

Neural Machine Translation for Spoken Language Domains



Thang Luong

IWSLT 2015

(Joint work with Chris Manning)

Neural Machine Translation (NMT)

- **End-to-end** neural approach to MT:
 - Simple and coherent.
- Achieved **state-of-the-art** WMT results:
 - *English-French*: (Luong et al., 2015a).
 - *English-German*: (Jean et al., 2015a, Luong et al., '15b).
 - *English-Czech*: (Jean et al., 2015b).

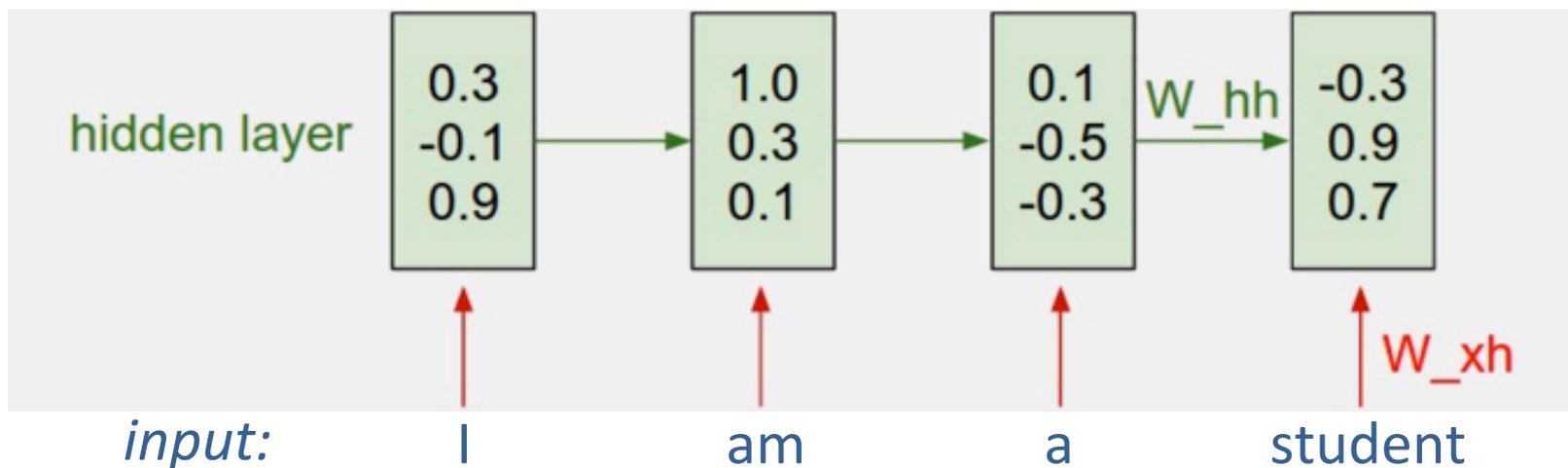
Not much work explores NMT for spoken language domains.

Outline

- A quick introduction to NMT.
 - Basics.
 - Attention mechanism.
- Our work in IWSLT.

We need to understand
Recurrent Neural Networks first!

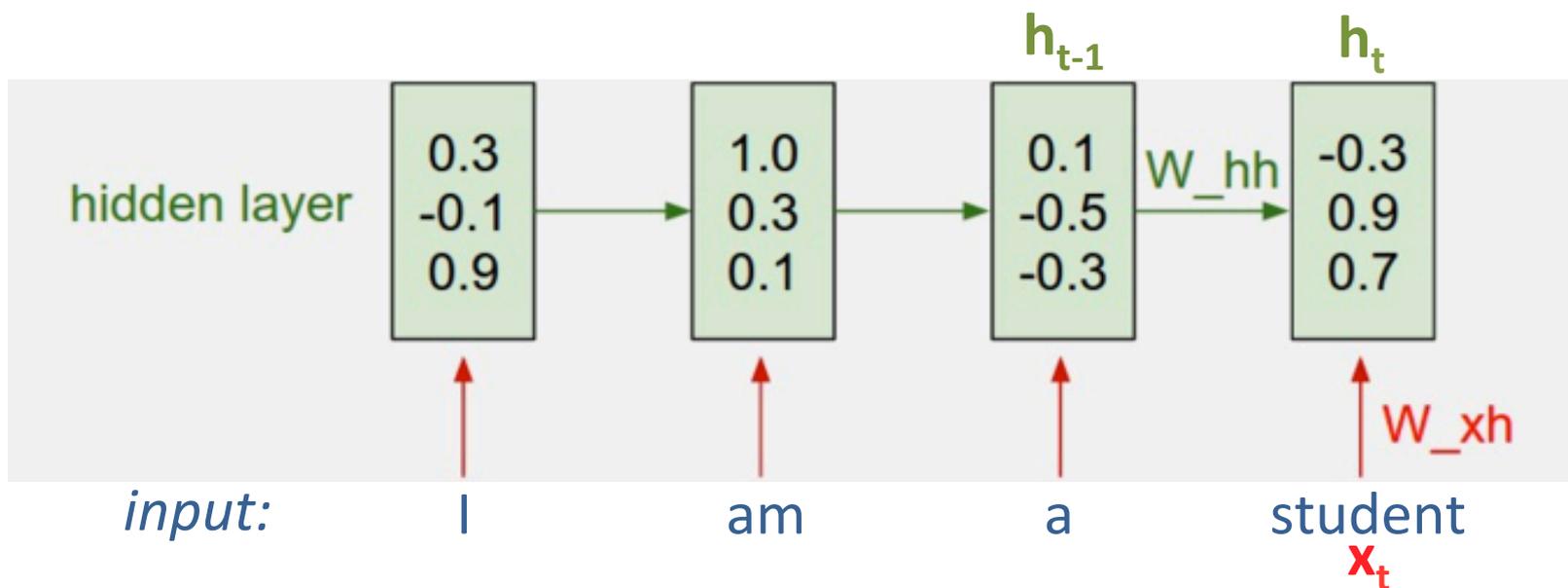
Recurrent Neural Networks (RNNs)



(Picture adapted from Andrej Karparthy)

Recurrent Neural Networks (RNNs)

$$h_t = \sigma (W_{xh}x_t + W_{hh}h_{t-1})$$



RNNs to represent sequences!

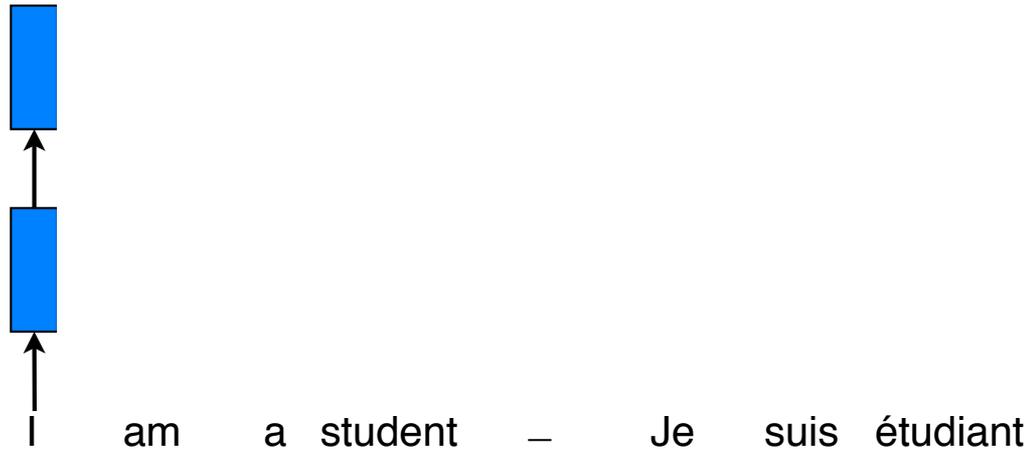
(Picture adapted from Andrej Karparthy)

Neural Machine Translation (NMT)

I am a student – Je suis étudiant

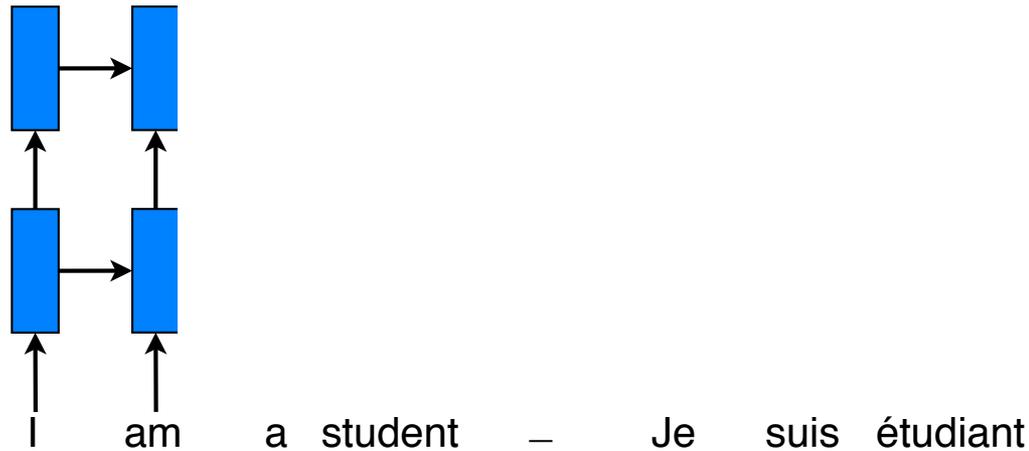
- Model $P(\text{target} \mid \text{source})$ directly.

Neural Machine Translation (NMT)



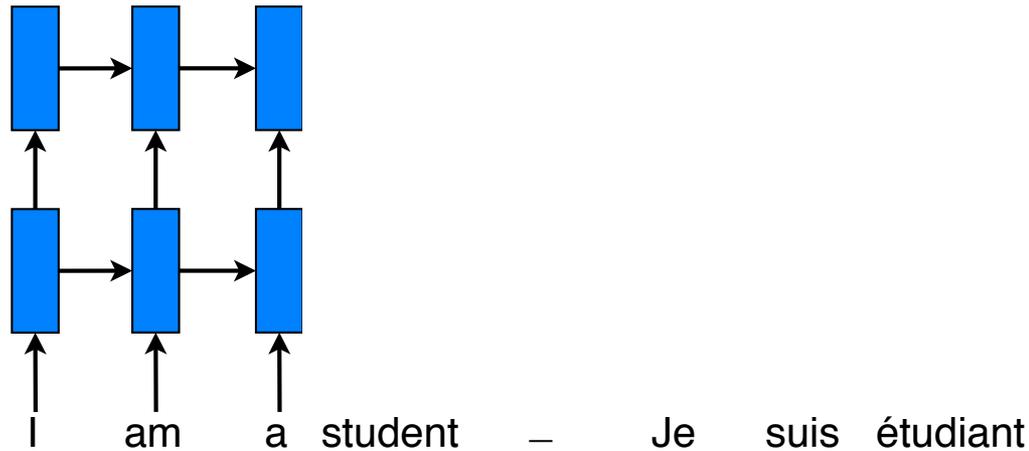
- RNNs trained **end-to-end** (Sutskever et al., 2014).

Neural Machine Translation (NMT)



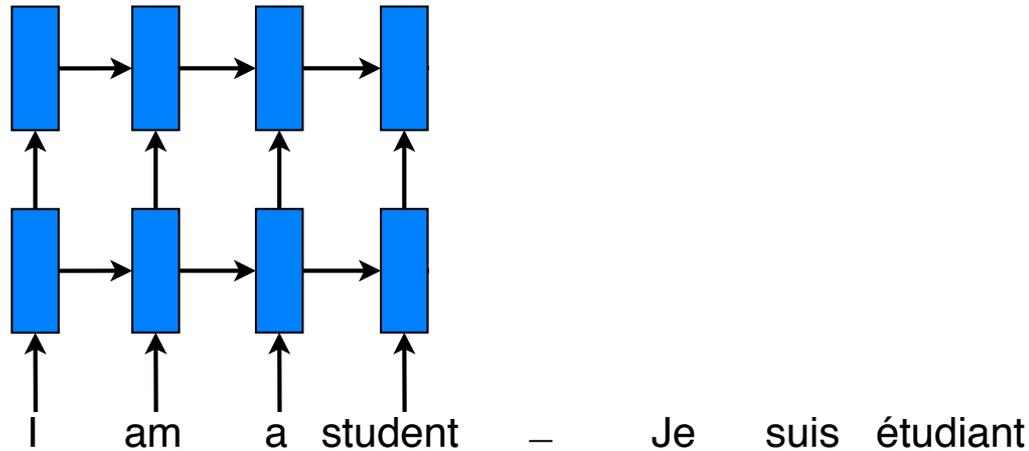
- RNNs trained **end-to-end** (Sutskever et al., 2014).

Neural Machine Translation (NMT)



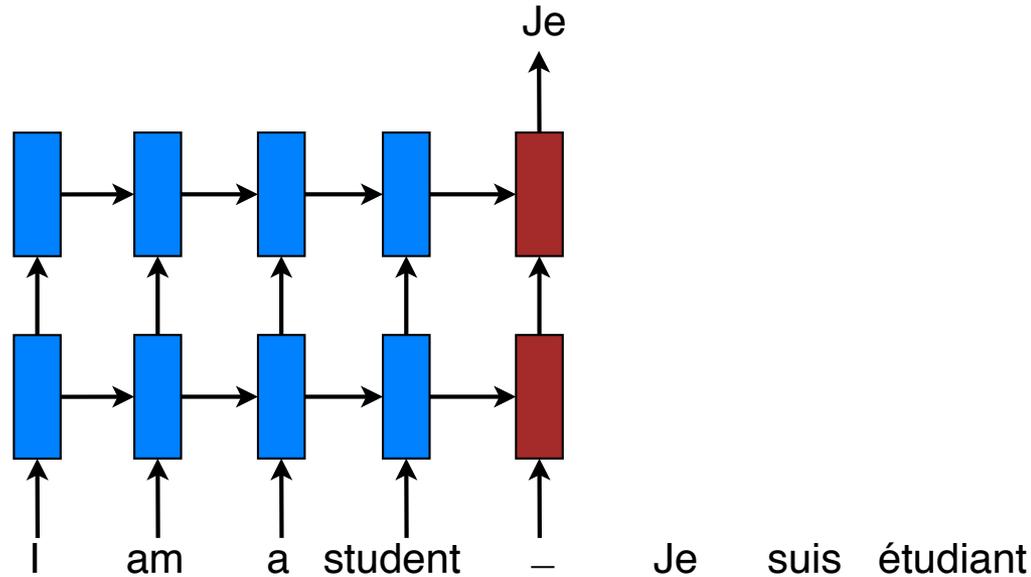
- RNNs trained **end-to-end** (Sutskever et al., 2014).

Neural Machine Translation (NMT)



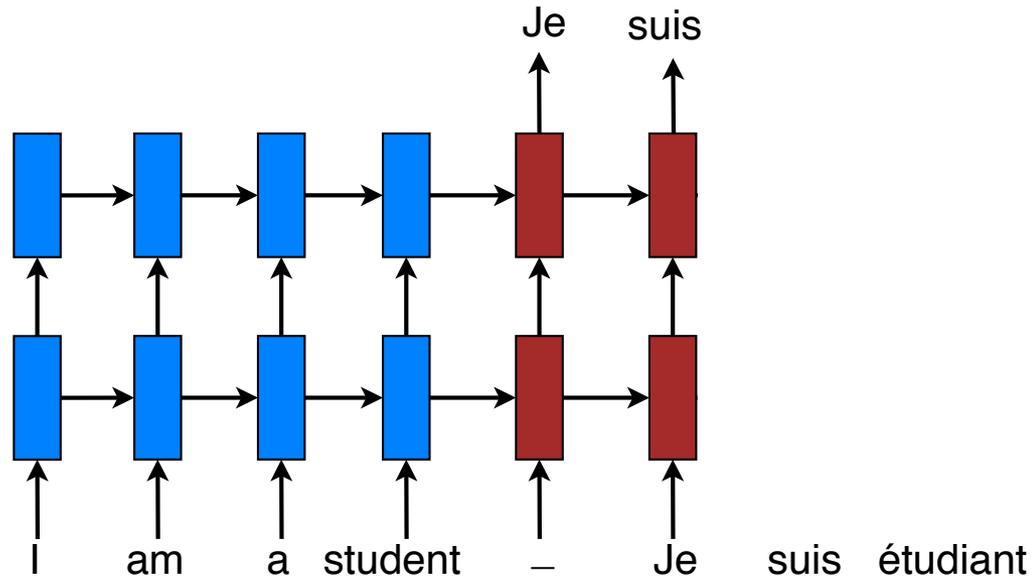
- RNNs trained **end-to-end** (Sutskever et al., 2014).

Neural Machine Translation (NMT)



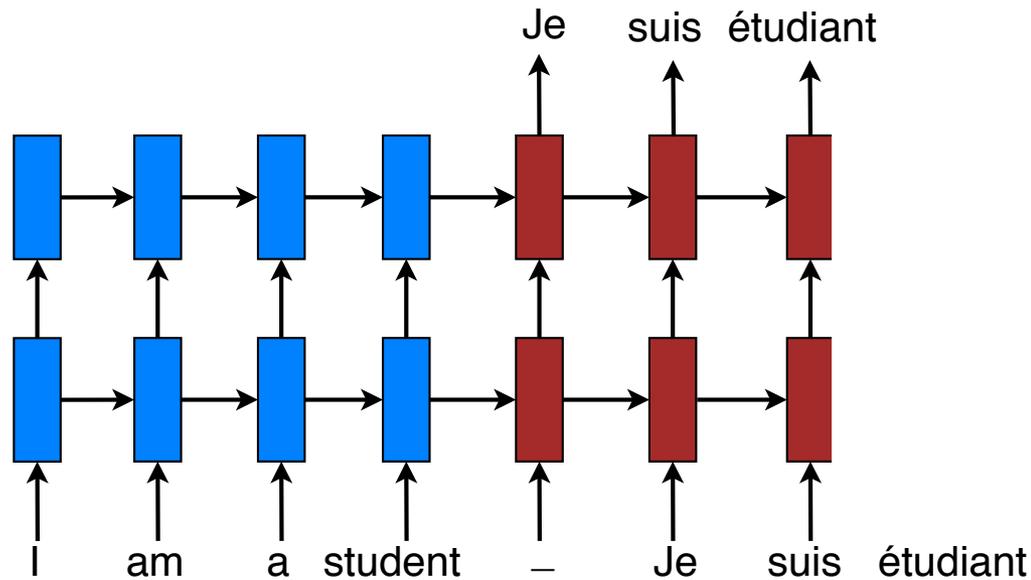
- RNNs trained **end-to-end** (Sutskever et al., 2014).

Neural Machine Translation (NMT)



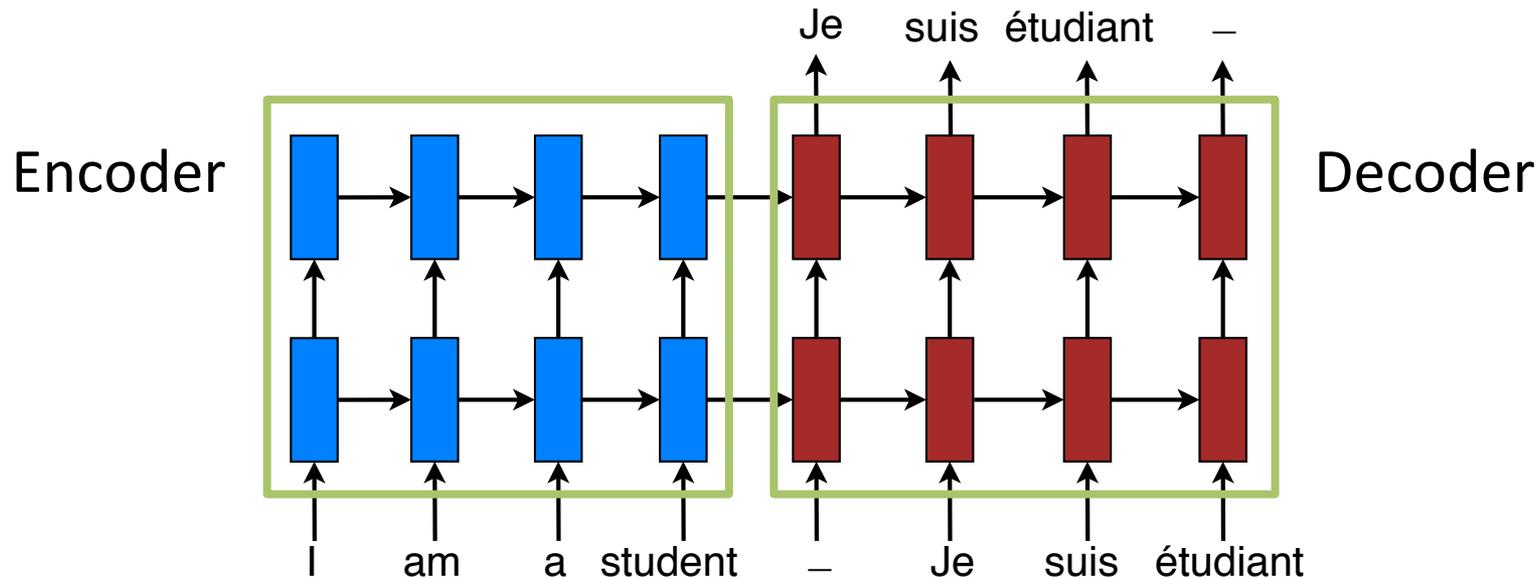
- RNNs trained **end-to-end** (Sutskever et al., 2014).

Neural Machine Translation (NMT)



- RNNs trained **end-to-end** (Sutskever et al., 2014).

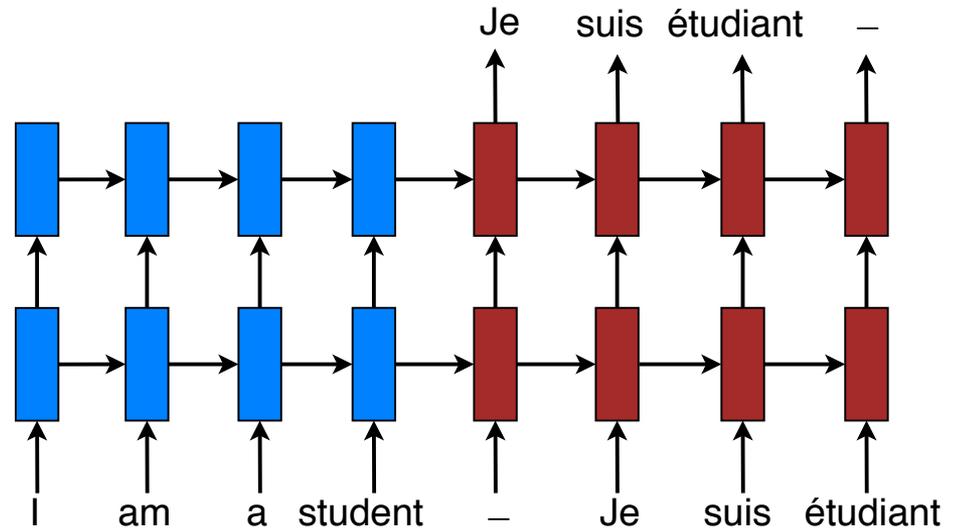
Neural Machine Translation (NMT)



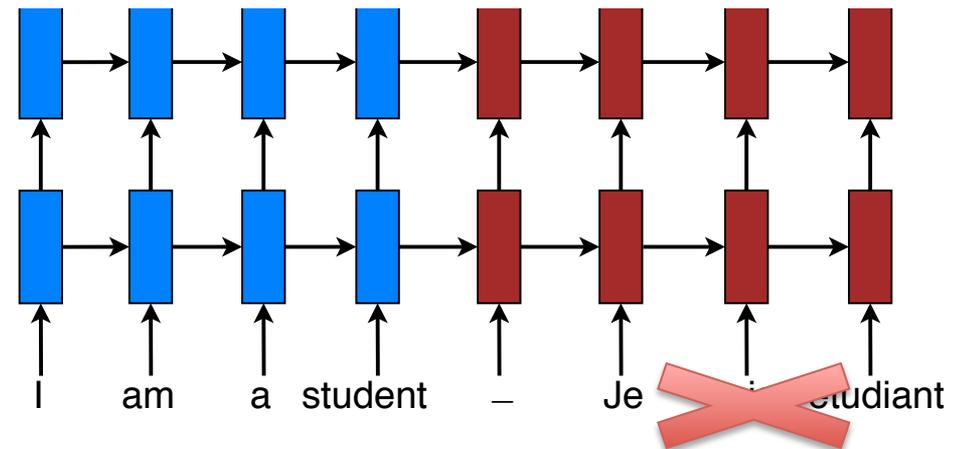
- RNNs trained **end-to-end** (Sutskever et al., 2014).
- **Encoder-decoder** approach.

Training vs. Testing

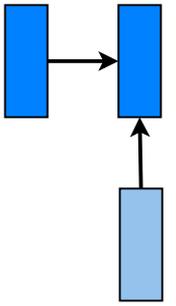
- *Training*
 - Correct translations are available.



- *Testing*
 - Only source sentences are given.



Recurrent types – vanilla RNN



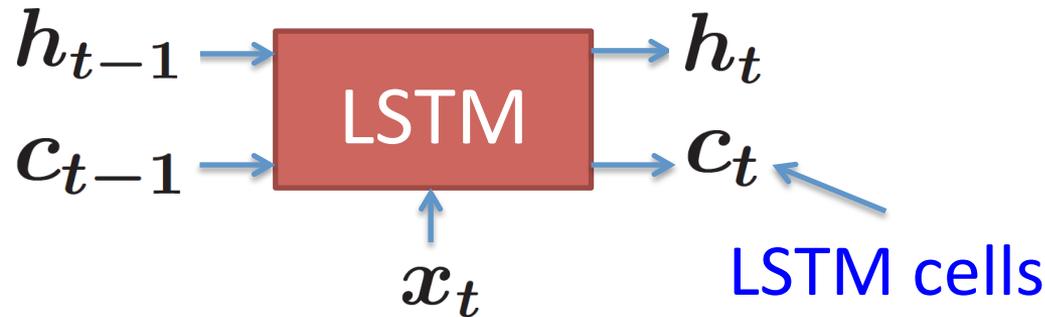
$$h_{t-1} \rightarrow \text{RNN} \rightarrow h_t = \sigma(\mathbf{W}_{xh}x_t + \mathbf{W}_{hh}h_{t-1})$$

x_t

Vanishing gradient problem!

Recurrent types – LSTM

C'mon, it's been around for 20 years!



- Long-Short Term Memory (LSTM)
 - (Hochreiter & Schmidhuber, 1997)
- LSTM cells are **additively** updated
 - Make backprop through time easier.

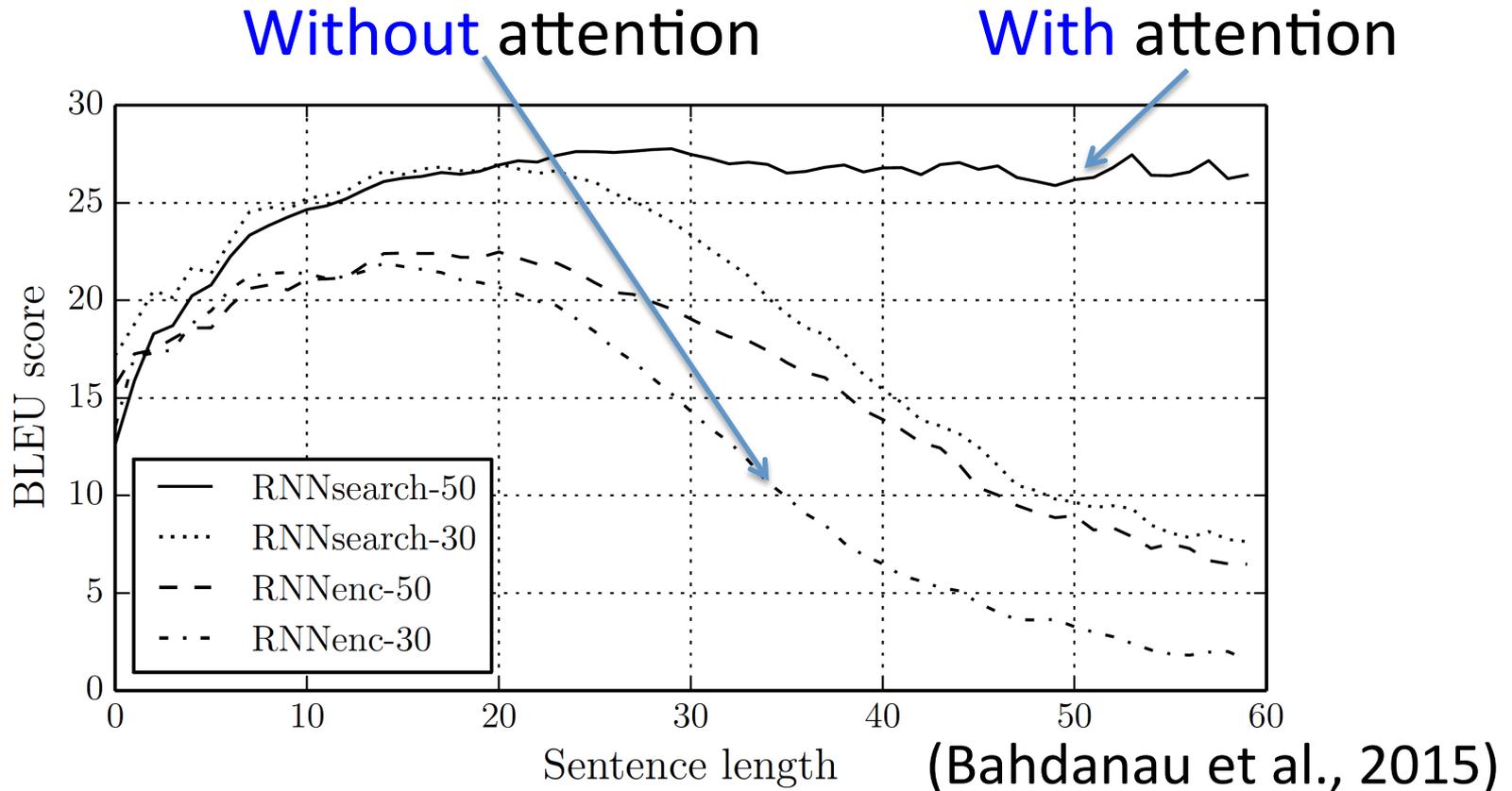
Summary – *NMT*

- Few linguistic assumptions.
- Simple beam-search decoders.
- Good generalization to long sequences.

Outline

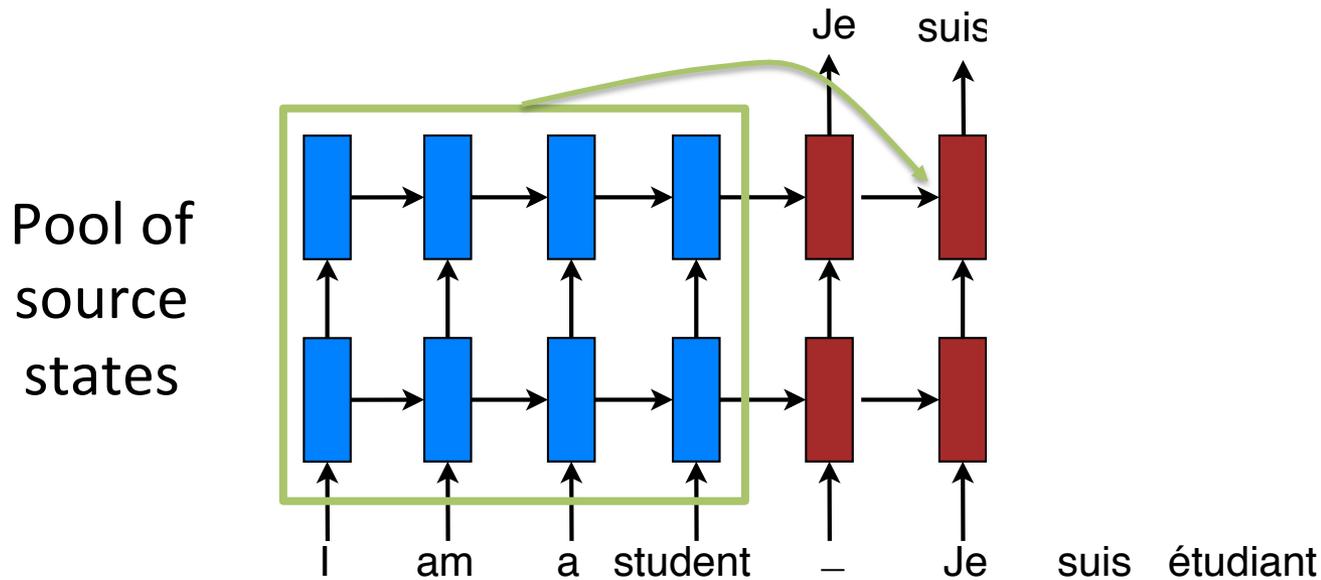
- A quick introduction to NMT.
 - Basics.
 - **Attention mechanism.**
- Our work in IWSLT.

Sentence Length Problem



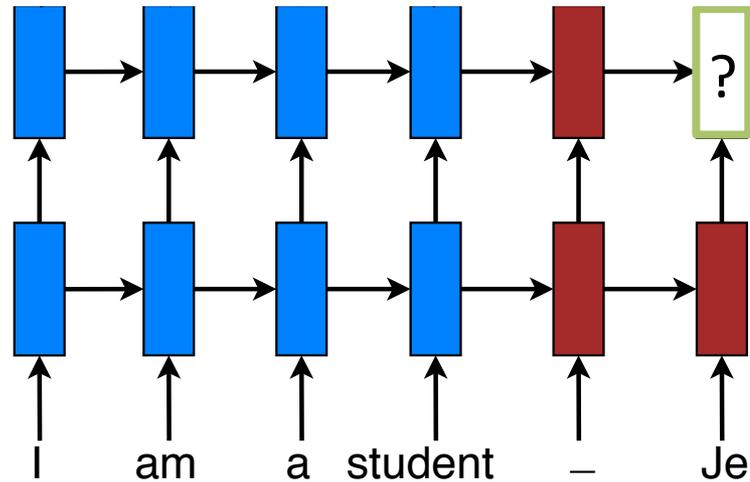
Problem: sentence meaning is represented by a fixed-dimensional vector.

Attention Mechanism



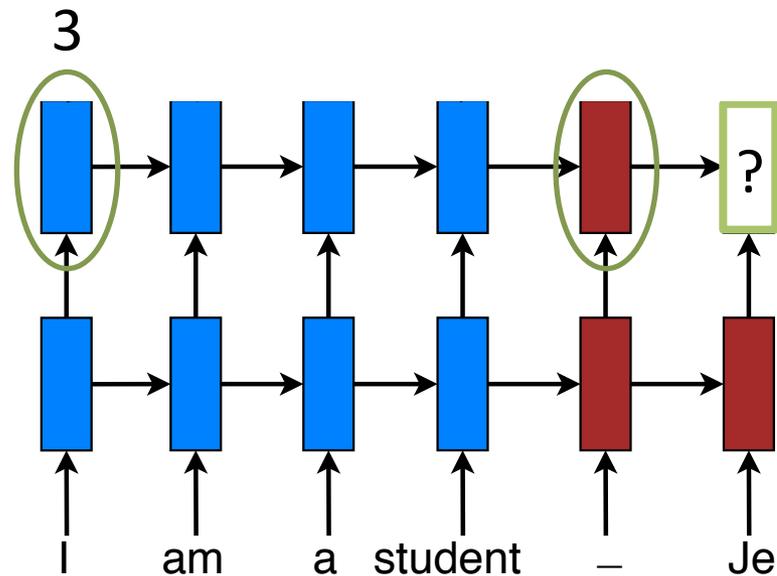
- **Solution:** random access memory
 - Retrieve as needed.

Attention Mechanism



Attention Mechanism – *Scoring*

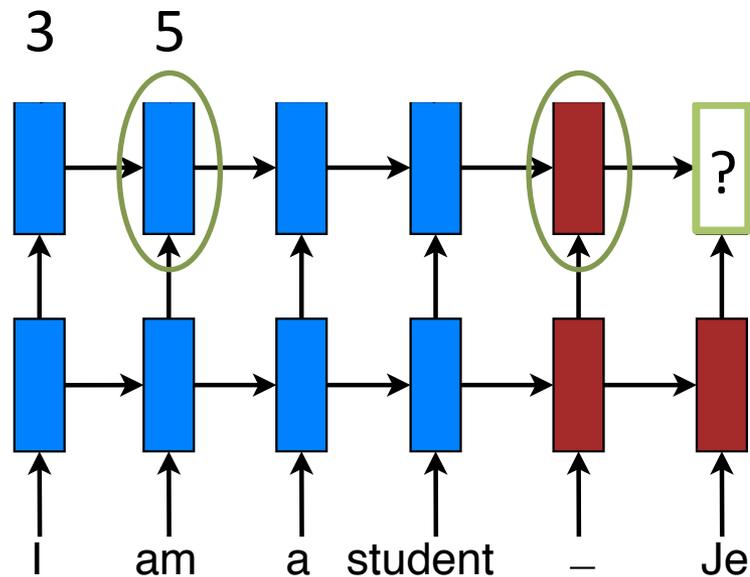
$$\text{score}(h_t, \bar{h}_s)$$



- Compare target and source hidden states.

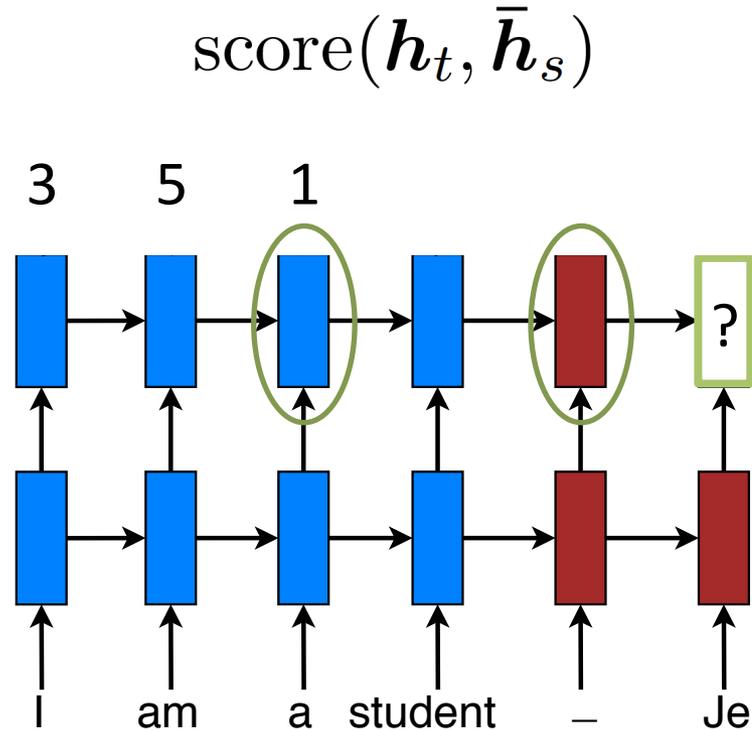
Attention Mechanism – *Scoring*

$$\text{score}(h_t, \bar{h}_s)$$



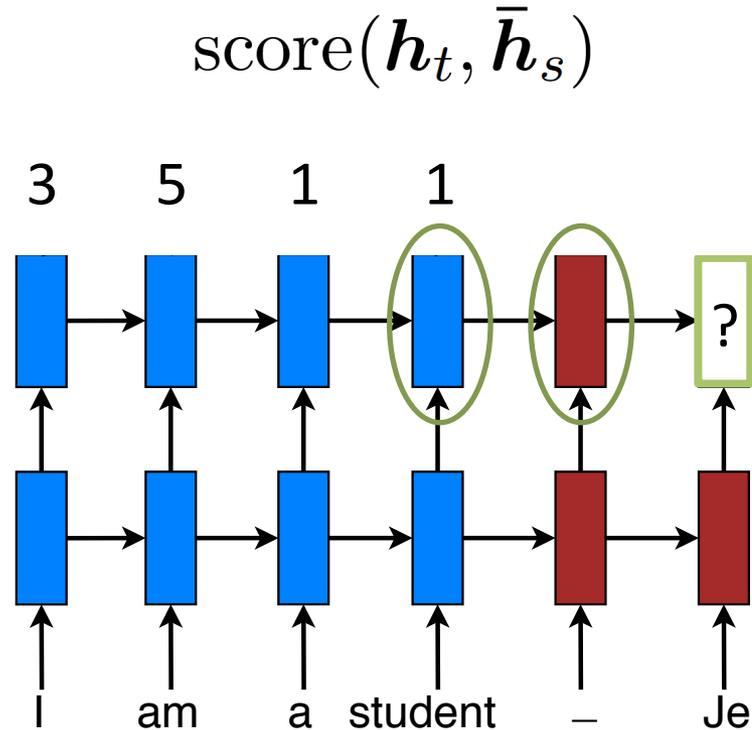
- Compare target and source hidden states.

Attention Mechanism – *Scoring*



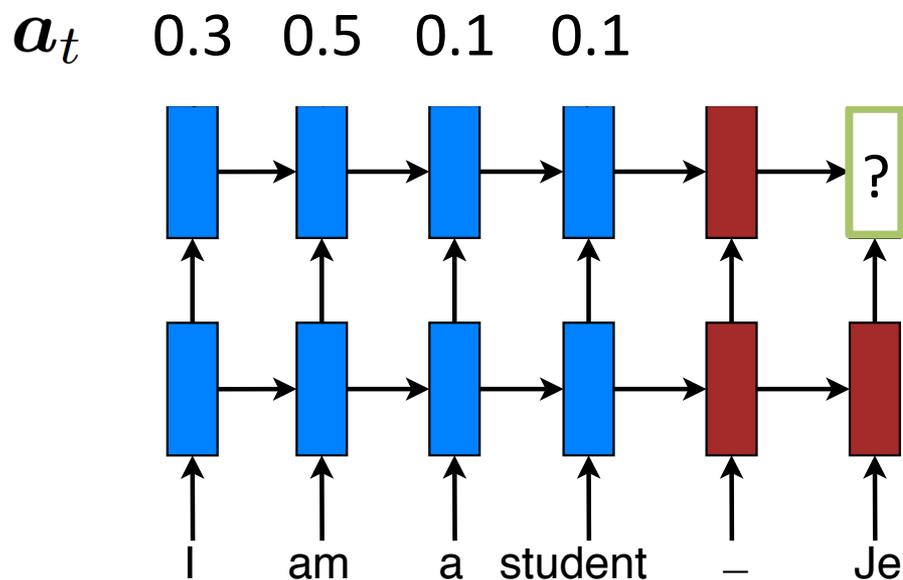
- Compare target and source hidden states.

Attention Mechanism – *Scoring*



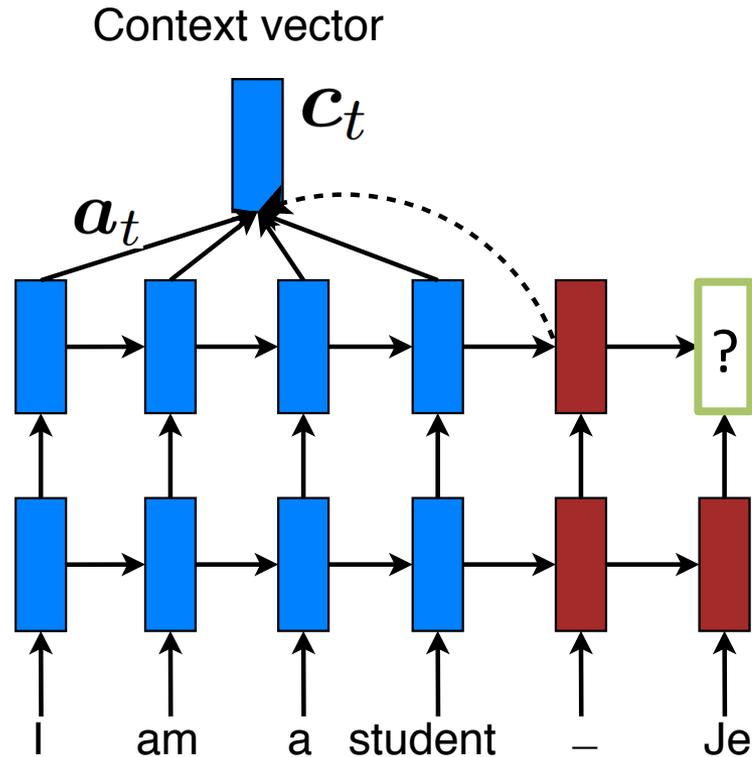
- Compare target and source hidden states.

Attention Mechanism – *Normalization*



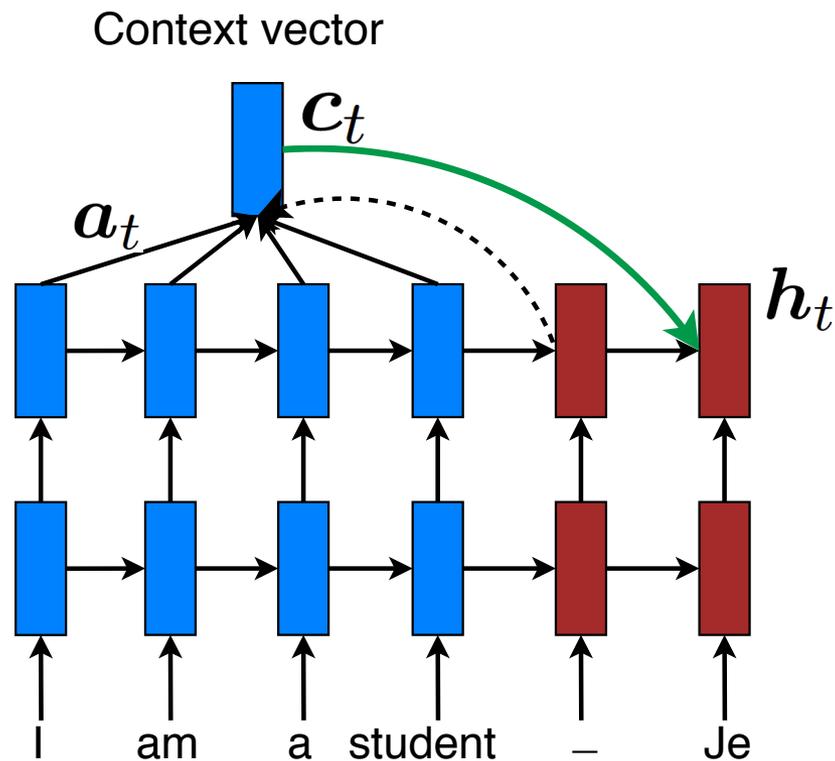
- Convert into **alignment weights**.

Attention Mechanism – *Context vector*



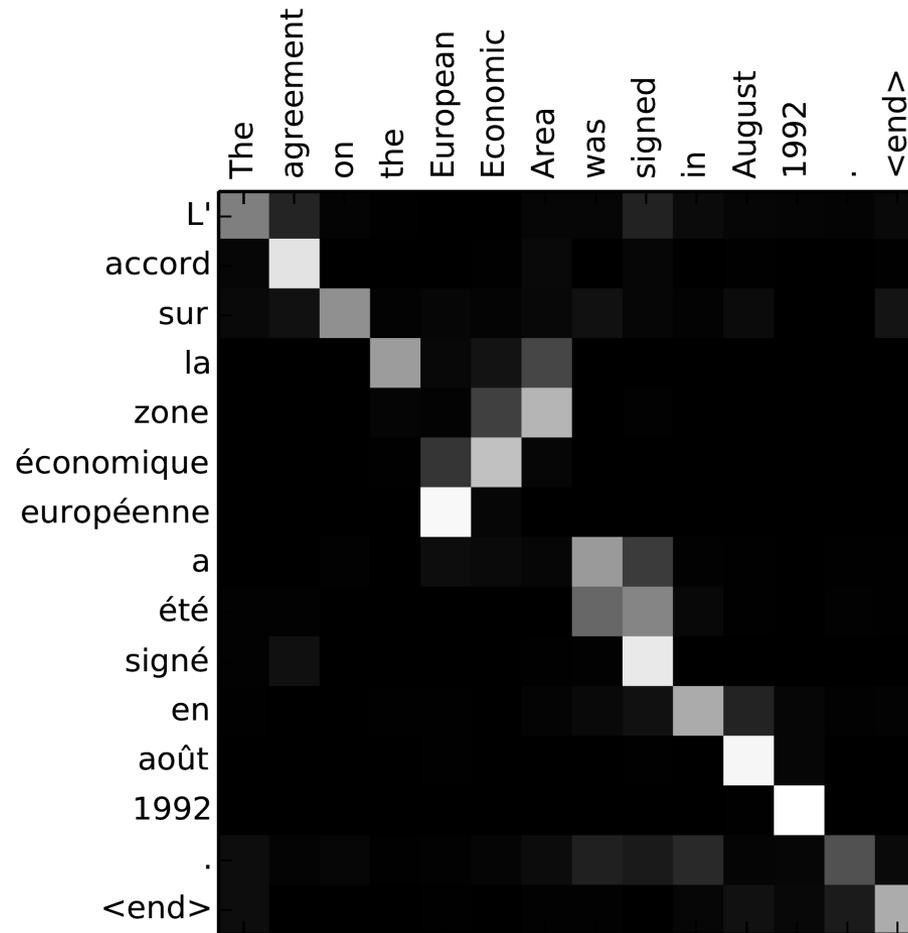
- Build **context** vector: weighted average.

Attention Mechanism – *Hidden state*



- Compute the **next hidden state**.

Alignments as a by-product



(Bahdanau et al., 2015)

Summary – *Attention*

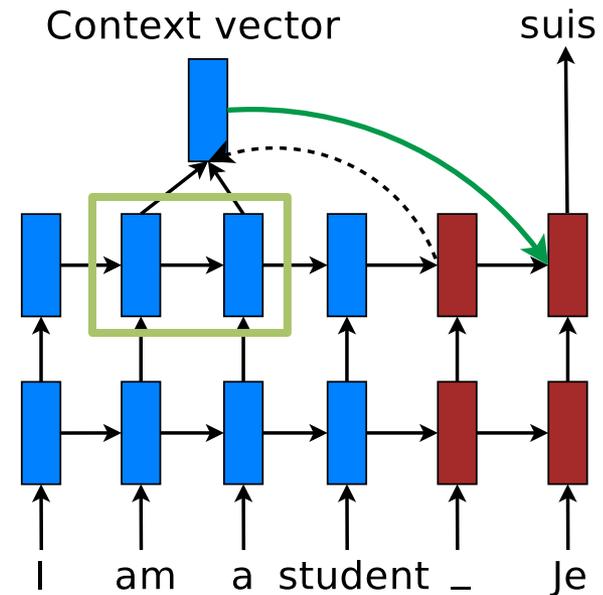
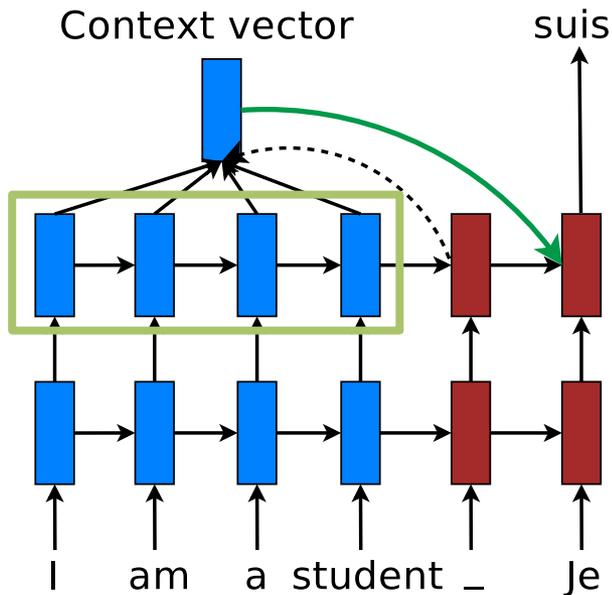
- A random access memory.
 - Help translate long sentences.
- Produce alignments.

Outline

- A quick introduction to NMT.
- Our work in IWSLT:
 - Models.
 - NMT adaptation.
 - NMT for low-resource translation.

Models

- Attention-based models (Luong et al., 2015b):
 - Global & local attention.

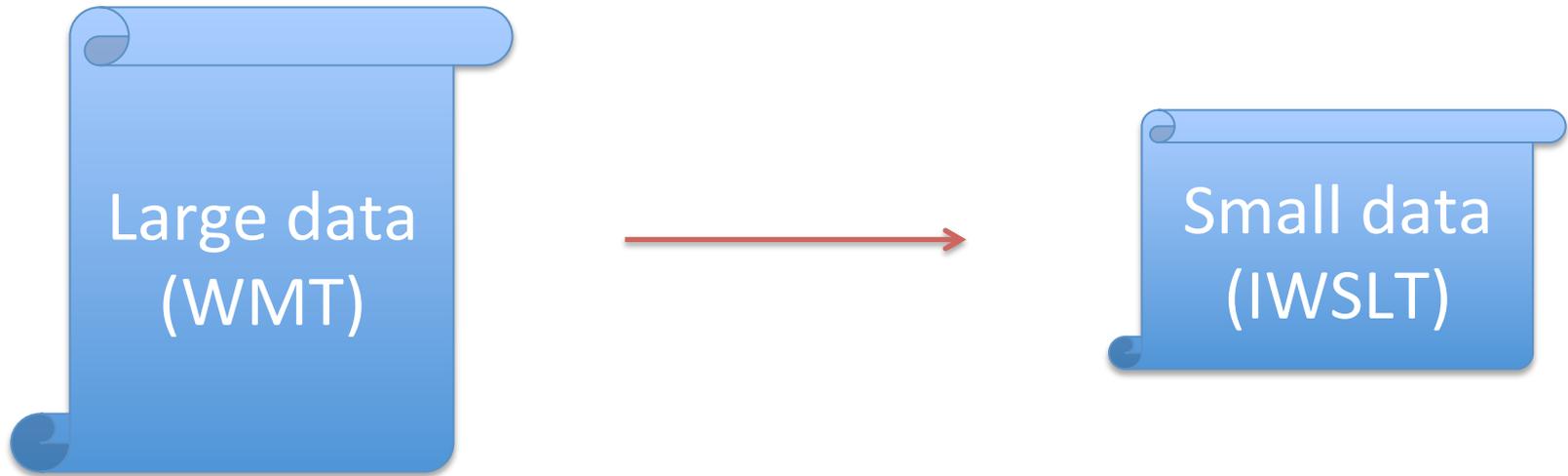


Global: **all** source states.

Local: **subset** of source states.

Train both types of models to ensemble later.

NMT Adaptation



Can we adapt existing models?

Existing models

- **State-of-the-art English** \mapsto **German** NMT system
 - Trained on WMT data (4.5M sent pairs)
 - Tesla K40, 7-10 days.
- **Ensemble** of 8 models (Luong et al., 2015b).
 - Global / local attention +/- dropout.
 - Source reversing.
 - 4 LSTM layers, 1000 dimensions.
 - 50K top frequent words.

Adaptation

- Further train on IWSLT data.
 - 200K sentence pairs.
 - 12 epochs with SGD: 3-5 hours on GPU.
- Same settings:
 - Source reversing.
 - 4 LSTM layers, 1000 dimensions.
- Same vocab:
 - 50K top frequent words.
 - Would be useful to update vocab!

Results – *TED tst2013*

Systems	BLEU
IWSLT'14 best entry (Freitag et al., 2014)	26.2
<i>Our NMT systems</i>	
Single NMT (unadapted)	25.6

Results – *TED tst2013*

Systems	BLEU
IWSLT'14 best entry (Freitag et al., 2014)	26.2
<i>Our NMT systems</i>	
Single NMT (unadapted)	25.6
Single NMT (adapted)	29.4 (+3.8)



Adaptation is effective.

Results – *TED tst2013*

Systems	BLEU
IWSLT'14 best entry (Freitag et al., 2014)	26.2
<i>Our NMT systems</i>	
Single NMT (unadapted)	25.6
Single NMT (adapted)	29.4 (+3.8)
Ensemble NMT (adapted)	31.4 (+2.0)



English \mapsto German Evaluation Results

Systems	tst2014	tst2015
IWSLT'14 best entry (Freitag et al., 2014)	23.3	
IWSLT'15 baseline	18.5	20.1

English \mapsto German Evaluation Results

Systems	tst2014	tst2015
IWSLT'14 best entry (Freitag et al., 2014)	23.3	
IWSLT'15 baseline	18.5	20.1
Our NMT ensemble	27.6 (+9.1)	30.1 (+10.0)



NMT generalizes well!

Sample English-German translations

src

We desperately need **great communication** from our scientists and engineers in order to change the world.

ref

Wir brauchen unbedingt **großartige Kommunikation** von unseren Wissenschaftlern und Ingenieuren, um die Welt zu verändern.

unadapt

Wir benötigen dringend eine **große Mitteilung** unserer Wissenschaftler und Ingenieure, um die Welt zu verändern.

adapted

Wir brauchen dringend eine **großartige Kommunikation** unserer Wissenschaftler und Ingenieure, um die Welt zu verändern.

- **Adapted** models are better.

Sample English-German translations

src

We desperately need great communication **from our scientists** and engineers in order to change the world.

ref

Wir brauchen unbedingt großartige Kommunikation **von unseren Wissenschaftlern** und Ingenieuren, um die Welt zu verändern.

unadapt

Wir benötigen dringend eine große Mitteilung **unserer Wissenschaftler** und Ingenieure, um die Welt zu verändern.

adapted

Wir brauchen dringend eine großartige Kommunikation **unserer Wissenschaftler** und Ingenieure, um die Welt zu verändern.

best

Wir brauchen dringend eine großartige Kommunikation **von unseren Wissenschaftlern** und Ingenieuren, um die Welt zu verändern.

- **Ensemble** models are best.
 - Correctly translate the plural noun “**scientists**”.

Sample English-German translations

src	Yeah. Yeah. So what will happen is that, during the call you have to indicate whether or not you have the disease or not, you see. Right.
ref	Was passiert ist, dass der Patient während des Anrufes angeben muss, ob diese Person an Parkinson leidet oder nicht. Ok.
base	Ja dass dass dass.
adapted	Ja. Ja. Es wird also passieren, dass man während des Gesprächs angeben muss, ob man krank ist oder nicht. Richtig.
best	Ja. Ja. Was passiert, ist, dass Sie während des zu angeben müssen, ob Sie die Krankheit haben oder nicht, oder nicht. Richtig.

- **Unadapted** models screwed up.

Sample English-German translations

src	Yeah. Yeah. So what will happen is that, during the call you have to indicate whether or not you have the disease or not, you see. Right.
ref	Was passiert ist, dass der Patient während des Anrufes angeben muss, ob diese Person an Parkinson leidet oder nicht. Ok.
base	Ja dass dass dass.
adapted	Ja. Ja. Es wird also passieren, dass man während des Gesprächs angeben muss, ob man krank ist oder nicht. Richtig.
best	Ja. Ja. Was passiert, ist, dass Sie während des zu angeben müssen, ob Sie die Krankheit haben oder nicht, oder nicht. Richtig.

- **Adapted** models produce more reliable translations.

Outline

- A quick introduction to NMT.
- Our work in IWSLT:
 - Models.
 - NMT adaptation.
 - **NMT for low-resource translation.**

NMT for low-resource translation

- So far, NMT systems have been trained on **large WMT data**:
 - *English-French*: 12-36M sentence pairs.
 - *English-German*: 4.5M sentence pairs.
- Not much work utilizes **small corpora**:
 - (Gülçehre et al., 2015): IWSLT Turkish \mapsto English, but use large English monolingual data.

Train English \mapsto Vietnamese systems

Setup

- Train English \mapsto Vietnamese models **from scratch**:
 - 133K sentence pairs: Moses tokenizer, true case.
 - Words occur at least 5 times.
 - 17K English words & 7.7K Vietnamese words.
- Use **smaller networks**:
 - 2 LSTM layers, 500 dimensions.
 - Tesla K40, 4-7 hours on GPU.
- **Ensemble** of 9 models:
 - Global / local attention +/- dropout.

English \mapsto Vietnamese Results – *BLEU*

Systems	tst2013
Single NMT	23.3
Ensemble NMT	26.9

Systems	tst2015
IWSLT'15 baseline	27.0
Our system	26.4

Results are competitive.

Latest Results – tst2015!

TED : MT English-Vietnamese (MT_{EnVi})

System	<i>case sensitive</i>		
	BLEU	NIST	TER
PJAIT	28.39	6.6650	56.01
JAIST	28.17	6.7092	55.84
KIT	26.60	6.4014	58.26
SU	26.41	6.5986	55.60
UNETI	22.93	6.0218	60.33
BASELINE	27.01	6.4716	58.42

We score
top in TER!

- Observation by (Neubig et al., 2015):
 - NMT is good at getting the syntax right, not much about lexical choices.

Sample English-Vietnamese translations

src	However, Gaddafi left behind a heavy burden, a legacy of tyranny , corruption and seeds of diversions.
ref	Tuy nhiên, Gaddafi đã để lại một gánh nặng, một di sản của chính thể chuyên chế , tham nhũng và những mầm mống chia rẽ.
single	Tuy nhiên, Gaddafi đằng sau gánh nặng nặng nề, một di sản di sản , tham nhũng và hạt giống.
multi	Tuy nhiên, Gaddafi bỏ lại sau một gánh nặng nặng nề, một di sản của sự chuyên chế , tham nhũng và hạt giống.

- **Ensemble** models are better.

Sample English-Vietnamese translations

src	From 1971 to 1977 - I look young, but I'm not - I worked in Zambia, Kenya, Ivory Coast, Algeria, Somalia, in projects of technical cooperation with African countries.
ref	Từ năm 1971 đến 1977- Trông tôi trẻ thế chứ không phải vậy đâu - Tôi đã làm việc tại Zambia, Kenya, Ivory Coast , Algeria, Somalia, trong những dự án hợp tác về kỹ thuật với những quốc gia Châu Phi
single	Từ 1971 đến năm 1977, tôi còn trẻ, nhưng tôi không - tôi làm việc ở Zambia, Kenya, Bờ Biển Ngà, Somalia, Somalia, trong các dự án về kỹ thuật
multi	Từ 1971 đến năm 1977 - tôi trông trẻ, nhưng tôi không - Tôi làm việc ở Zambia, Kenya, Bờ Biển Ngà, Algeria, Somalia, trong các dự án hợp tác với các nước châu Phi.

- **Ensemble** models are better.

Sample English-Vietnamese translations

src | From 1971 to 1977 -- I look young, but I'm not -- -- I worked in Zambia, Kenya, **Ivory Coast**, Algeria, Somalia, in projects of technical cooperation with African countries.

ref | Từ năm 1971 đến 1977-- Trông tôi trẻ thế chứ không phải vậy đâu ---- Tôi đã làm việc tại Zambia, Kenya, **Ivory Coast** , Algeria, Somalia, trong những dự án hợp tác về kỹ thuật với những quốc gia Châu Phi

single | Từ 1971 đến năm 1977, tôi còn trẻ, nhưng tôi không -- -- tôi làm việc ở Zambia, Kenya, **Bờ Biển Ngà**, Somalia, Somalia, trong các dự án về kỹ thuật

multi | Từ 1971 đến năm 1977 -- tôi trông trẻ, nhưng tôi không -- Tôi làm việc ở Zambia, Kenya, **Bờ Biển Ngà**, Algeria, Somalia, trong các dự án hợp tác với các nước châu Phi.

- **Sensible name** translations.



Thank you!

Conclusion

- NMT in spoken language domains:
 - Domain adaptation
 - Low-resource translation.
- Domain adaptation is useful
 - New SOTA results for English-German.
- Low-resource translation
 - Competitive results for English-Vietnamese.