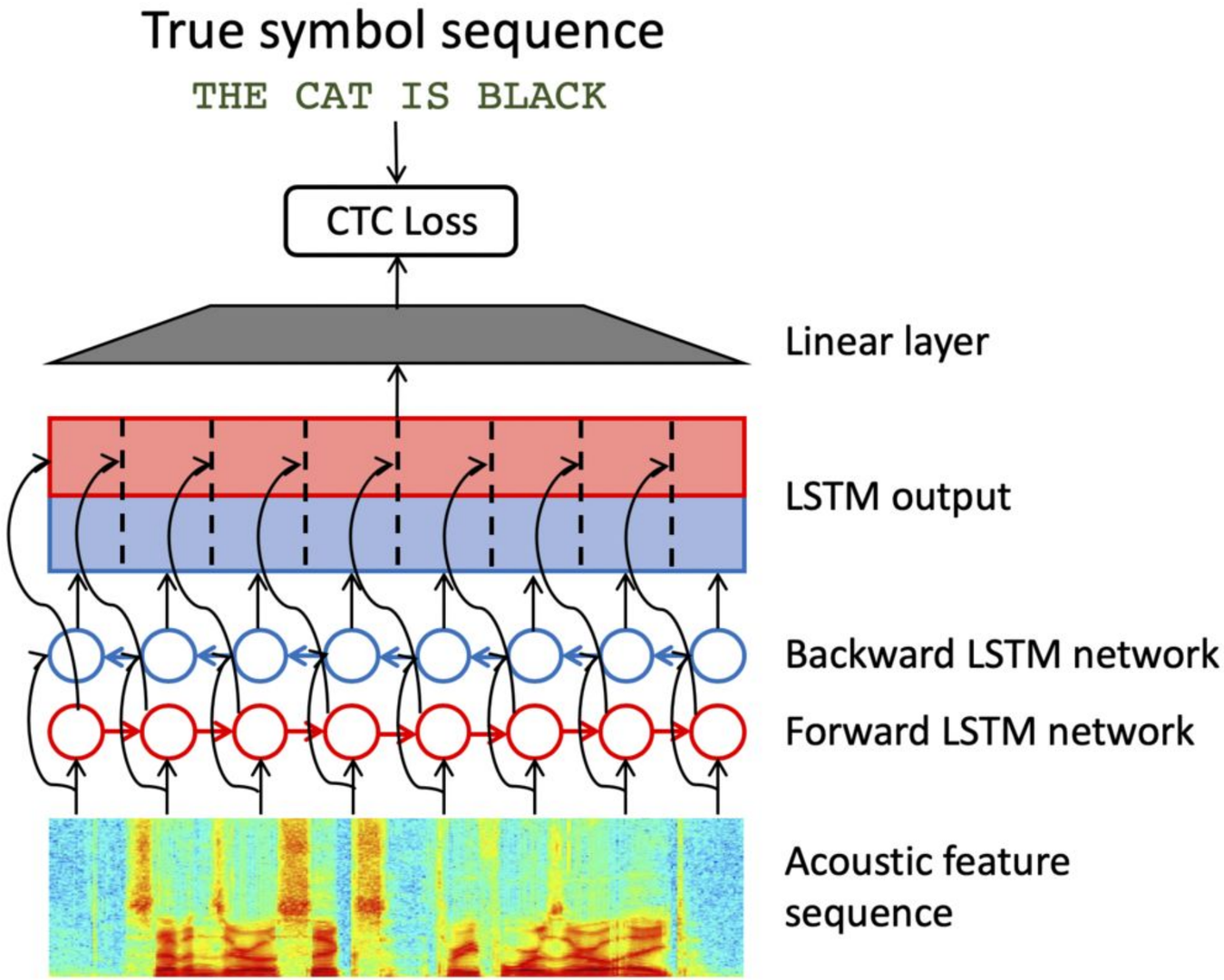
CTC allows the network to have a lot of outputs and still end up with less that represent the transcript



Graves, Alex, Santiago Fernández, Faustino Gomez, and Jürgen Schmidhuber. "Connectionist temporal classification: labelling unsegmented sequence data with recurrent neural networks." In Proceedings of the 23rd international conference on Machine learning, pp. 369-376. 2006.

Sources:
- Coursera

6. Speech recognition



Sources:
- https://www.ibm.com/blogs/research/2019/10/end-to-end-speech-recognition/

7. Trigger word detection

Trigger word or waking up word detection examples

Amazon Echo	Baidu DuerOS	Apple Siri	Google Home
(Alexa)	(xiaodunihao)	(Hey Siri)	(Okay Google)

Sources:
- Coursera

7. Trigger word detection

There is still an open discussion regarding which is the best trigger word detection algorithm.

In any case, we'll be implementing our own trigger word detection system.

Sources:
- Coursera

Let’s move to Google Colab!

Notebooks:

- *13_Trigger_word_detection.ipynb*

Sources:
- Coursera