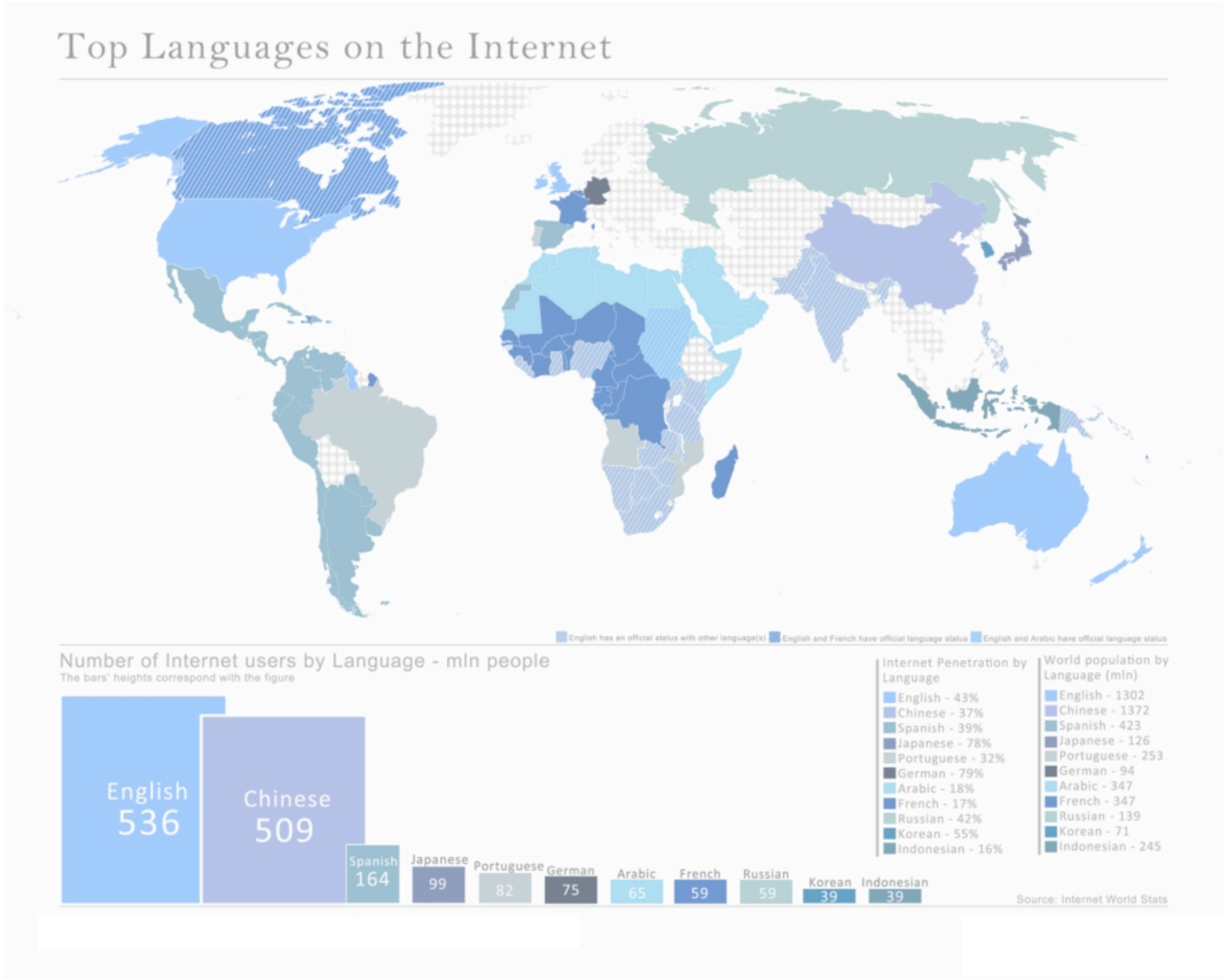
Deep Learning

Lecture 19

Neural Machine Translation

Part 1

7 billion people, 7000 languages

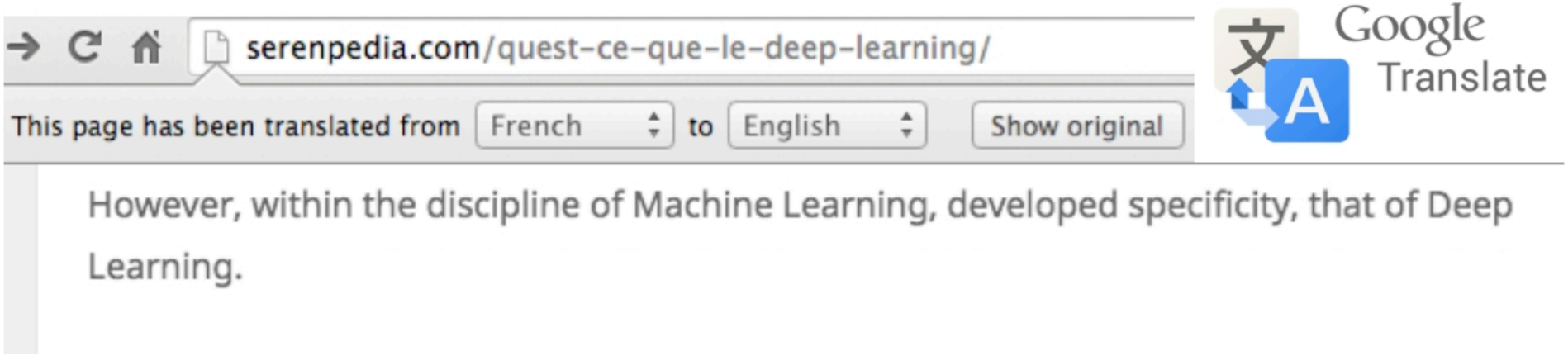


Charlie “Vote” Loyd
@vruba

Follow

Foreign candy always seems like the output of a slightly undertrained neural net.

Machine vs. Human Translation

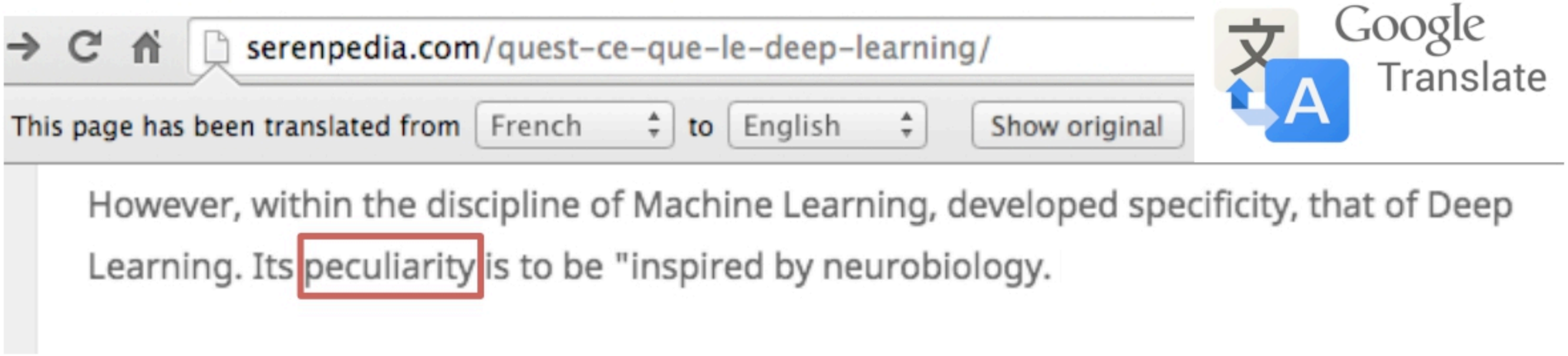


Faithful translation

"Nevertheless, within the discipline of machine learning, a specialization called deep learning has been developed.

Grammatically incorrect.

Machine vs. Human Translation



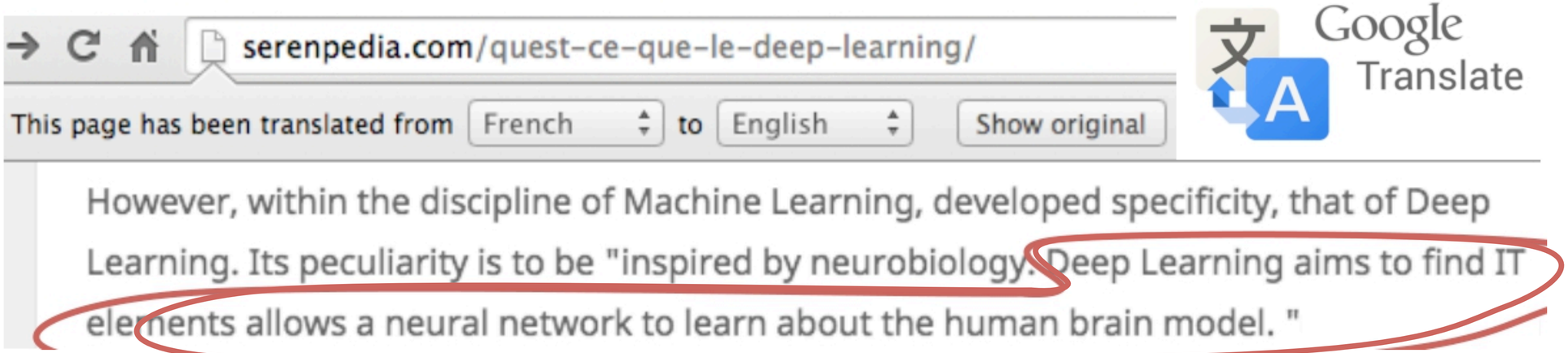
Faithful translation

"Nevertheless, within the discipline of machine learning, a specialization called deep learning has been developed. Its distinguishing feature is that it is inspired by neurobiology.

Poor word choice

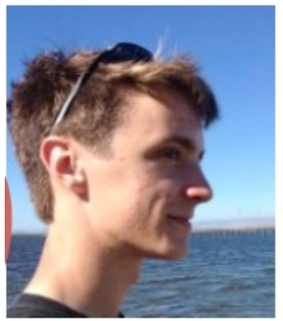
Used with permission from Thong Luong

Machine vs. Human Translation



Faithful translation

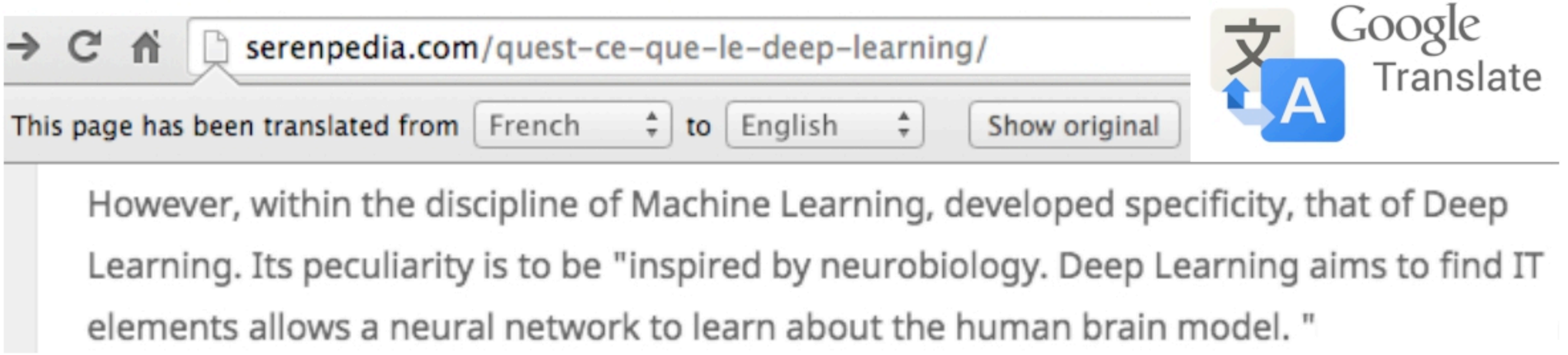
"Nevertheless, within the discipline of machine learning, a specialization called deep learning has been developed. Its distinguishing feature is that it is inspired by neurobiology. Deep learning deals with computational elements which allow a network of artificial neurons to learn a model of the human brain."



- Bad sentence structures.

Used with permission from Thong Luong

Machine vs. Human Translation



Fluent translation

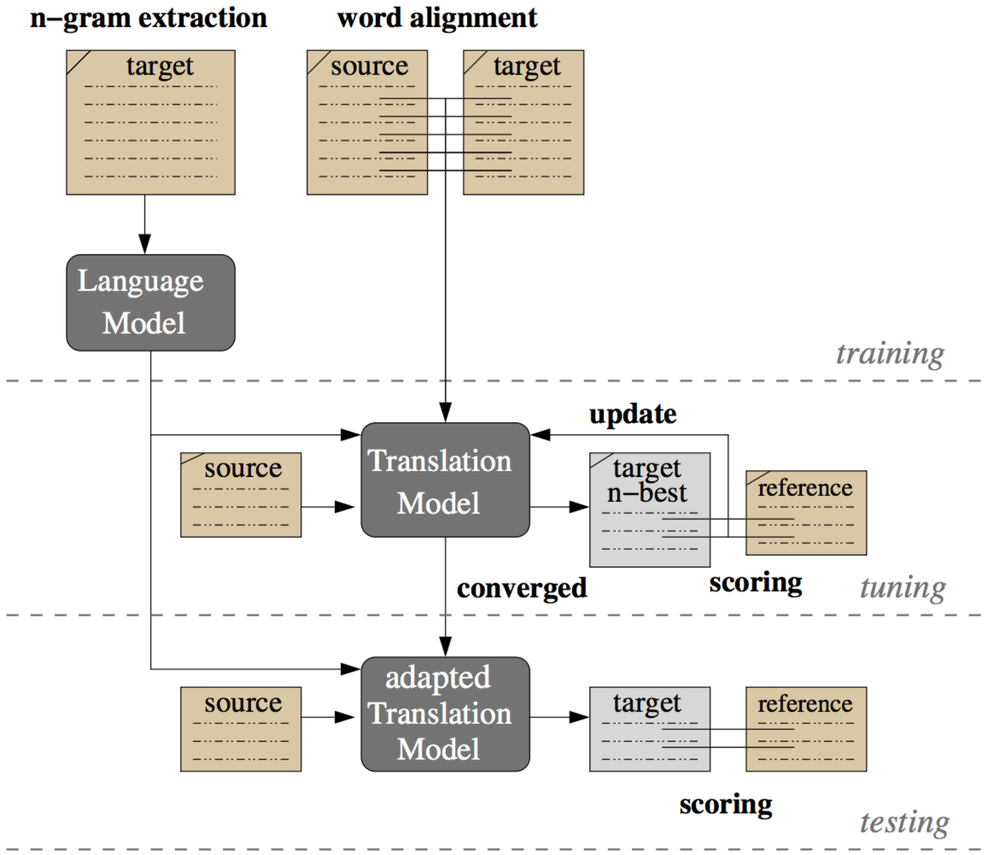
"However, in machine learning a specialization called deep learning has emerged. It can be recognized by its distinctive neurobiological influence. Deep learning is centered around networks of artificial neurons which can learn models of the human brain."

A big gap!

Used with permission from Thong Luong

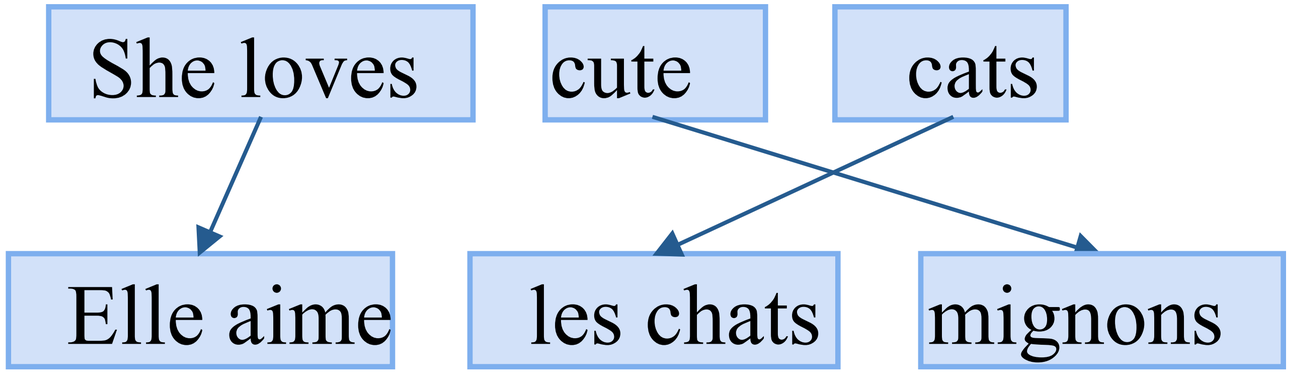
How has MT evolved?

Statistical Machine Translation (SMT)



- many components → pipeline architecture
- context-independent translations
- idependent models (TM, LM, RM)
- learn feature weights using minimum error rate training

Phrase-based MT



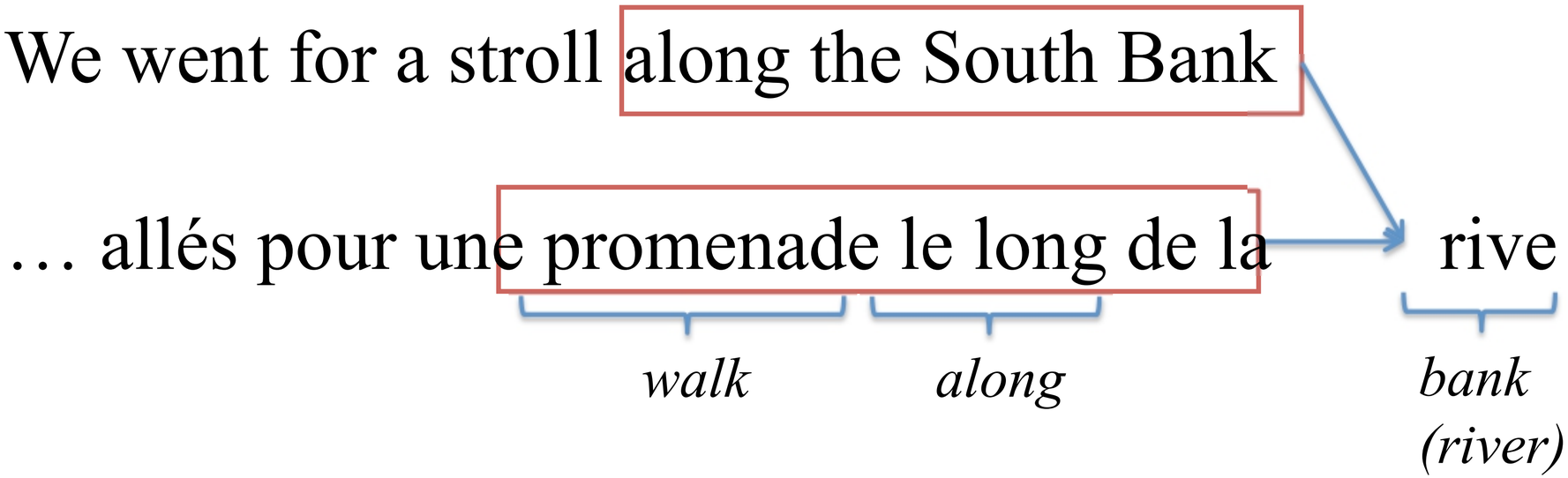
(Brown et al., 1993; Koehn et al., 2003; Och & Ney, 2004)

- Break sentences into chunks.
- *Translation model*: look up phrase translations.
- *Language model*: He phrases together.

Translate locally

LM uses only target words

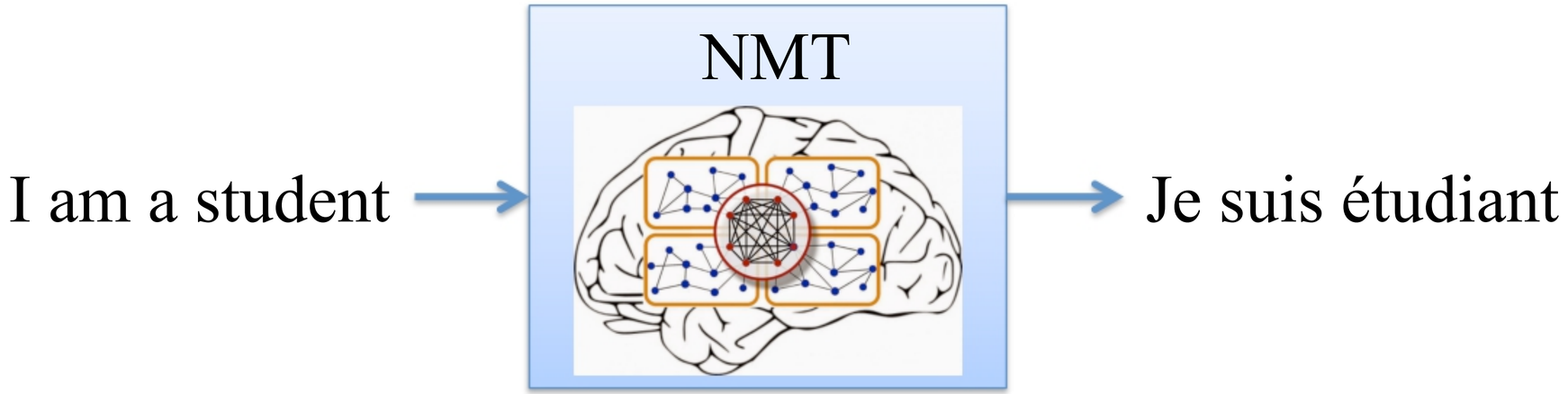
Joint Neural Language Model



- Conditioned on source words (Devlin et al., 2014)
- Still translate locally.

MT systems become more complex!

Neural Machine Translation to the rescue!



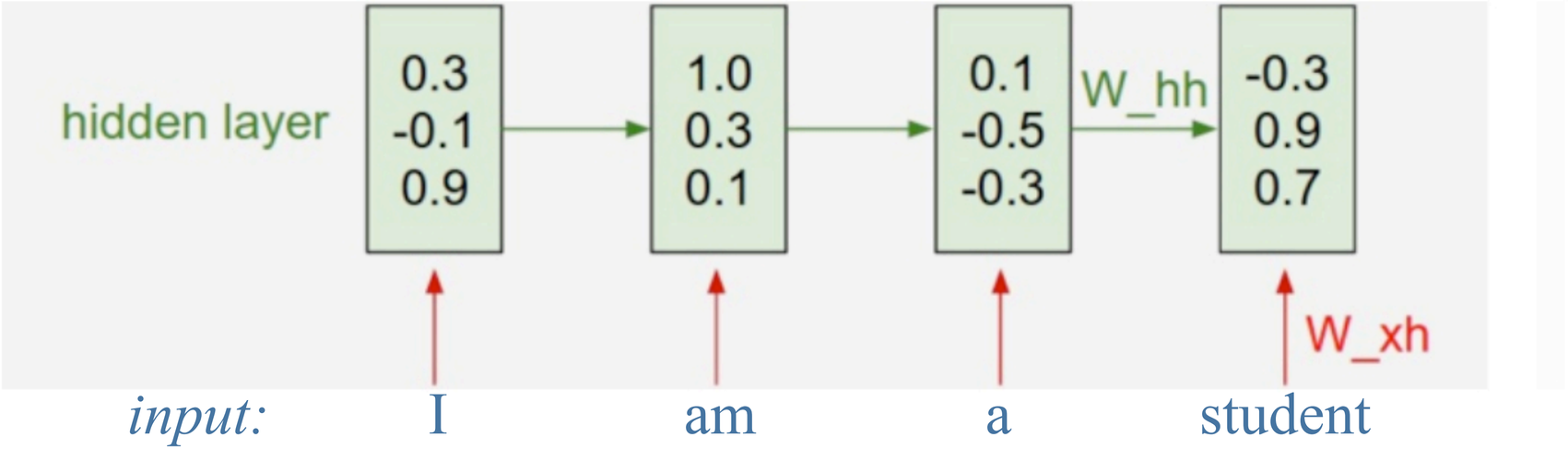
(Sutskever et al., 2014; Cho et al., 2014)

- *Sequence-to-sequence*: translate globally.
- *End-to-end*: simple & generalizable.

Outline

- Basic NMT
  - RNN Recap.
  - Encoder-Decoder.
  - Training.
  - Testing.
- Advanced NMT

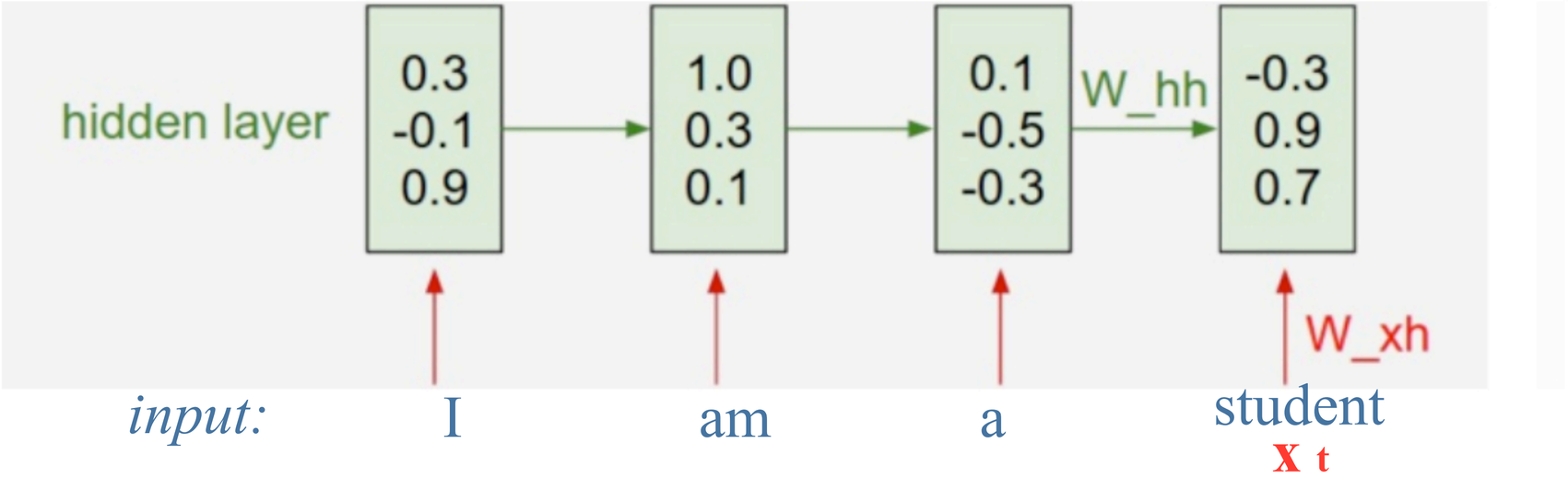
Recurrent Neural Networks (RNNs)



Recurrent Neural Networks (RNNs)

$$\boldsymbol{h}_t = \sigma\left(\boldsymbol{W}_{\boldsymbol{xh}}\boldsymbol{x}_t + \boldsymbol{W}_{\boldsymbol{hh}}\boldsymbol{h}_{t-1}\right)$$

h t-1 h t



RNNs to represent sequences!

Neural Machine Translation (NMT)

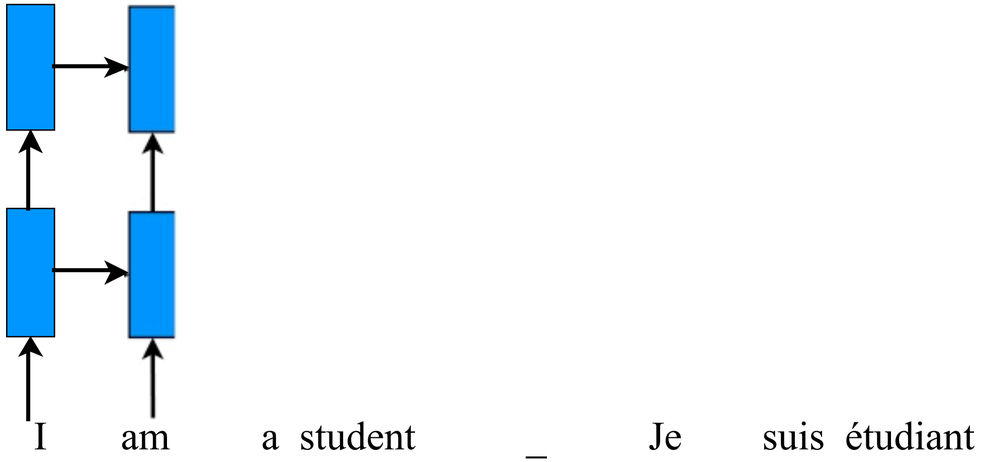
I am a student _ Je suis étudiant

- Recurrent Neural Networks:
  - Model P(target | source) directly.
  - Can be trained end-to-end.

Neural Machine Translation (NMT)

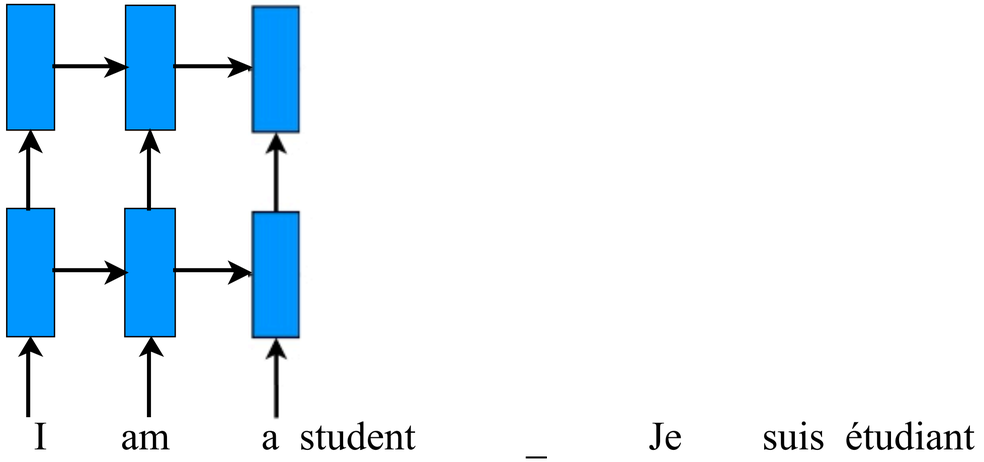
- Recurrent Neural Networks:
  - Model P(target | source) directly.
  - Can be trained end-to-end.

Neural Machine Translation (NMT)



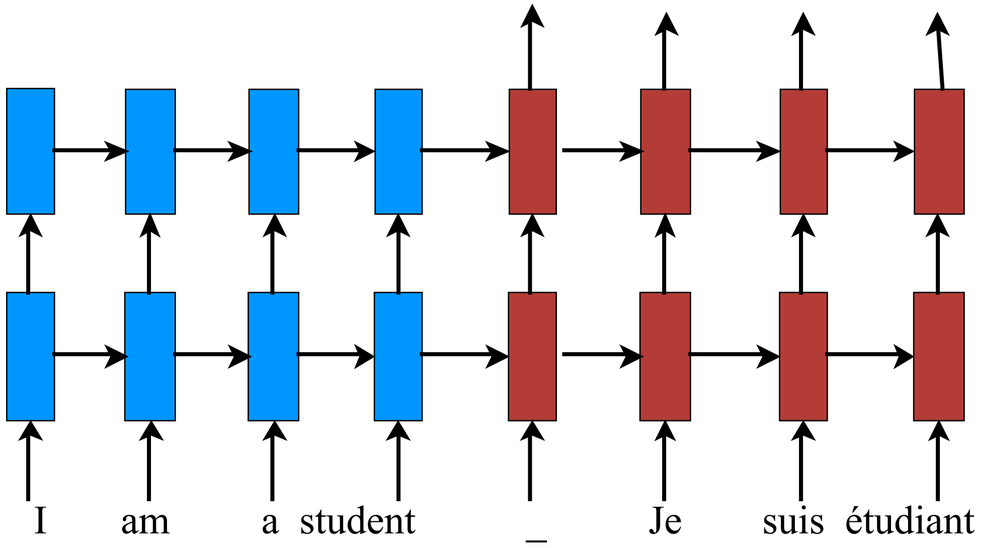
- Recurrent Neural Networks:
  - Model P(target | source) directly.
  - Can be trained end-to-end.

Neural Machine Translation (NMT)



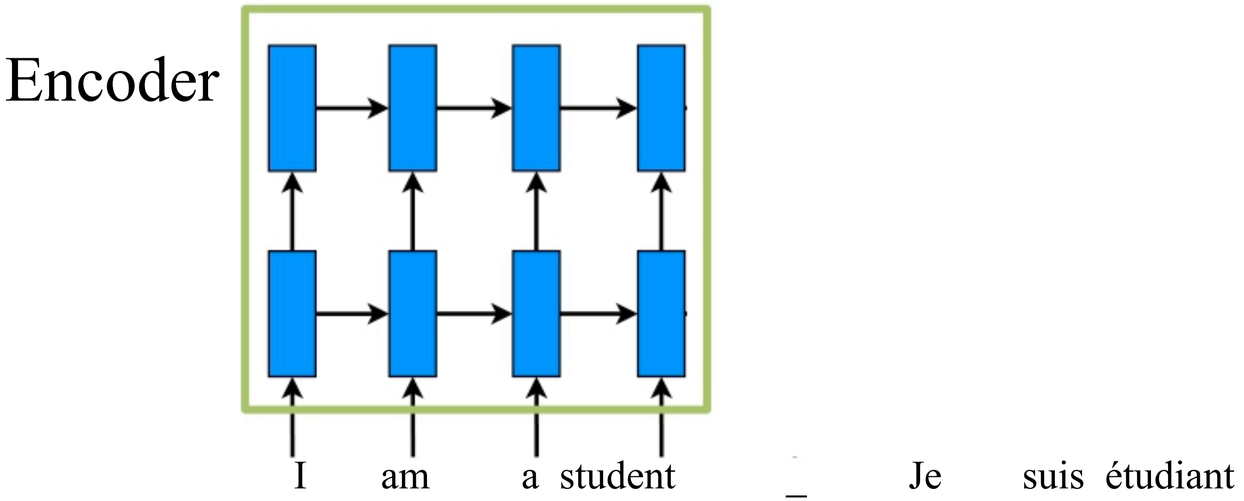
- Recurrent Neural Networks:
  - Model P(target | source) directly.
  - Can be trained end-to-end.

Neural Machine Translation (NMT)



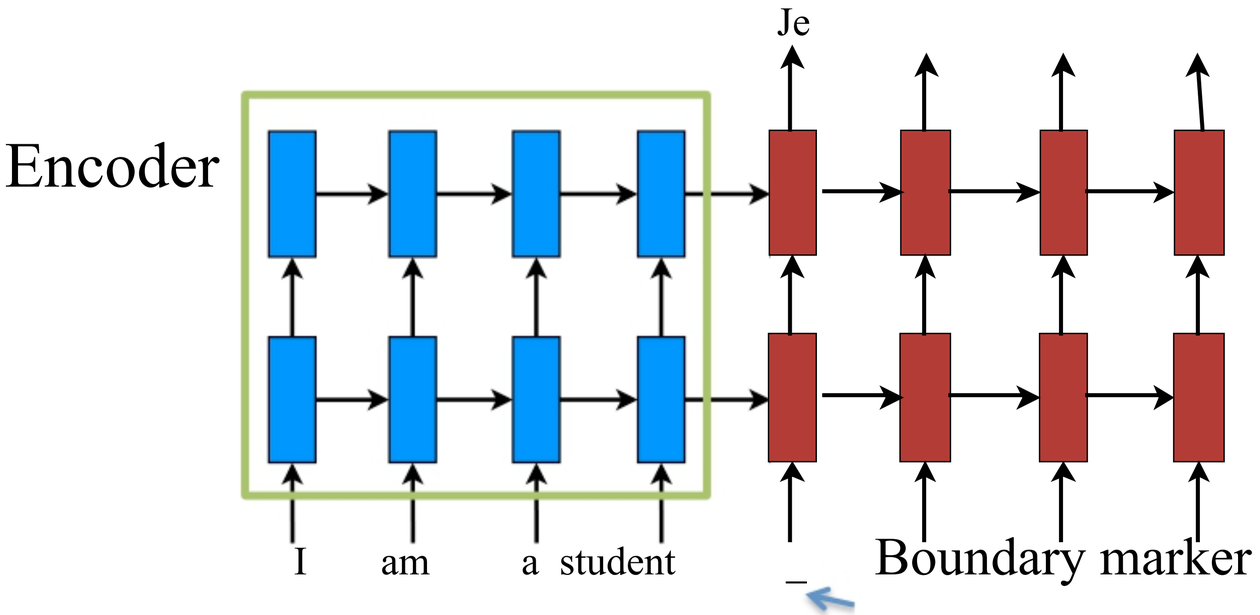
- Recurrent Neural Networks:
  - Model P(target | source) directly.
  - Can be trained end-to-end.

Neural Machine Translation (NMT)



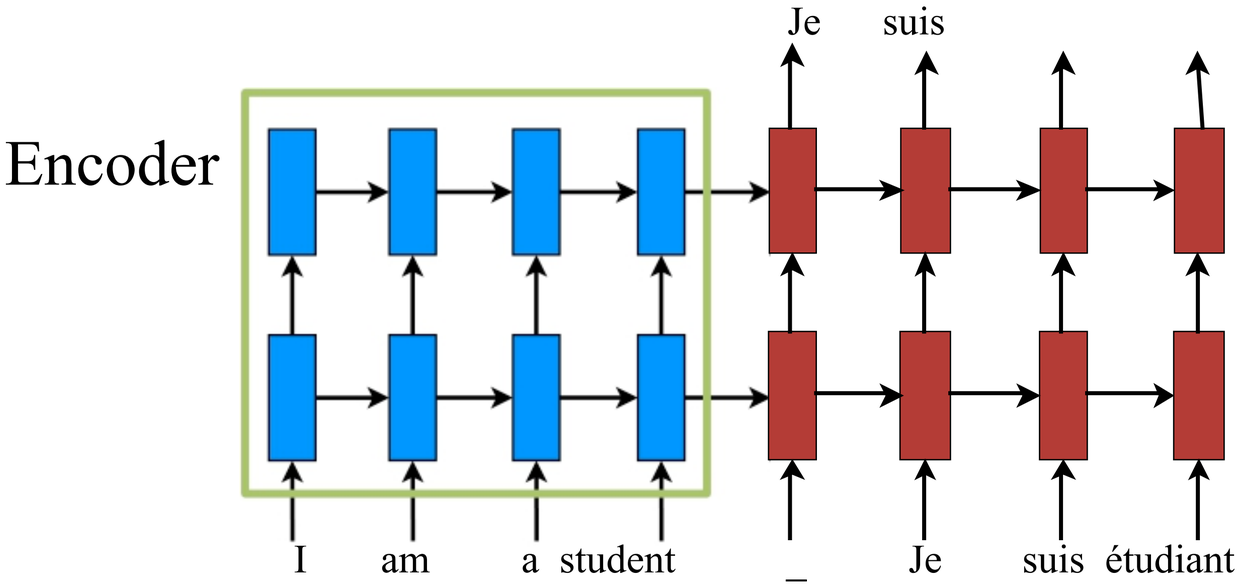
- Recurrent Neural Networks:
  - Model P(target | source) directly.
  - Can be trained end-to-end.

Neural Machine Translation (NMT)



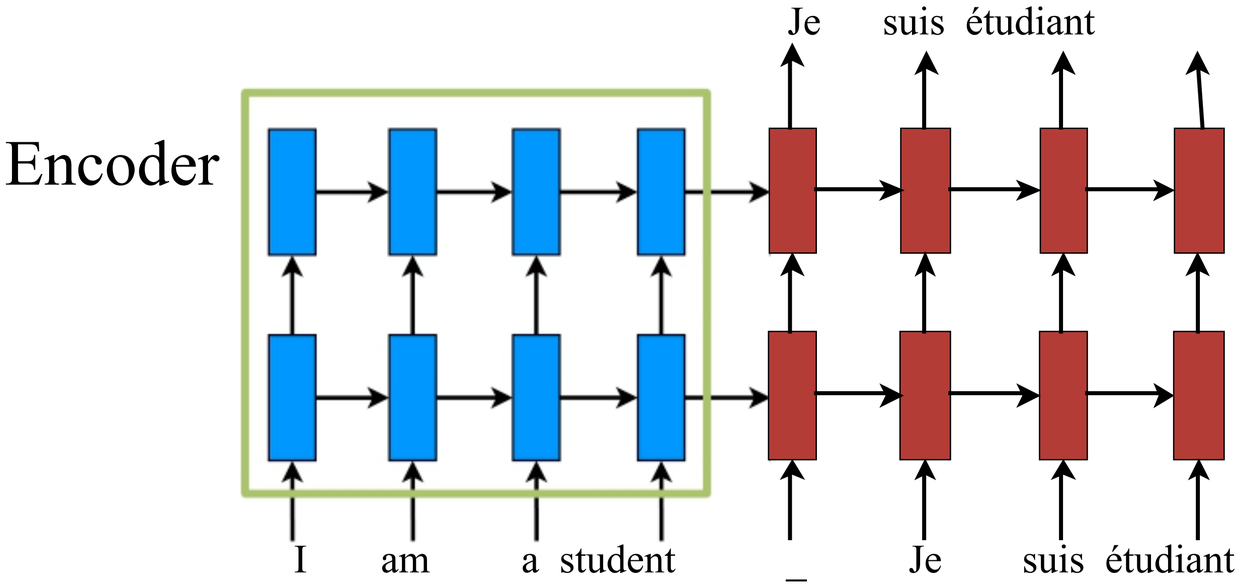
- Recurrent Neural Networks:
  - Model P(target | source) directly.
  - Can be trained end-to-end.

Neural Machine Translation (NMT)



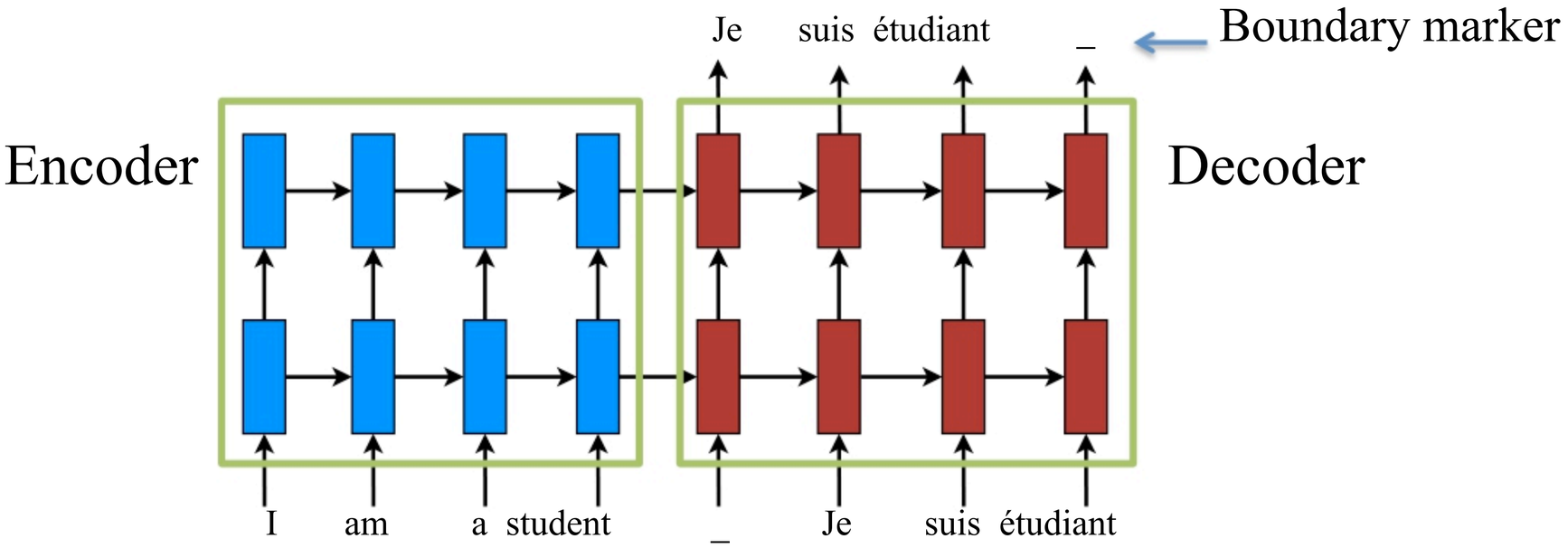
- Recurrent Neural Networks:
  - Model P(target | source) directly.
  - Can be trained end-to-end.

Neural Machine Translation (NMT)



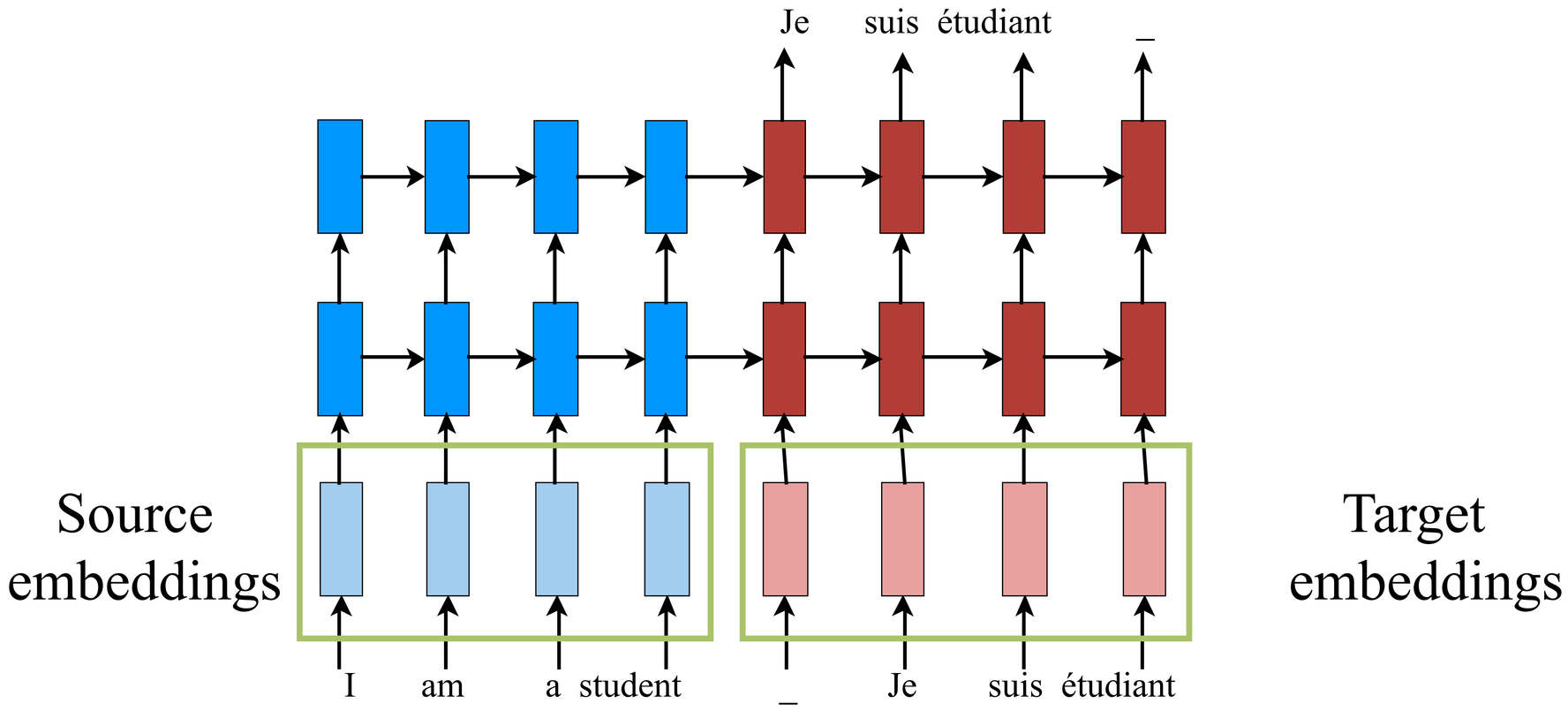
- Recurrent Neural Networks:
  - Model P(target | source) directly.
  - Can be trained end-to-end.

Neural Machine Translation (NMT)



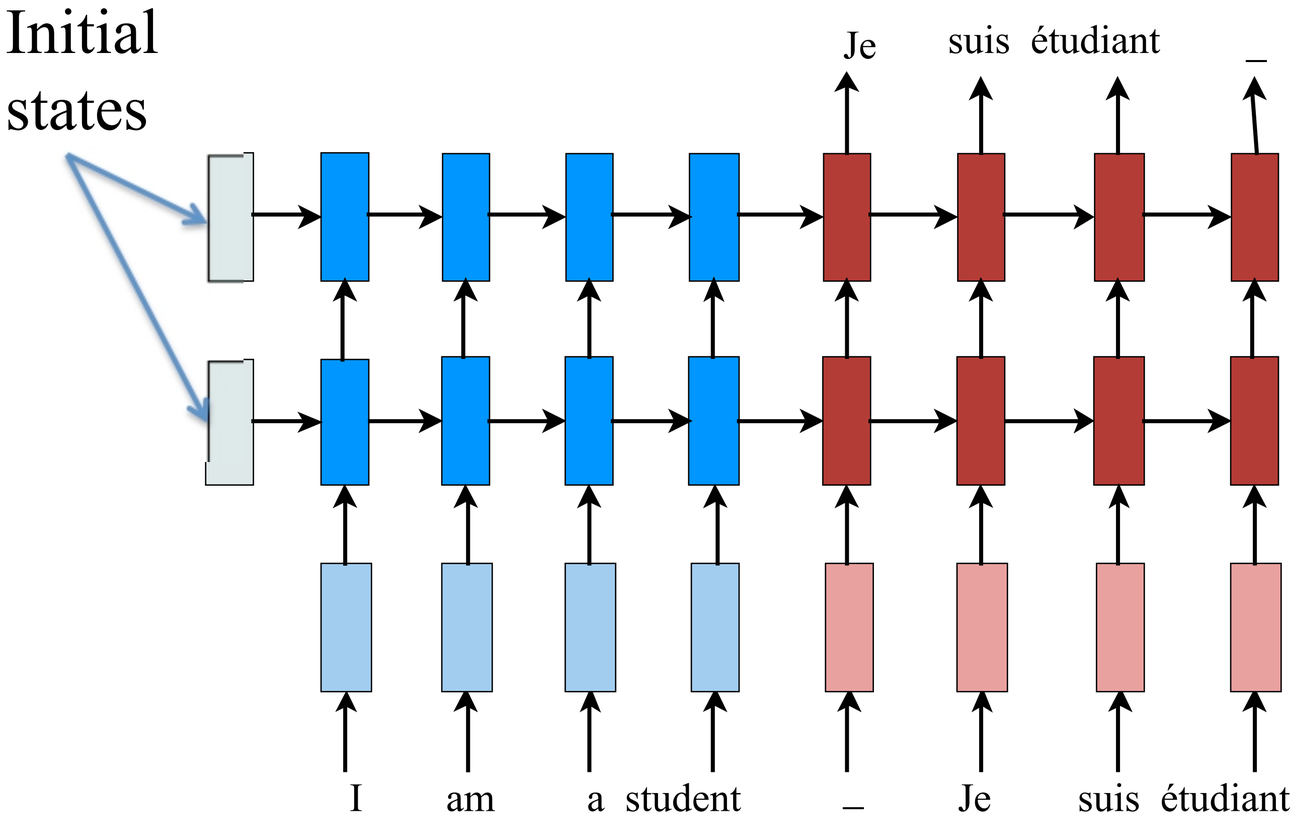
- Recurrent Neural Networks:
  - Model P(target | source) directly.
  - Can be trained end-to-end.

Word Embeddings



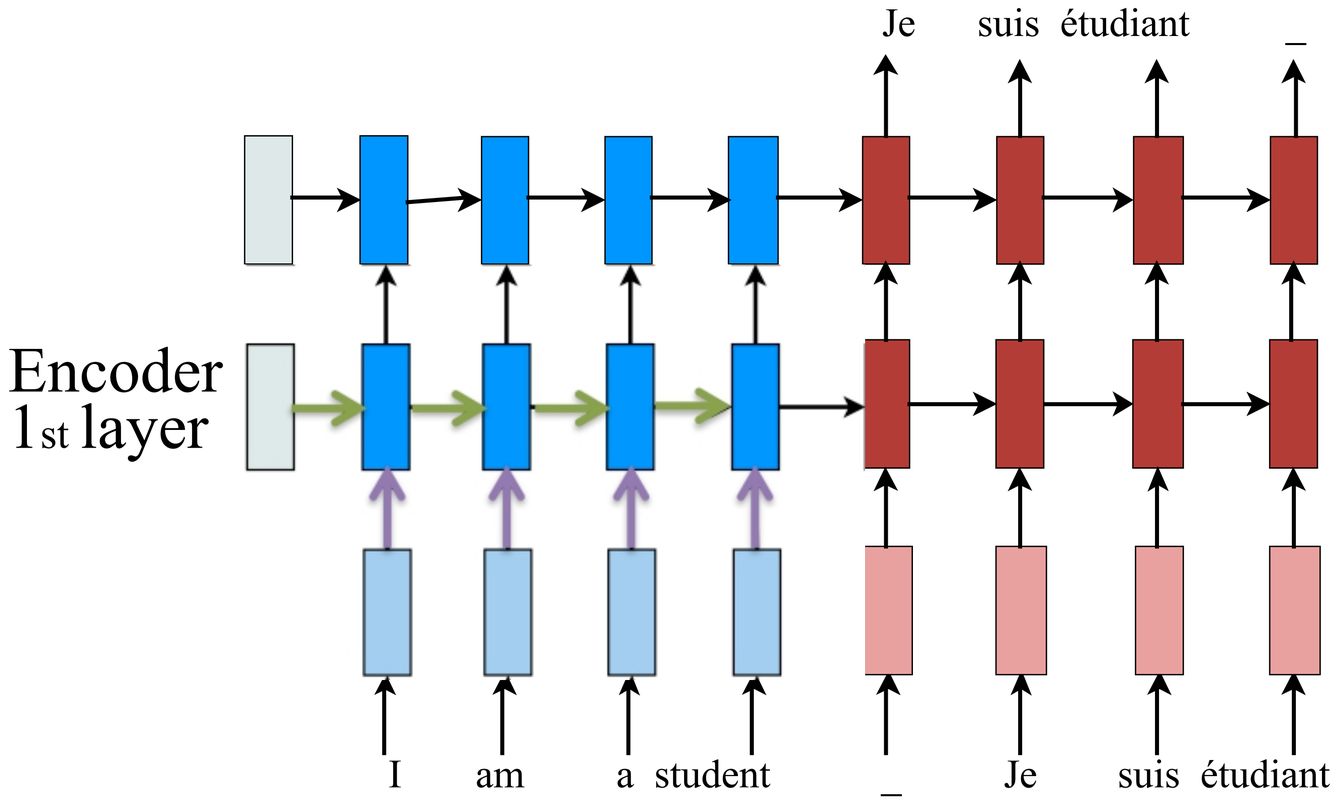
- One for each language: can learn from scratch.

Recurrent Connections



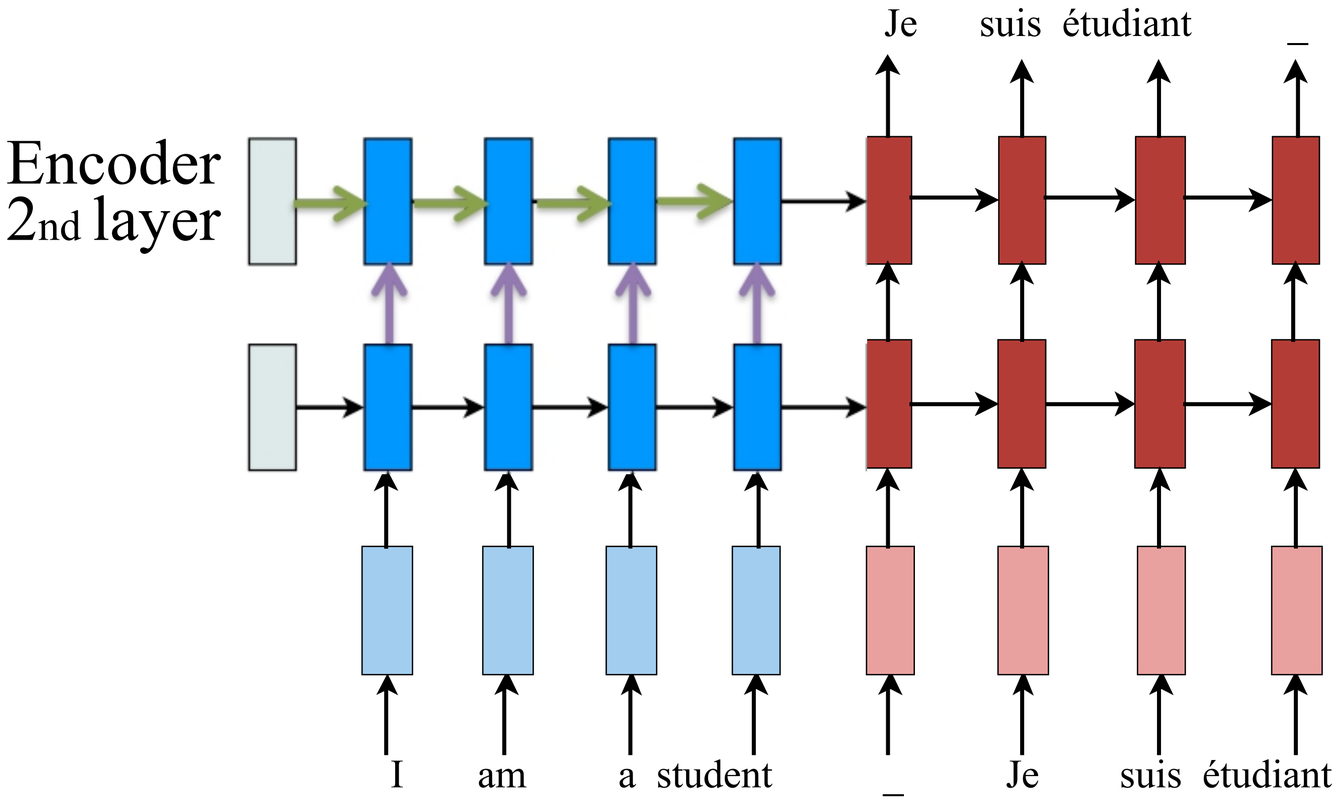
- Often set to 0.

Recurrent Connections



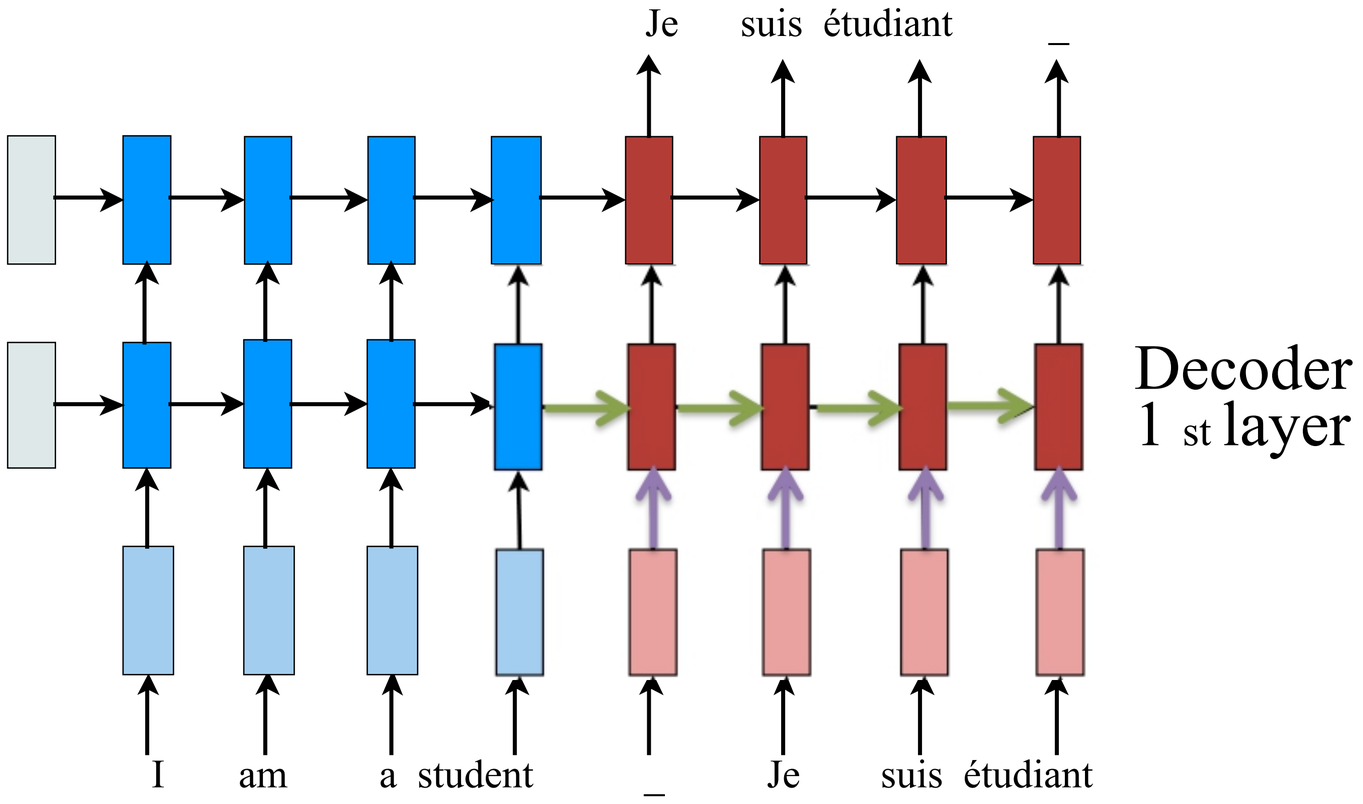
- Different : {1st layer, 2nd layer} x {encoder, decoder}.

Recurrent Connections



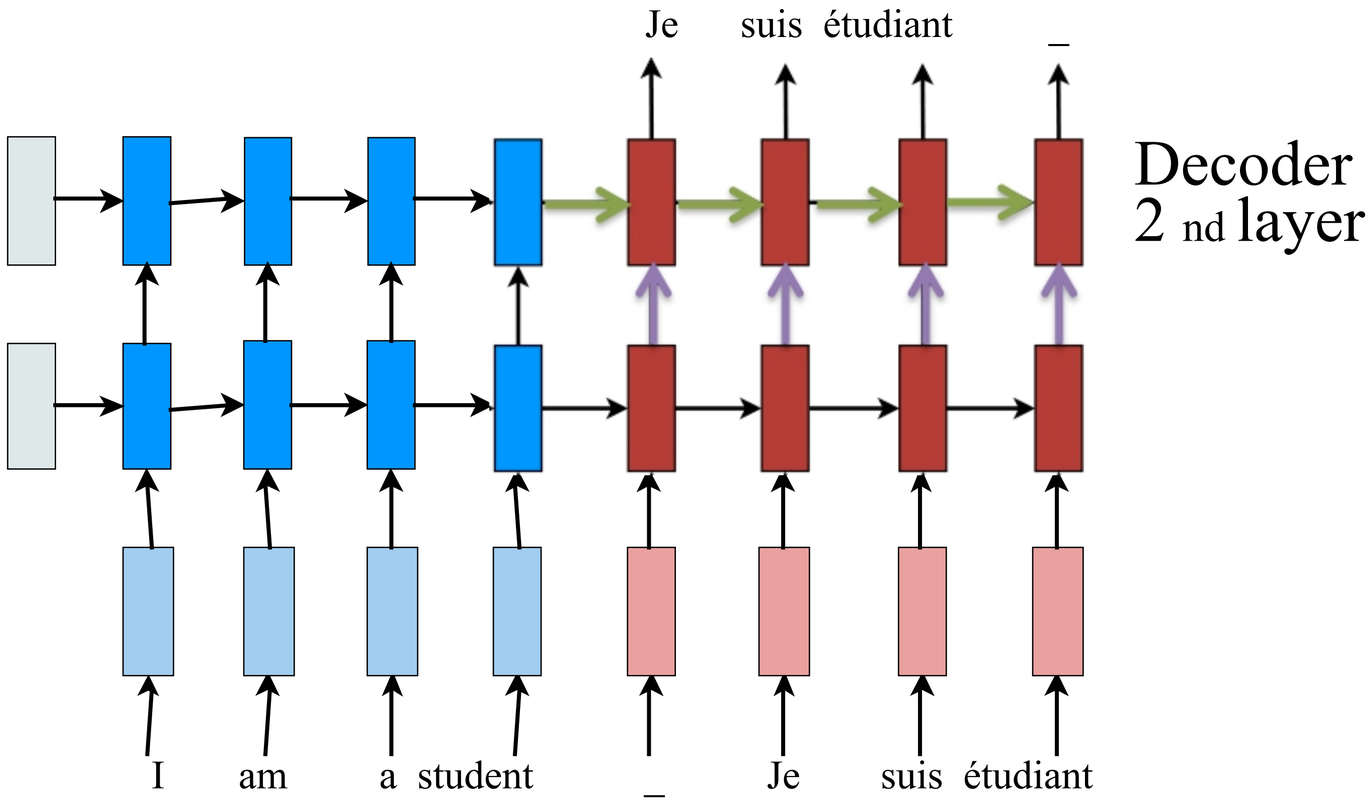
- Different : {1st layer, 2nd layer} x {encoder, decoder}.

Recurrent Connections



- Different : {1st layer, 2nd layer} x {encoder, decoder}.

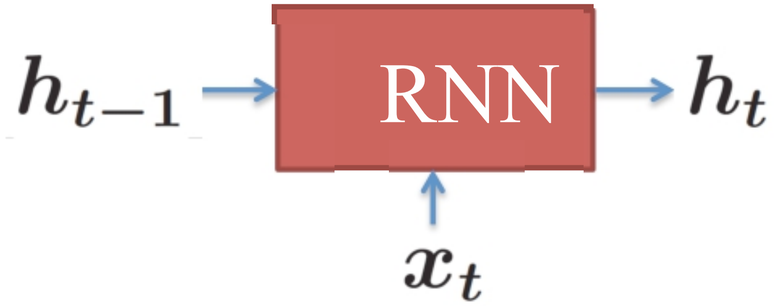
Recurrent Connections



- Different : {1st layer, 2nd layer} x {encoder, decoder}.

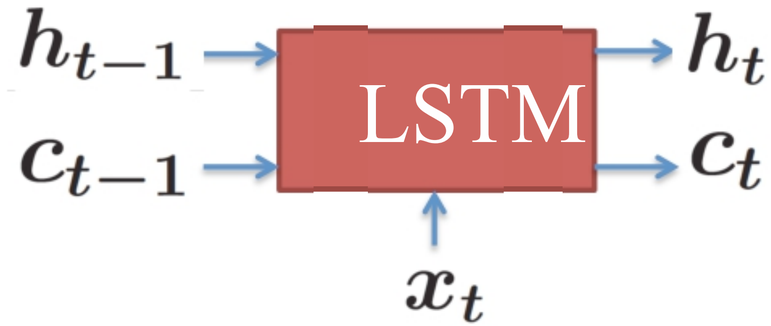
Recurrent Units

- Vanilla:

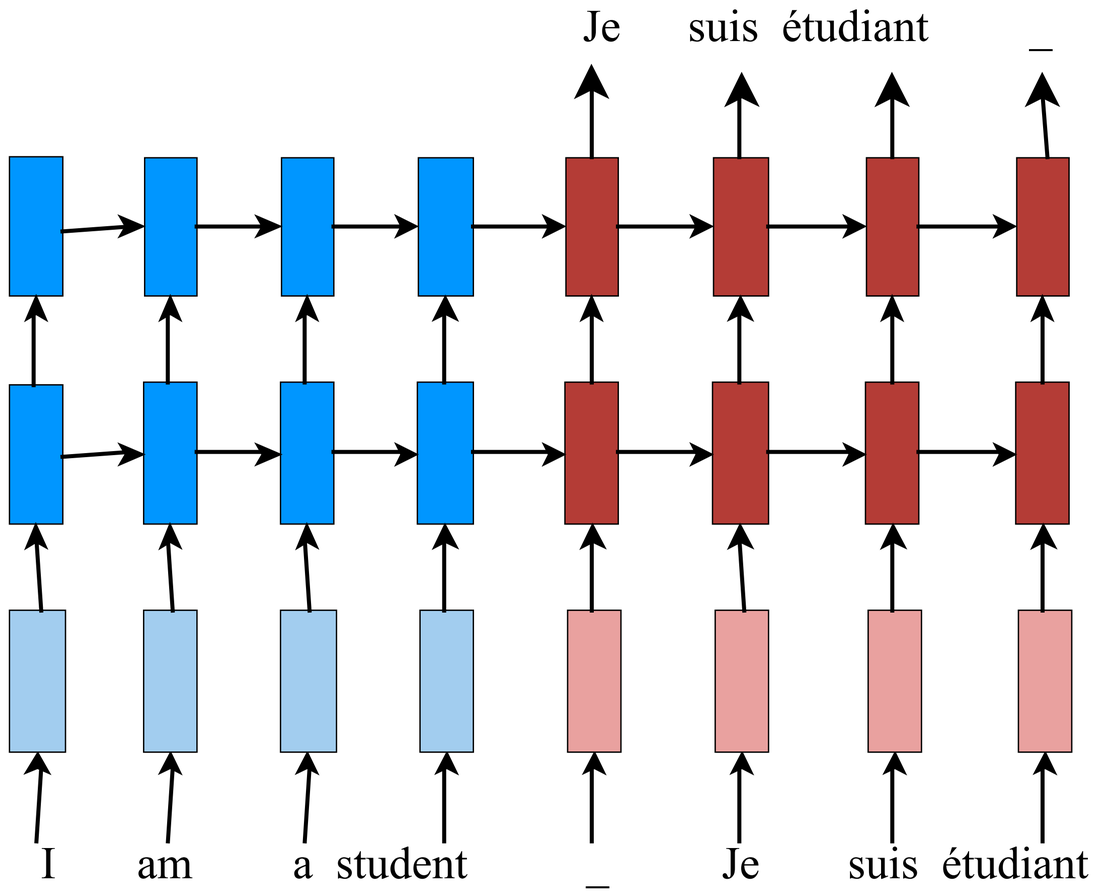


Vanishing gradient problem!

- LSTM:

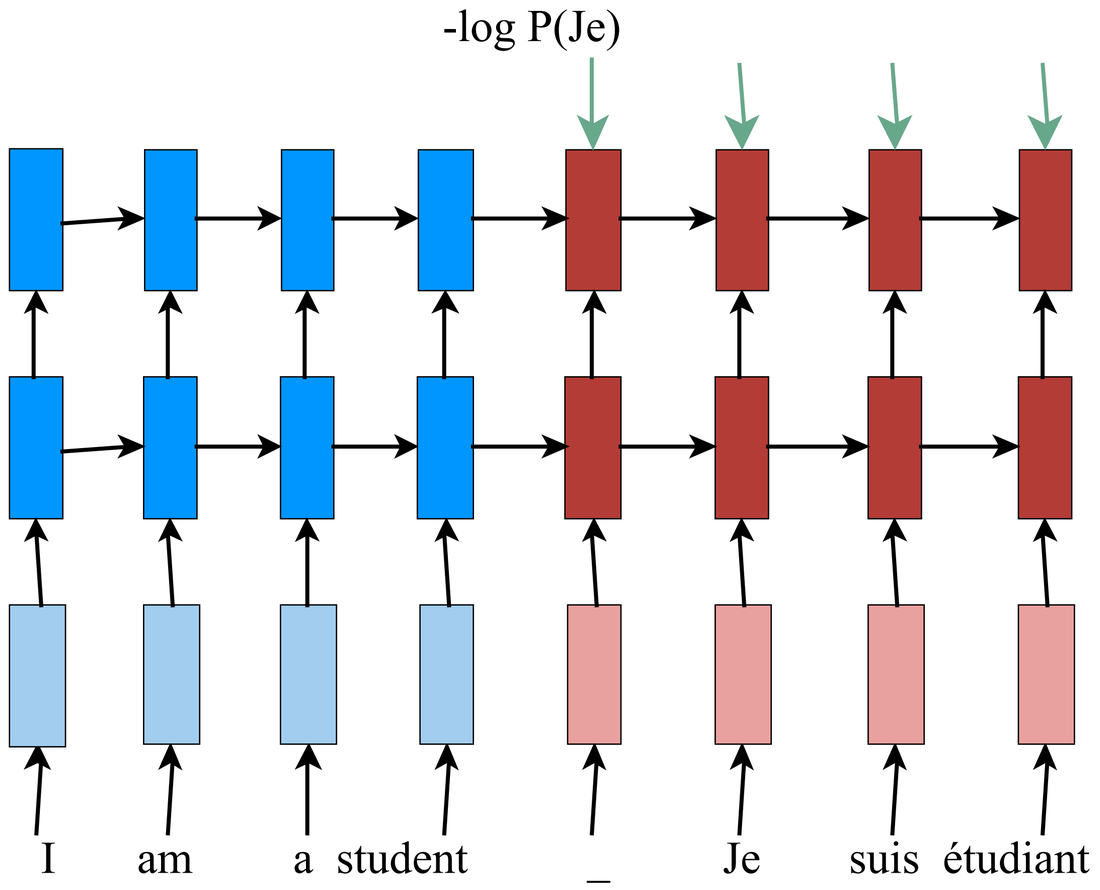


Training Loss



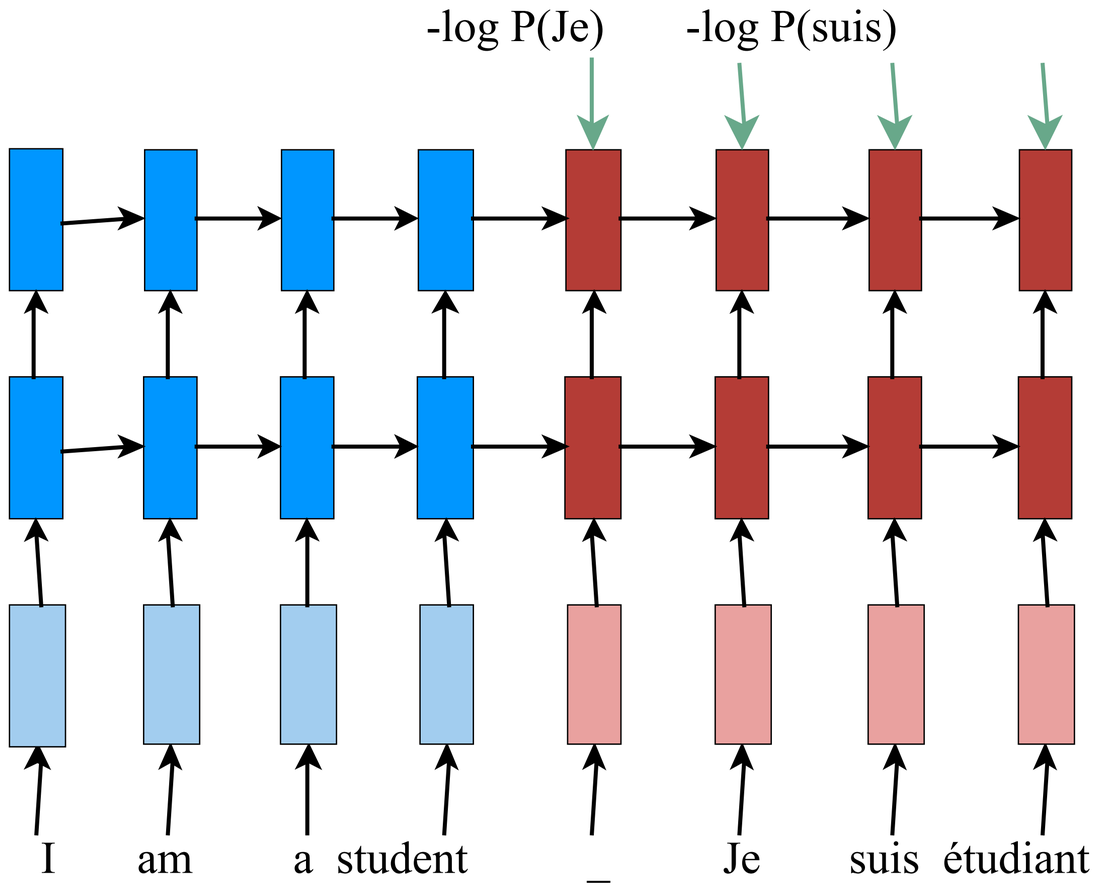
- Maximize P(target | source)

Training Loss



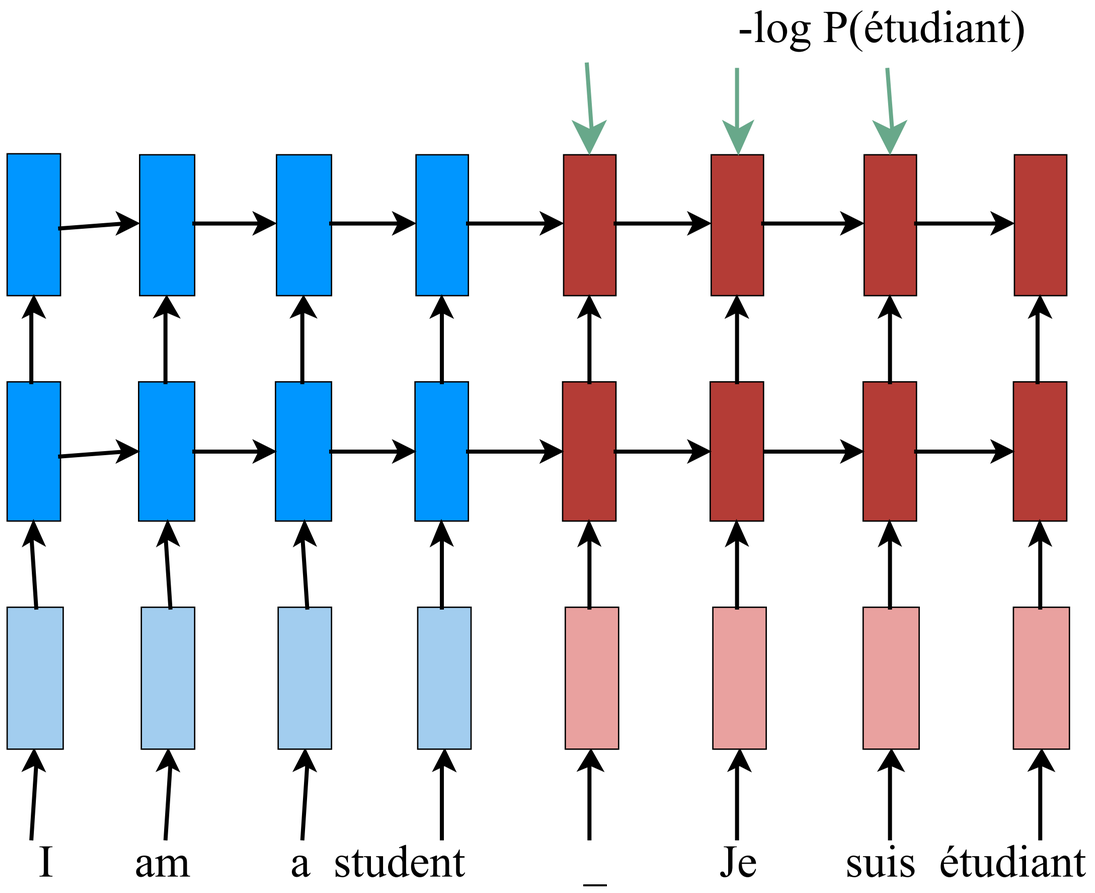
- Sum of all individual losses

Training Loss



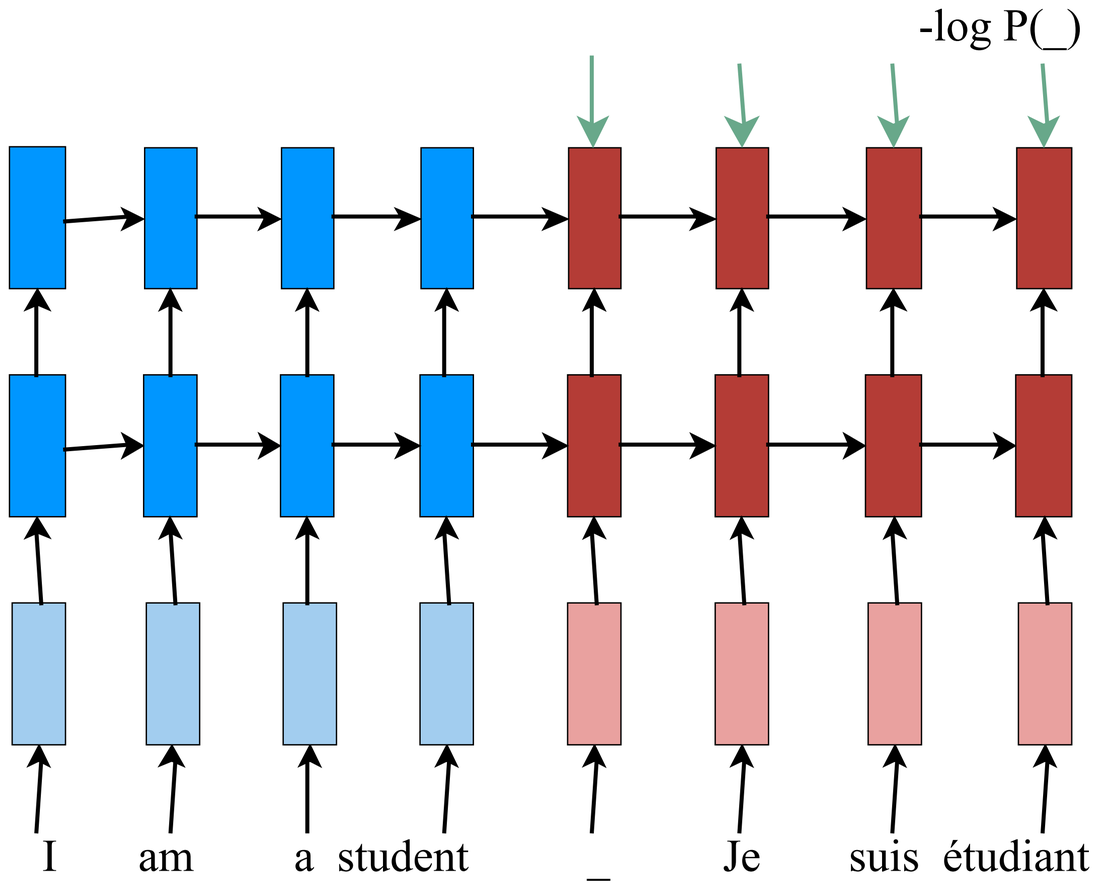
- Sum of all individual losses

Training Loss



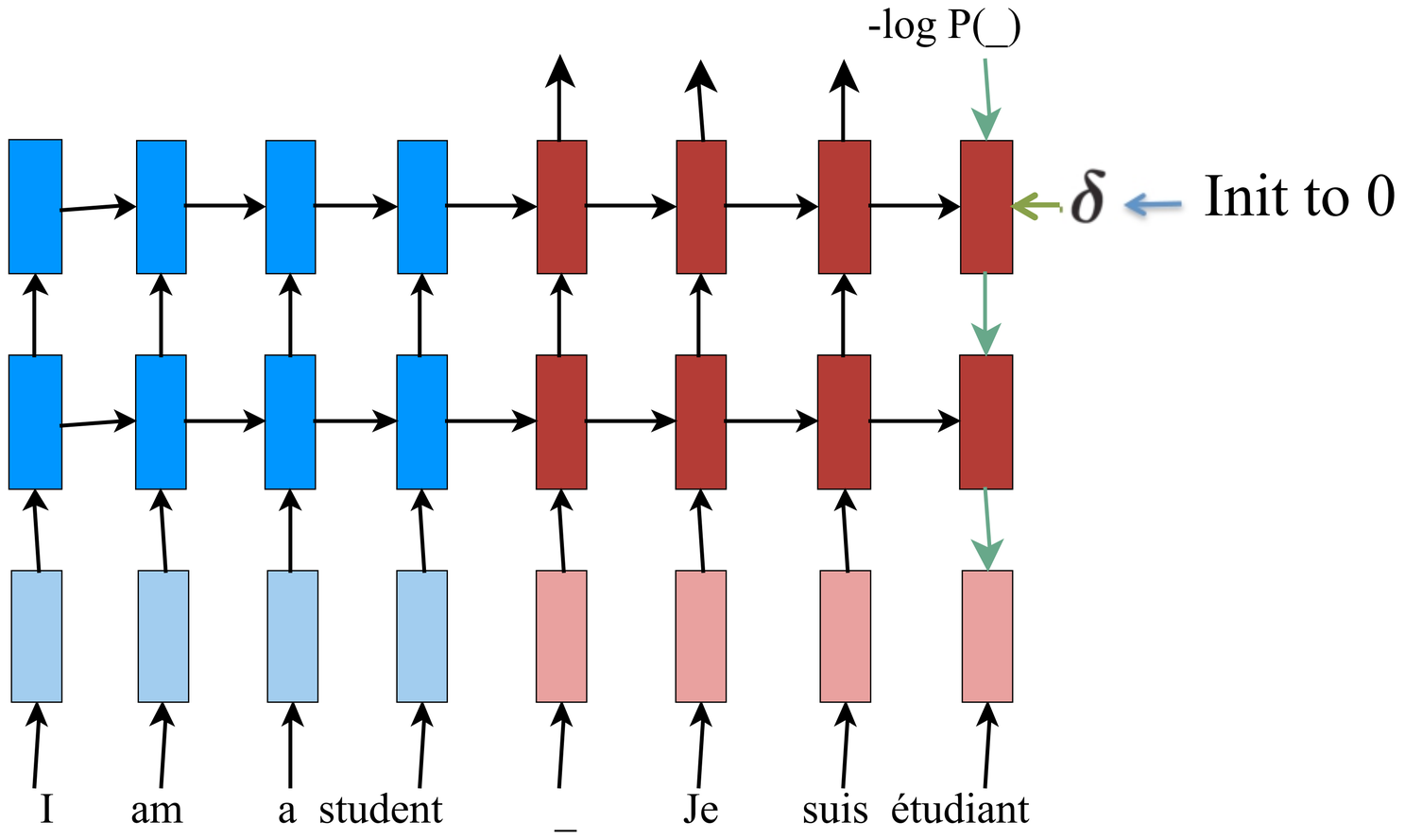
- Sum of all individual losses

Training Loss

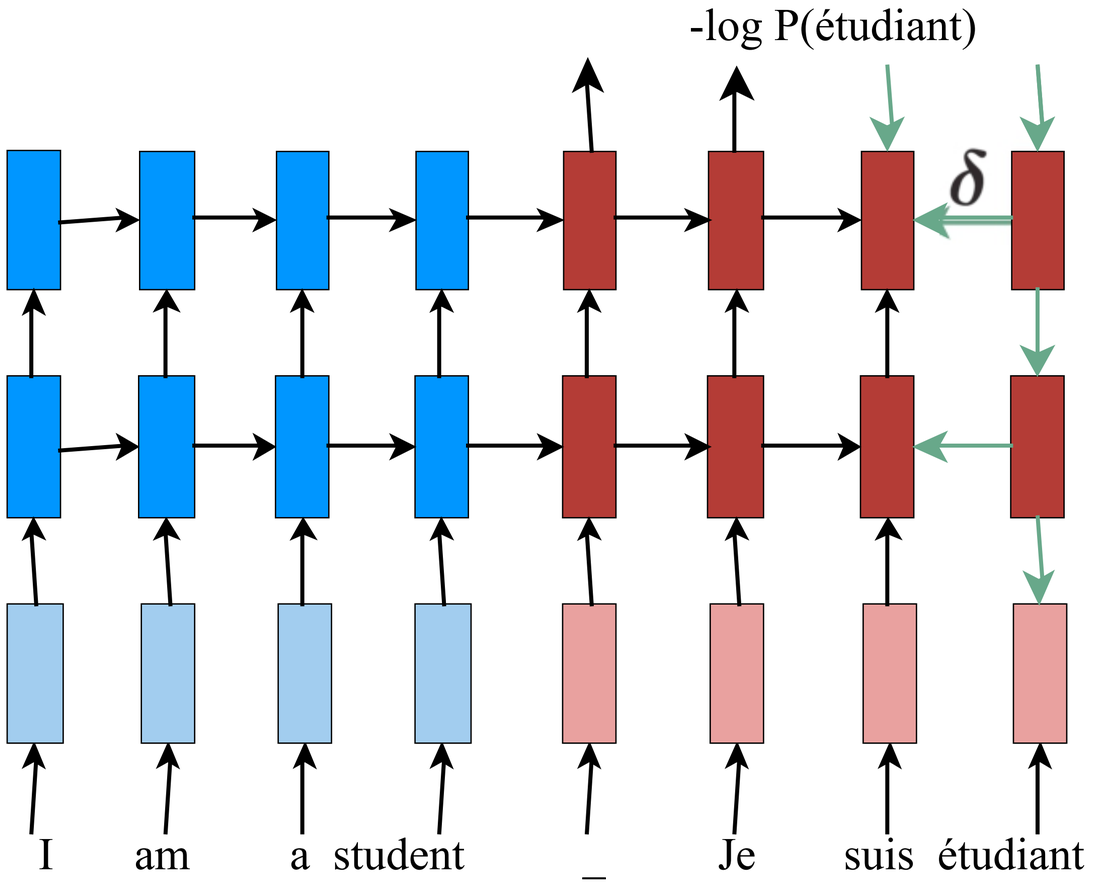


- Sum of all individual losses

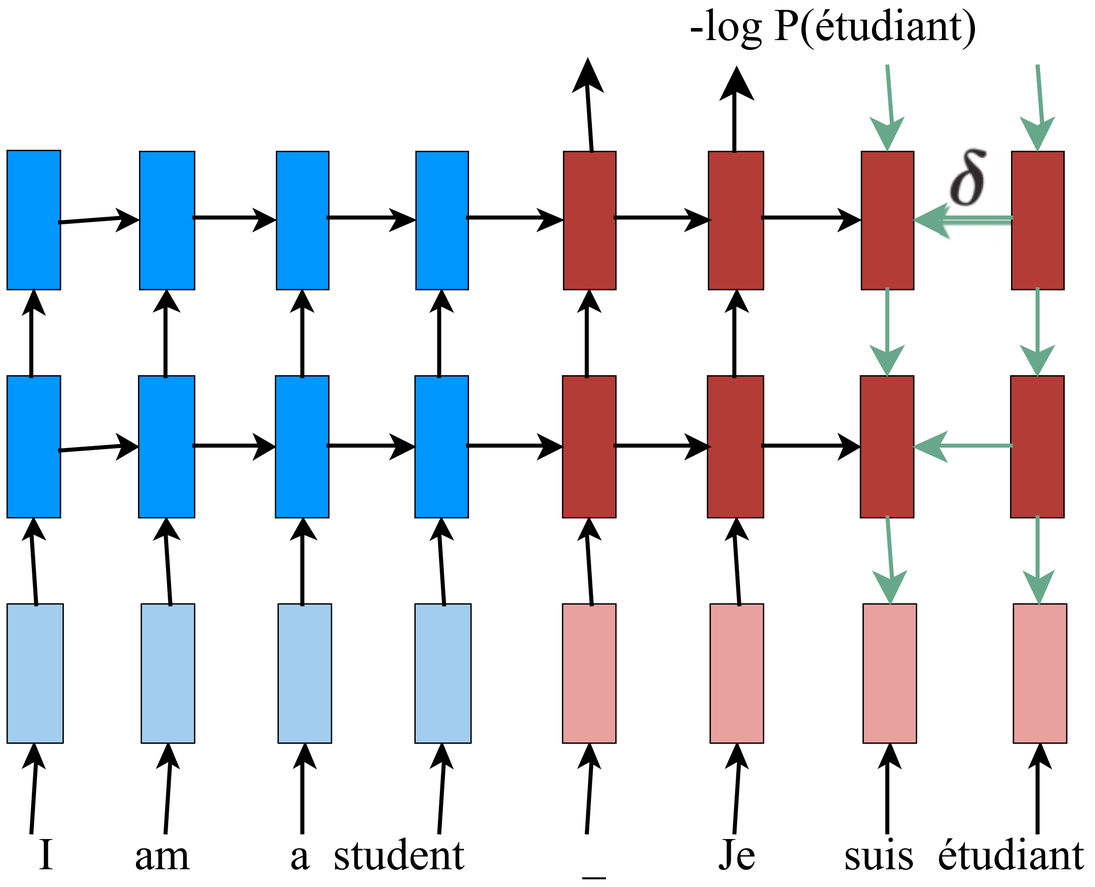
Backpropagation Through Time



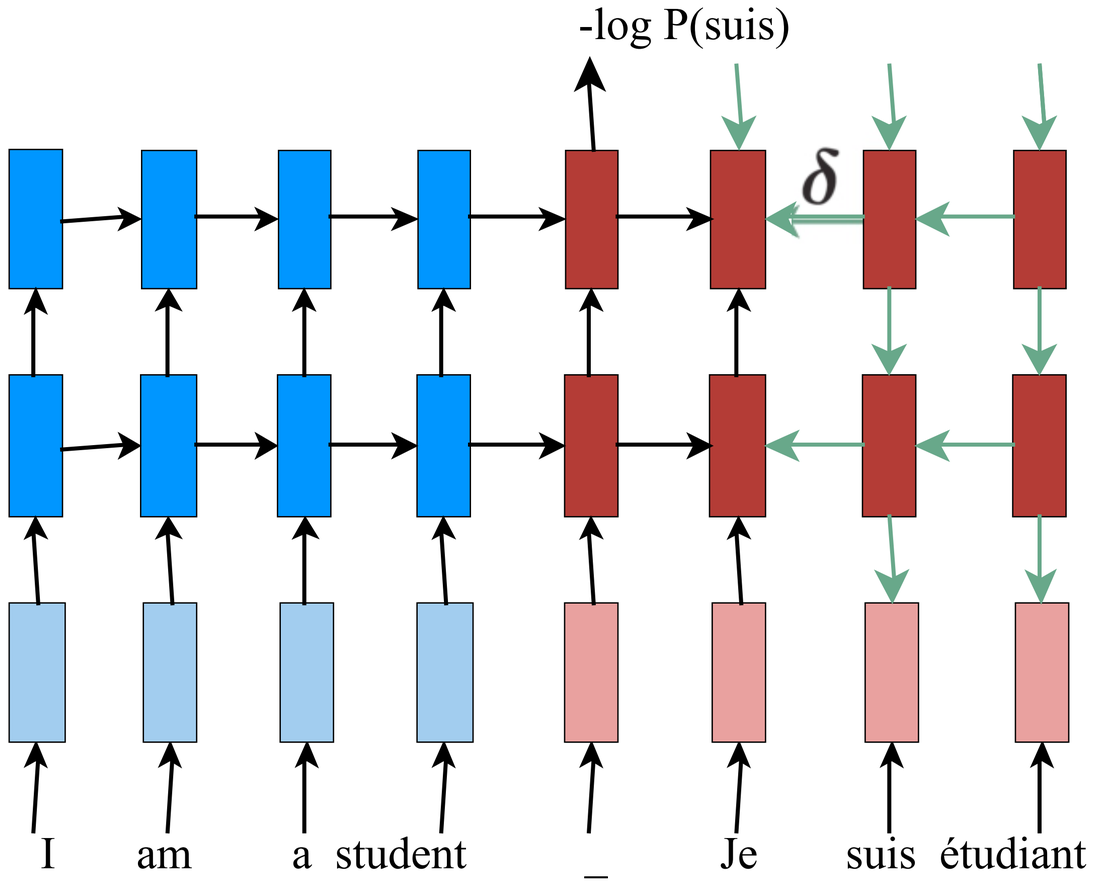
Backpropagation Through Time



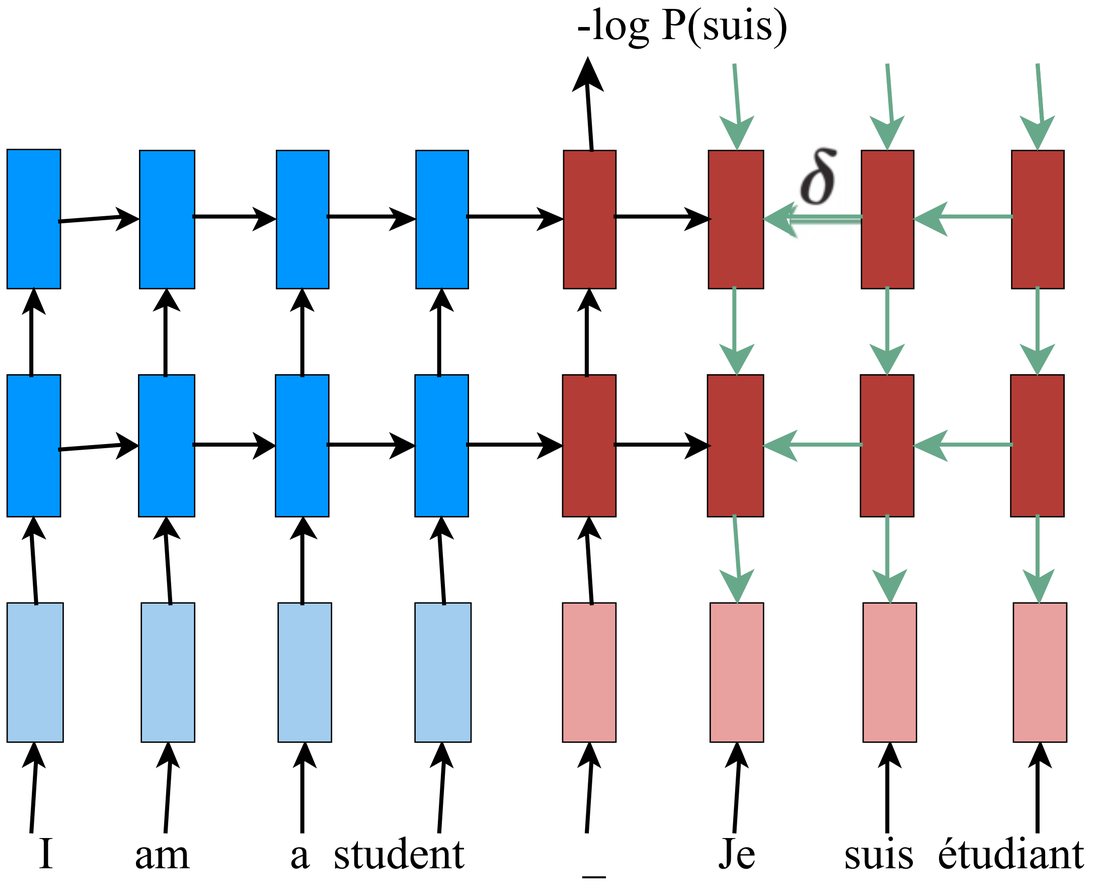
Backpropagation Through Time



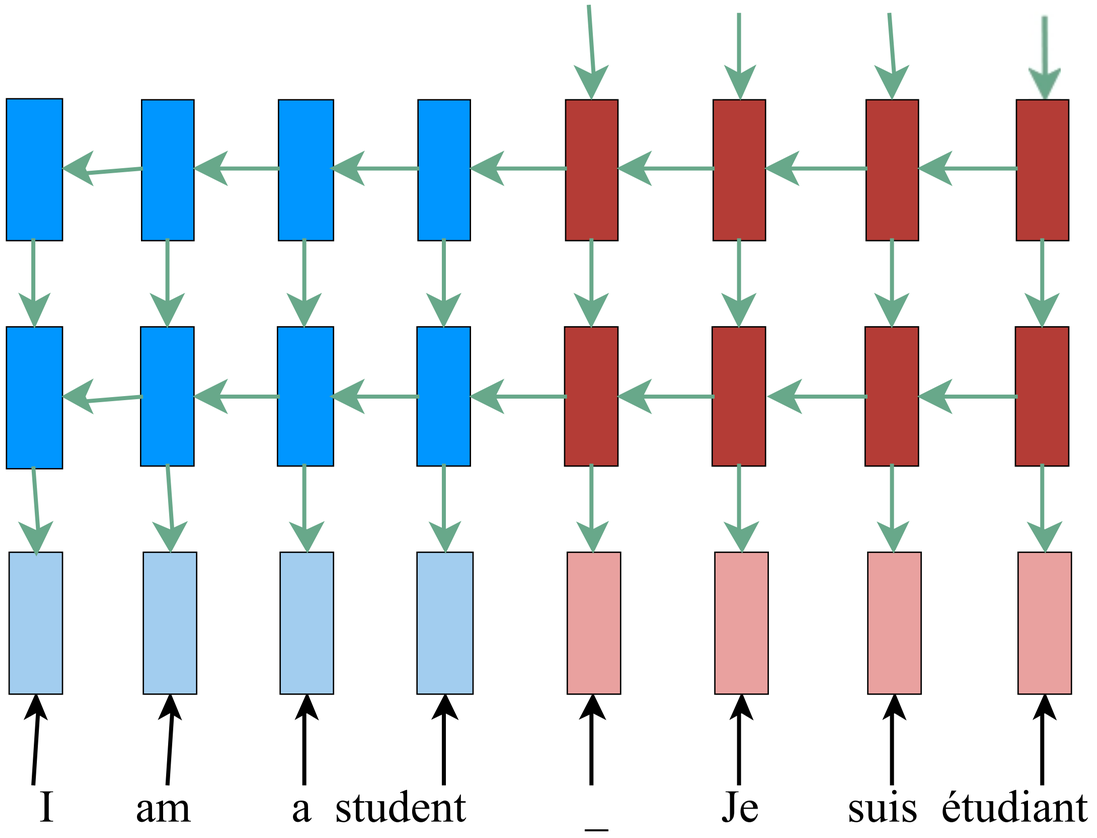
Backpropagation Through Time



Backpropagation Through Time



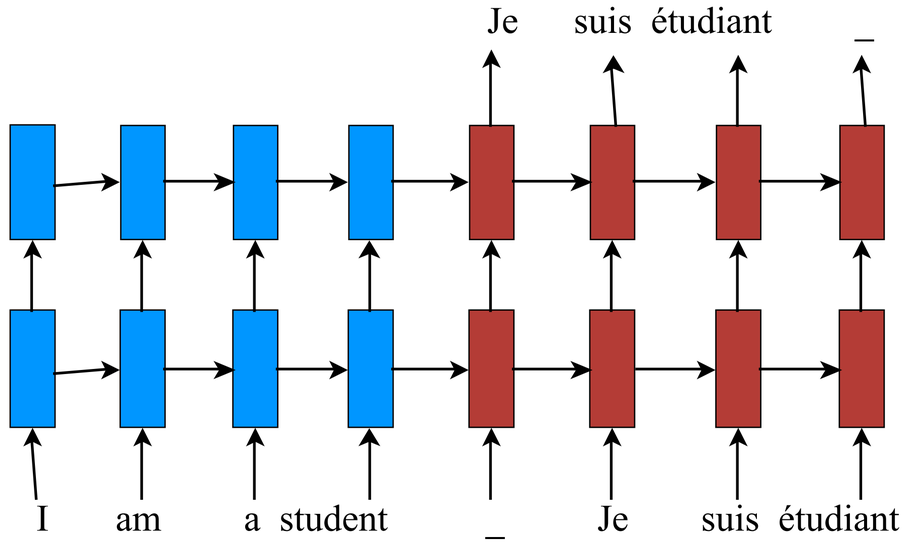
Backpropagation Through Time



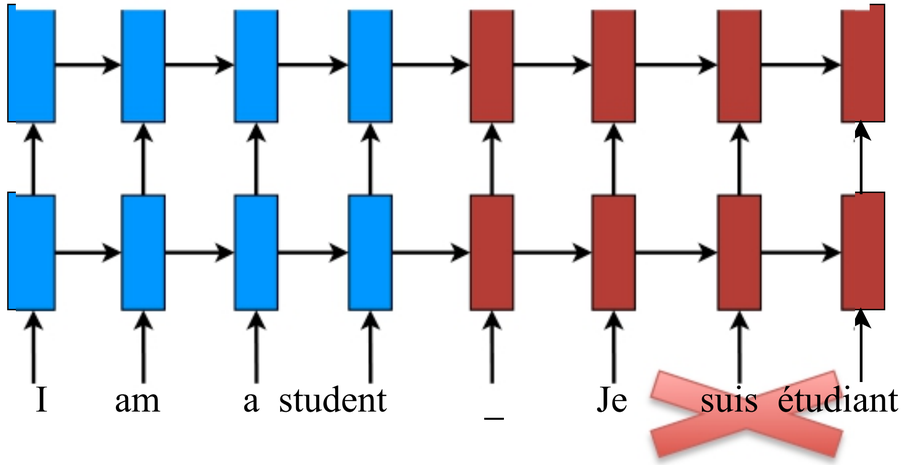
RNN gradients are accumulated.

Training vs. Testing

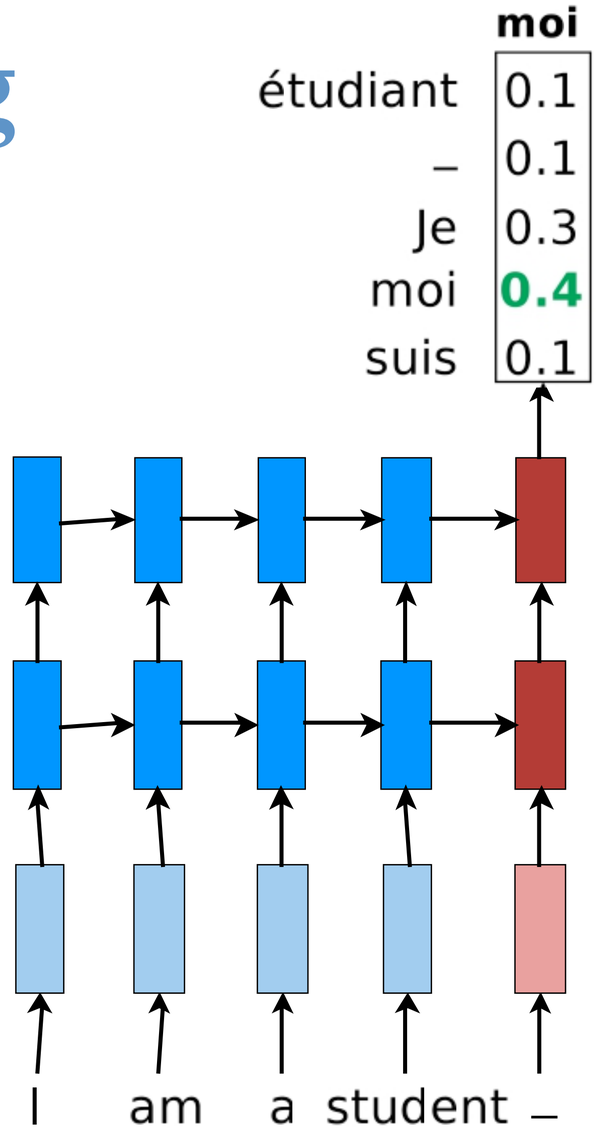
- *Training*
  - Correct translations are available.



- *Testing*
  - Only source sentences are given.

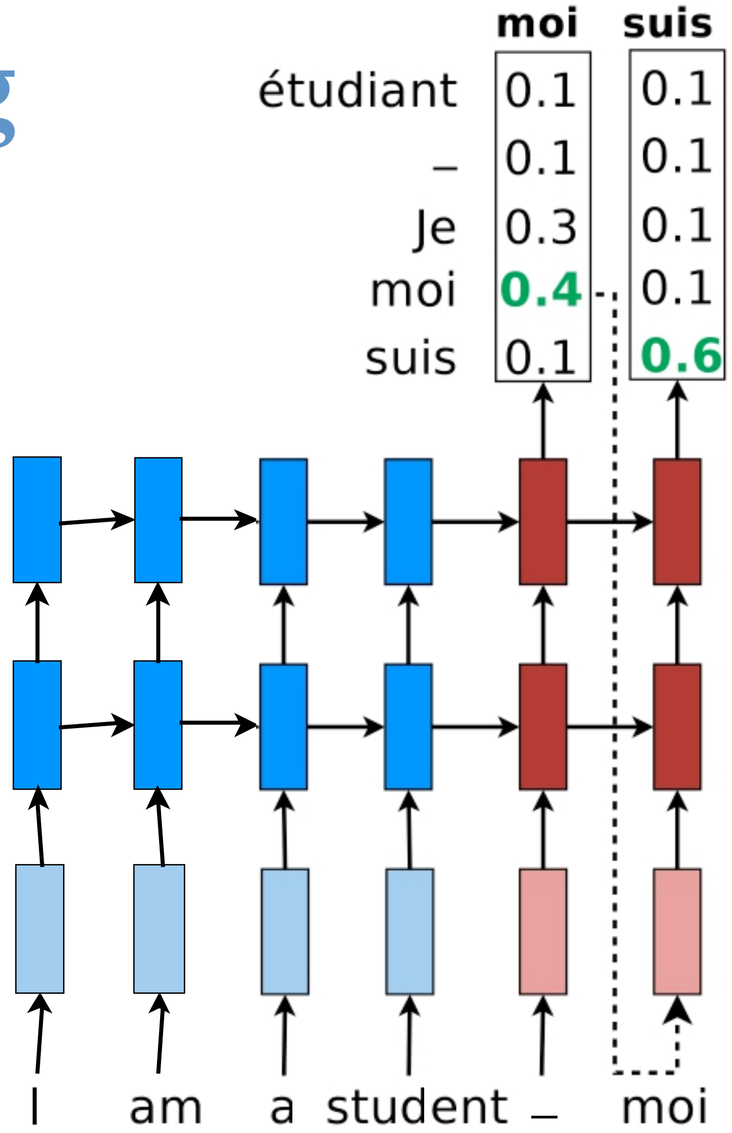


Testing



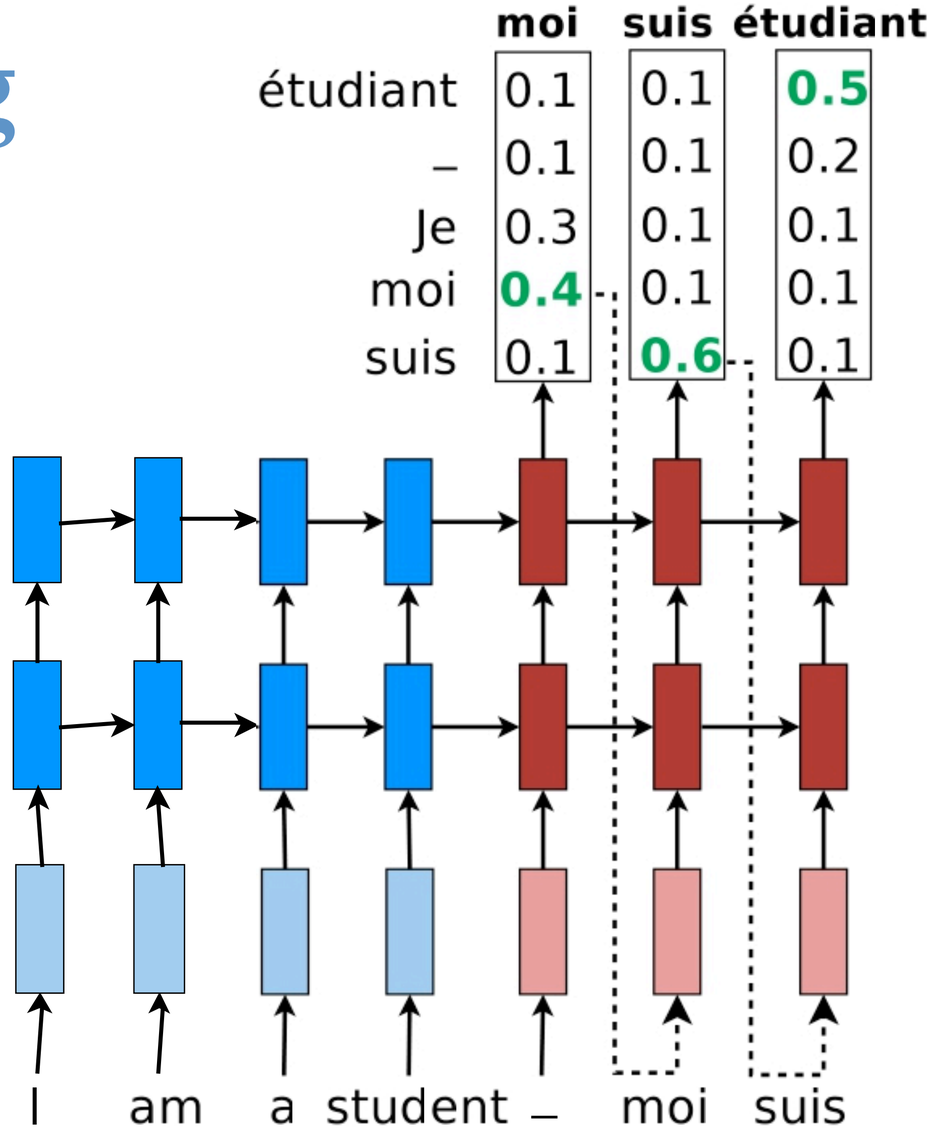
- Feed the most likely word

Testing



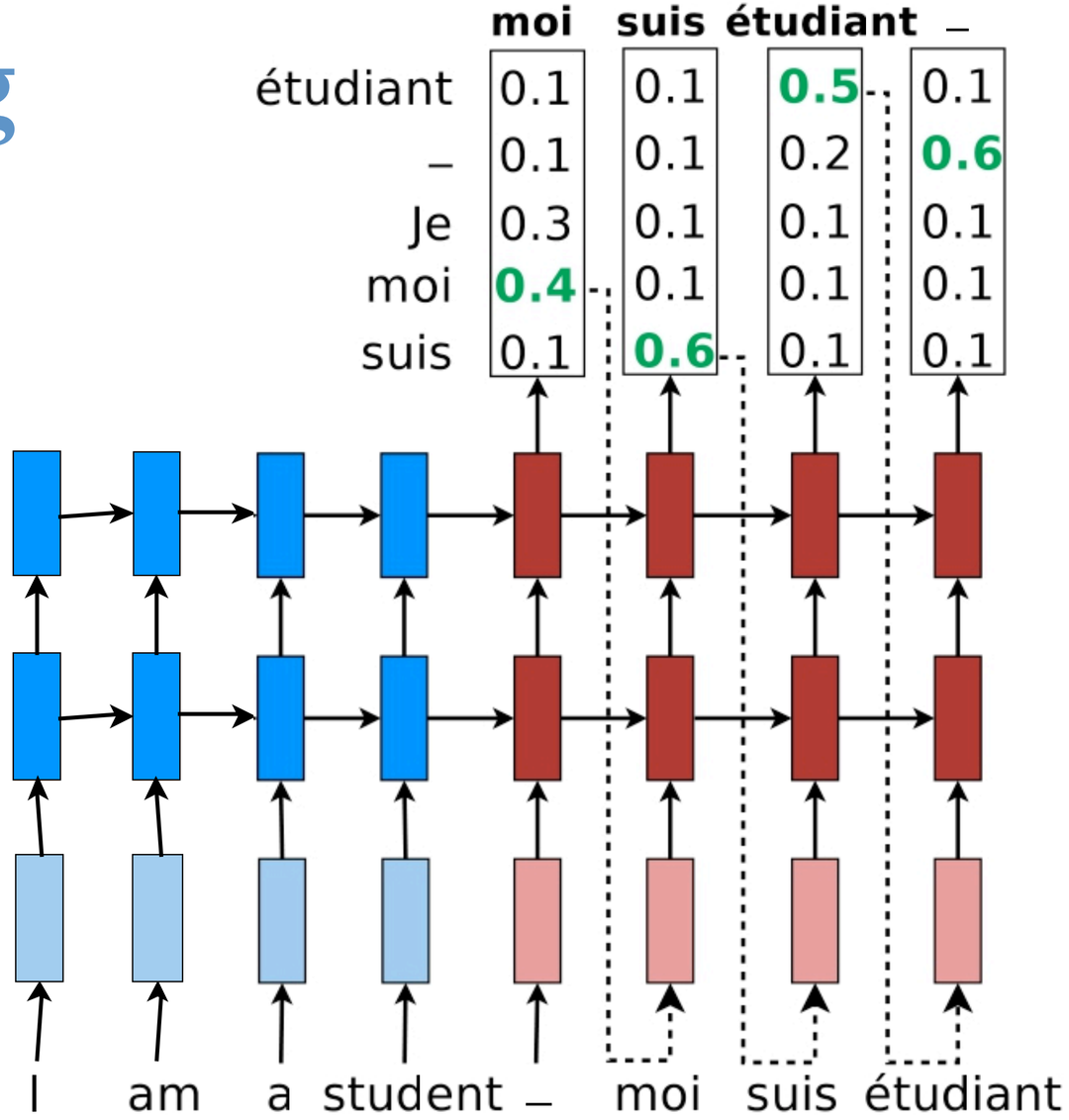
- Feed the most likely word

Testing



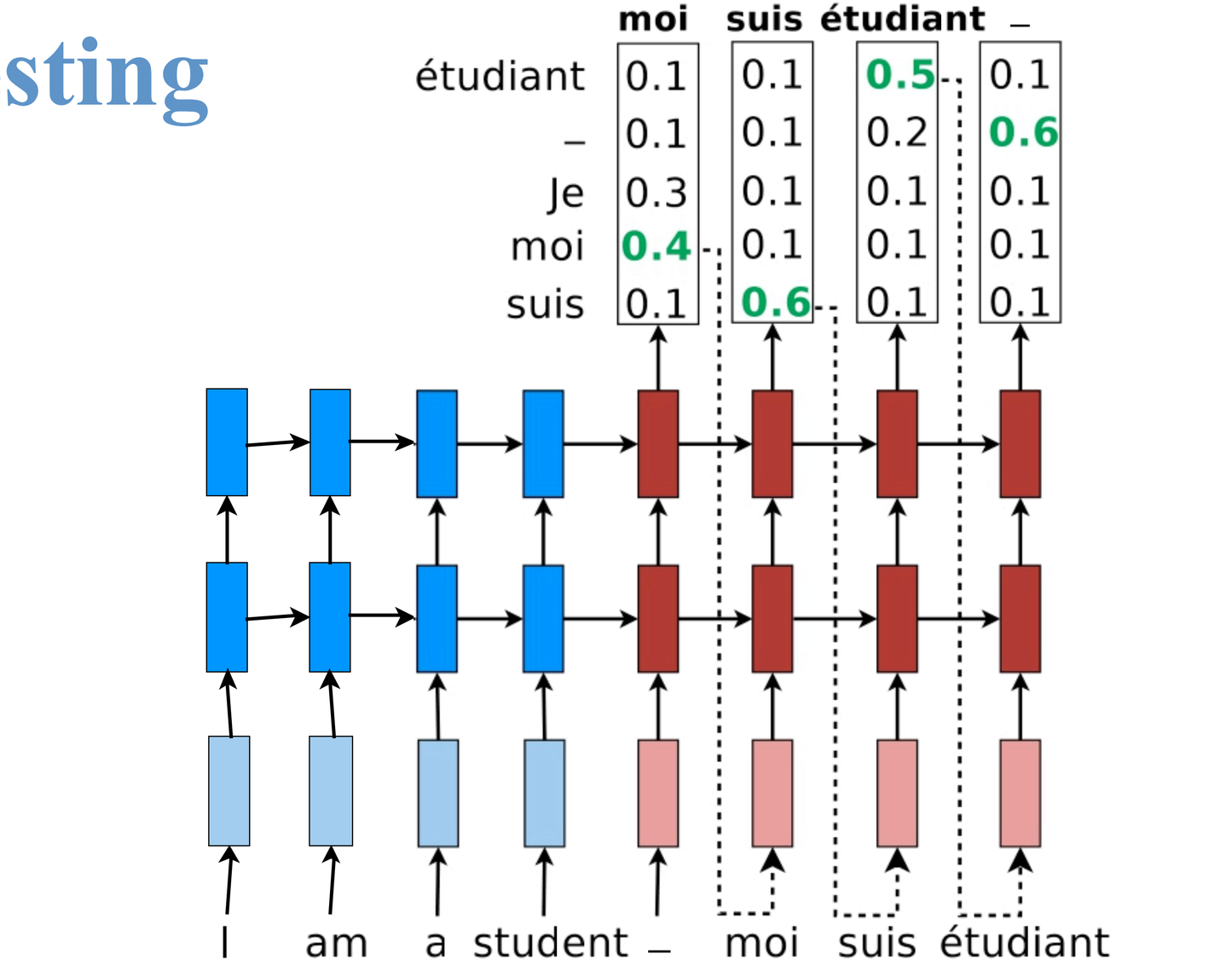
- Feed the most likely word

Testing



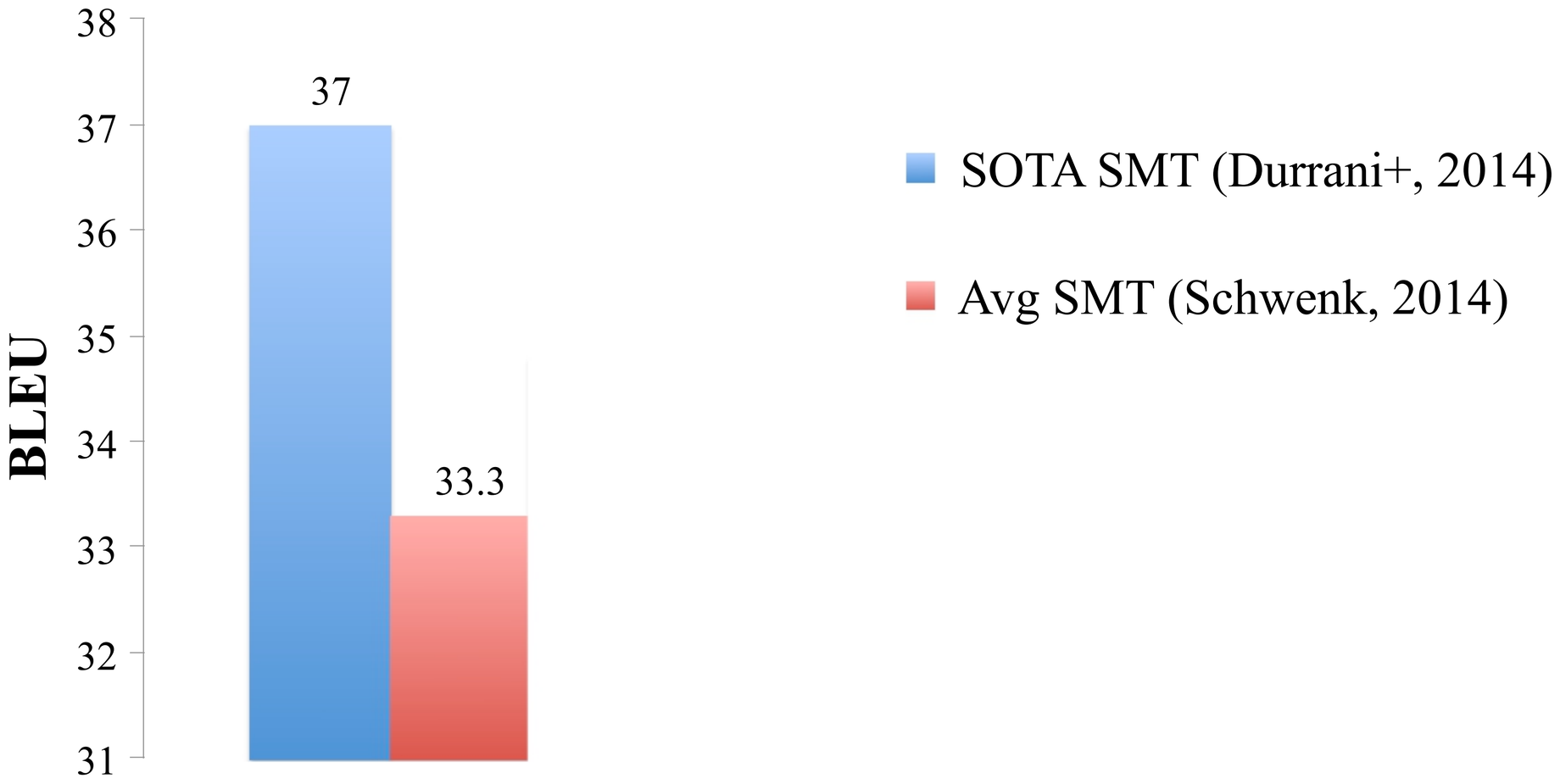
- Feed the most likely word

Testing



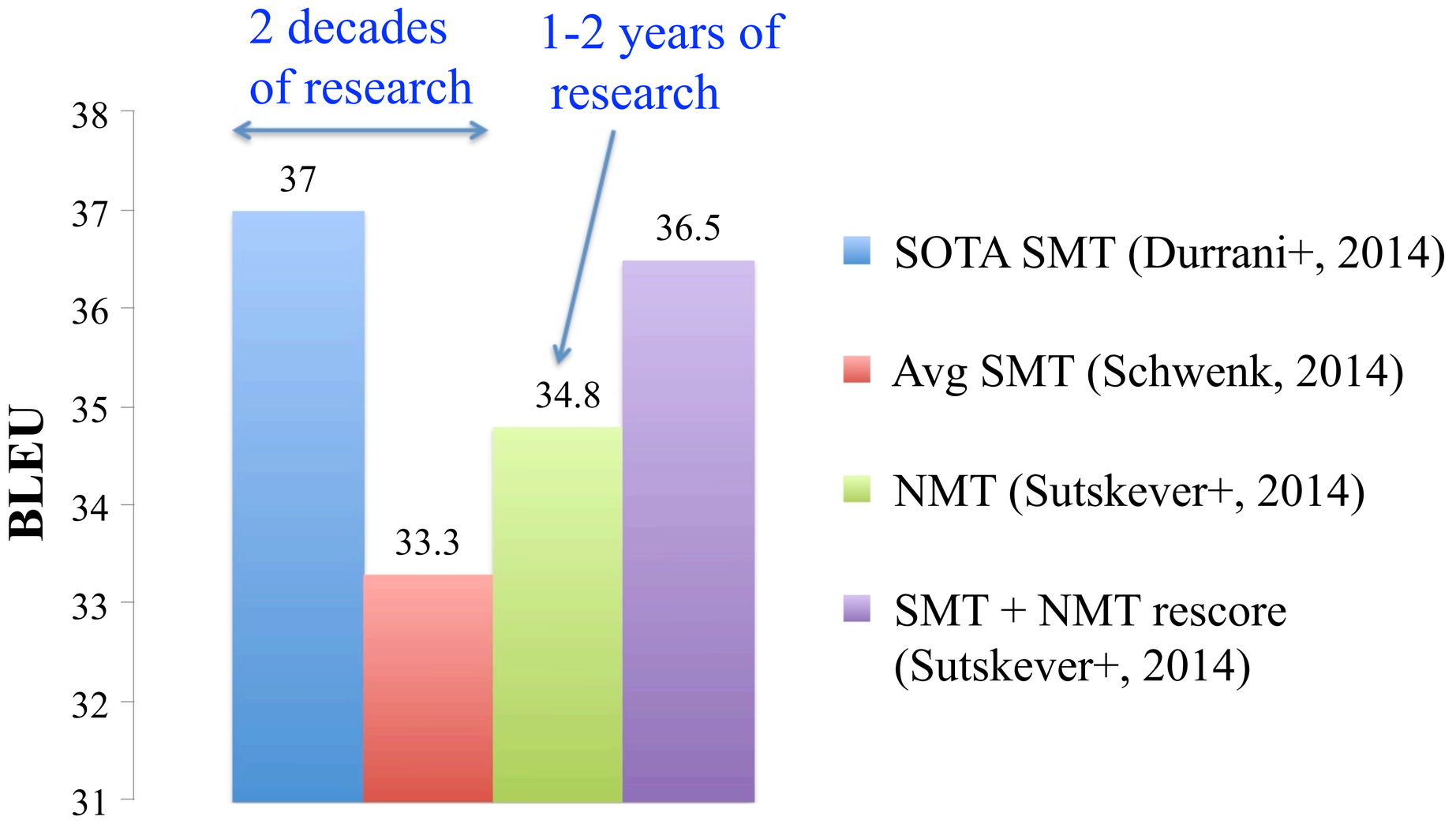
Simple beam-search decoders!

English-French WMT'14 results



Used with permission from Thong Luong

English-French WMT'14 results



Encoder-decoder Variants

	Encoder	Decoder
(Sutskever et al., 2014) seq2seq	Deep LSTM	Deep LSTM
(Cho et al., 2014) (Bahdanau et al., 2015) (Jean et al., 2015)	(Bidirectional) GRU	GRU
(Kalchbrenner & Blunsom, 2013)	CNN	(Inverse CNN) RNN

Next, advanced NMT!

Break time: when MT fails …

Sale of chicken murder

Go back toward your behind

Deep fried baby

Meat muscle stupid bean sprouts

Used with permission from Thong Luong

Limitations

- #1 : the ***vocabulary size*** problem
  - *Goal* : extend the vocabulary coverage.

- #2 : the ***sentence length*** problem
  - *Goal* : translate long sentences better.

- #3 : the ***language complexity*** problem
  - *Goal* : handle more language variations.