Deep Learning

Lecture 20

Neural Machine Translation

Part 2

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

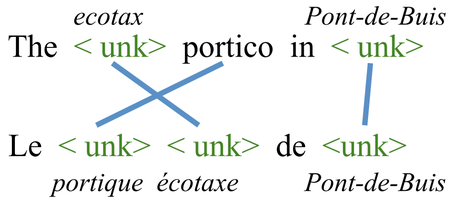
- #1 : the ***vocabulary size*** problem
  - *Goal* : extend the vocabulary coverage.

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

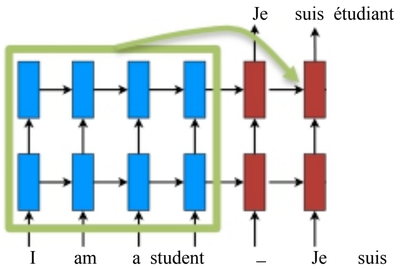
- #3 : the ***language complexity*** problem
  - *Goal* : handle more language variations.

Advanced NMT

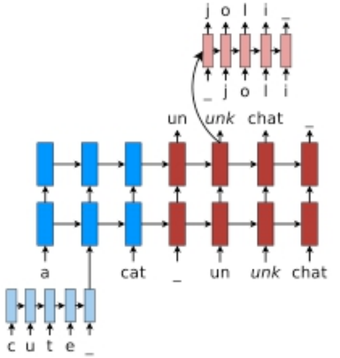
- #1 : the ***vocabulary size*** problem
  - *Sol* : "copy" mechanism.



- #2 : the ***sentence length*** problem
  - *Sol* : attention mechanism.



- #3 : the ***language complexity*** problem
  - *Sol* : character-level translation.



Computer Vision vs. NLP

1K categories

cat

Computer Vision

1M categories

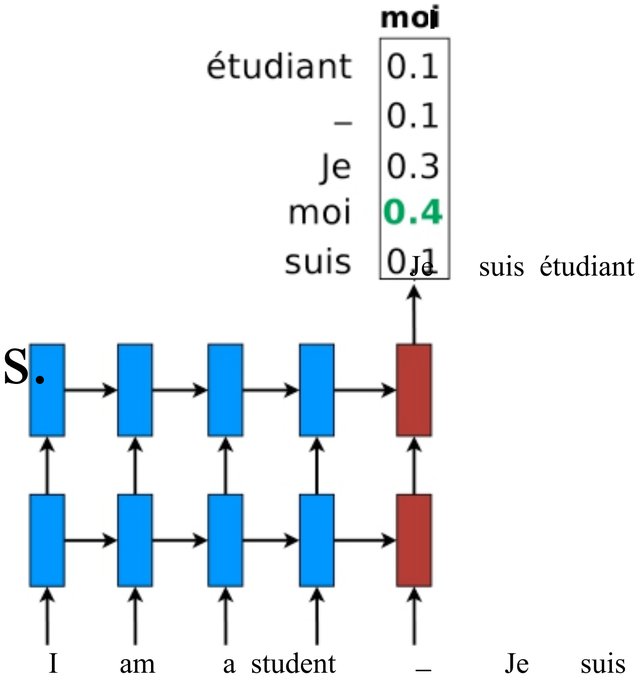
mat

NLP

The cat sat on a

#1 The Vocabulary Size Problem

- Word generation problem
  - Vocabs are modest: 50K.
  - Simple softmax : GPU friendliness.



The ecotax portico in Pont-de-Buis
Le portique écotaxe de Pont-de-Buis

The < unk> portico in <unk>
Le < unk> <unk> de <unk>

Google

- Propose “copy” mechanisms for < *unk*> .

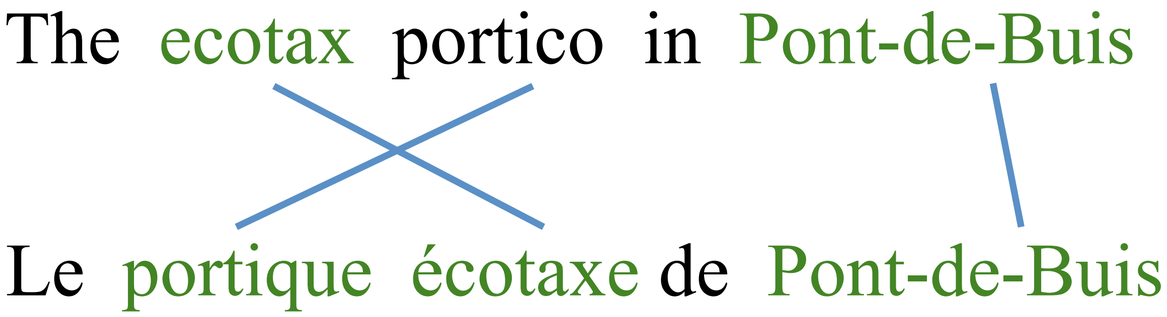
- Simple & effective
  - Treat any NMT as a black box.
  - Annotate training data.
  - Post-process translations.

SOTA for English-French translation.

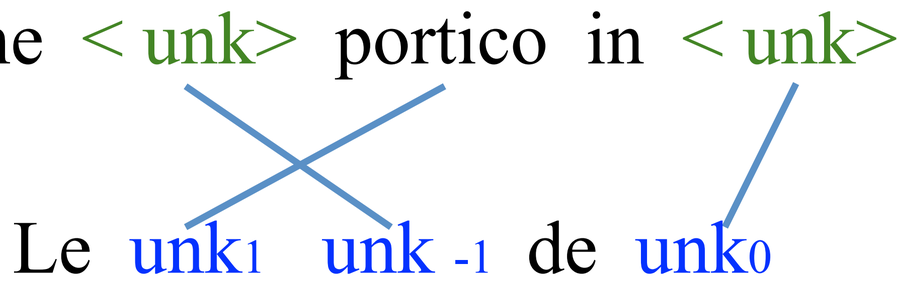
Thang Luong, Ilya Sutskever *, Quoc Le*, Oriol Vinyals, and Wojciech Zaremba.*
***Addressing the Rare Word Problem in Neural Machine Translation** . ACL 2015.*

Training annotation

- Learn alignments.



- Add relative positions.



"Attention" for rare words

Post-process

Test sentence

ecotax *Pont-de-Buis*

The < unk> portico in < unk>

Translation

Le portique unk -1 de unk0

Post-process

Test sentence: The *ecotax* < unk> portico in *Pont-de-Buis* < unk>

Translation: Le portique unk -1 de unk0

Dictionary translation

Post-edit Translation: Le portique écotaxe de Pont-de-Buis

Post-process

ecotax *Pont-de-Buis*

Test sentence The < unk> portico in < unk>

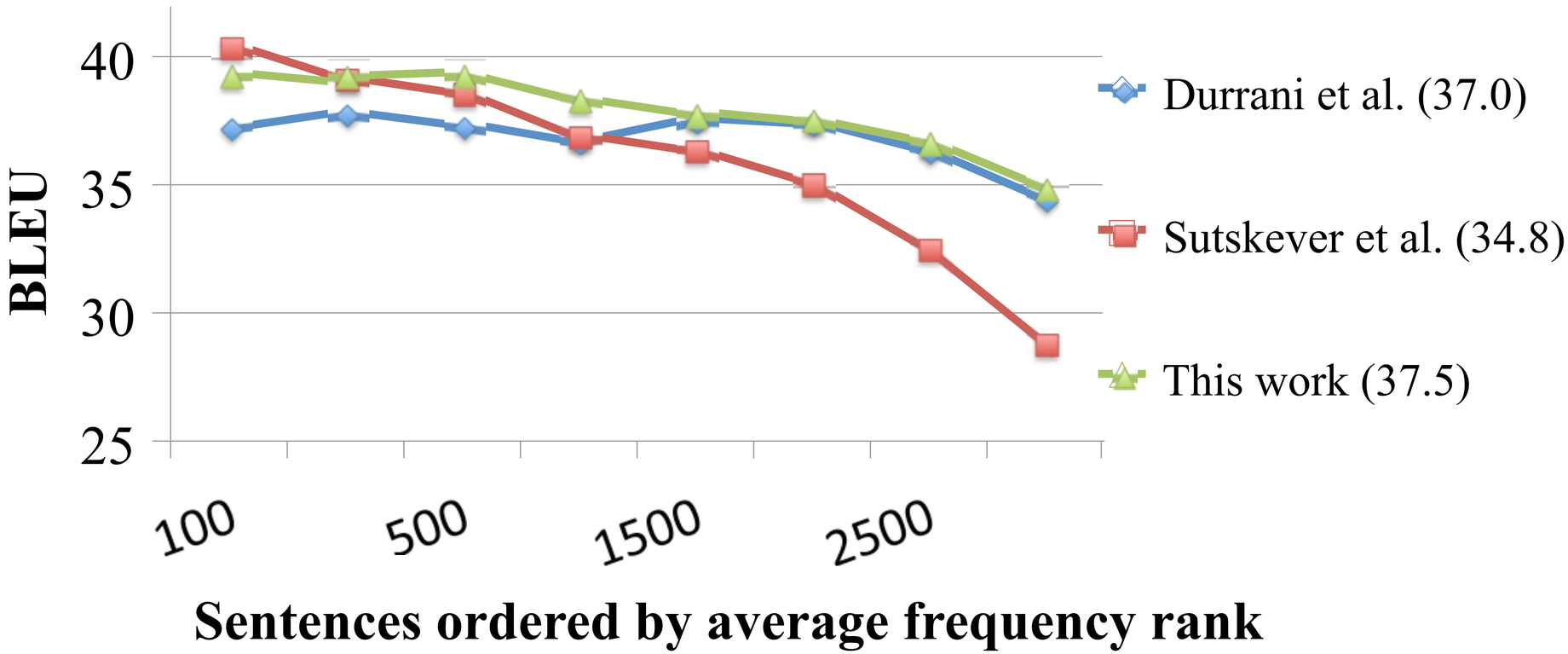
Translation Le portique unk -1 de unk0

Identity copy

Post-edit Translation Le portique écotaxe de Pont-de-Buis

Orthogonal to large-vocab techniques

Effects of Translating Rare Words



First SOTA NMT system!

Sample translations

source	An additional **2600** operations including **orthopedic** and **cataract** surgery will help clear a backlog .
human	**2600** opérations supplémentaires , notamment dans le domaine de la chirurgie **orthopédique** et de la **cataracte** , aideront à ravraper le retard .
trans	En outre , **unk 1** opérations supplémentaires , dont la chirurgie **unk 5** et la **unk 6** , permettront de résorber l' arriéré .
trans +unk	En outre , **2600** opérations supplémentaires , dont la chirurgie **orthopédiques** et la **cataracte** , permettront de résorber l' arriéré .

- Predict well long-distance alignments.
  - Correct: **cataract** vs. ***cataracte*** .

Sample translations

source	This **trader** , Richard **Usher** , lel RBS in **2010** and is understand to have be given leave from his current position as European head of forex spot trading at **JPMorgan** .
human	Ce **trader** , Richard **Usher** , a quitté **RBS** en 2010 et aurait été mis suspendu de son poste de responsable européen du trading au comptant pour les devises chez **JPMorgan** .
trans	Ce **unk 0** , Richard **unk 0** , a quivé **unk 1** en 2010 et a compris qu' il est autorisé à quiver son poste actuel en tant que leader européen du marché des points de vente au **unk 5** .
trans +unk	Ce **négociateur** , Richard **Usher** , a quivé **RBS** en 2010 et a compris qu' il est autorisé à quiver son poste actuel en tant que leader européen du marché des points de vente au **JPMorgan** .

- Translate well long sentences.
  - Correct: **JPMorgan** vs. ***JPMorgan*** .

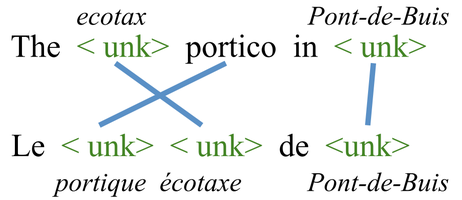
Sample translations

source	But concerns have grown aler Mr **Mazanga** was quoted as saying **Renamo was** abandoning the 1992 peace accord .
human	Mais l' inquiétude a grandi après que M. **Mazanga** a déclaré que la **Renamo** ***abandonnait*** l' accord de paix de 1992 .
trans	Mais les inquiétudes se sont accrues après que M. **unkpos 3** a déclaré que la **unk 3** **unk 3** l' accord de paix de 1992 .
trans +unk	Mais les inquiétudes se sont accrues après que M. **Mazanga** a déclaré que la **Renamo** **était** l' accord de paix de 1992 .

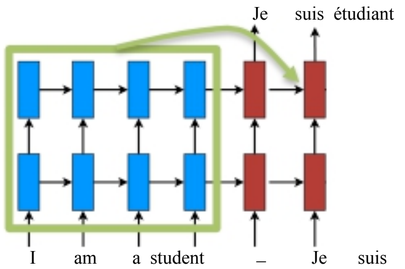
- Incorrect alignment prediction: **was** – **était** vs. ***abandonnait***.

Advancing NMT

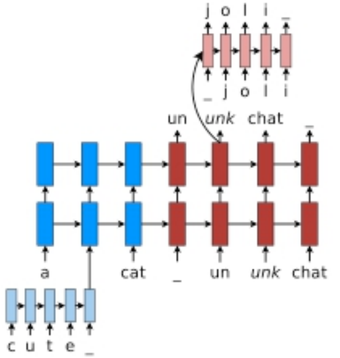
- #1: the *vocabulary size* problem
  - *Sol* : "copy" mechanism.



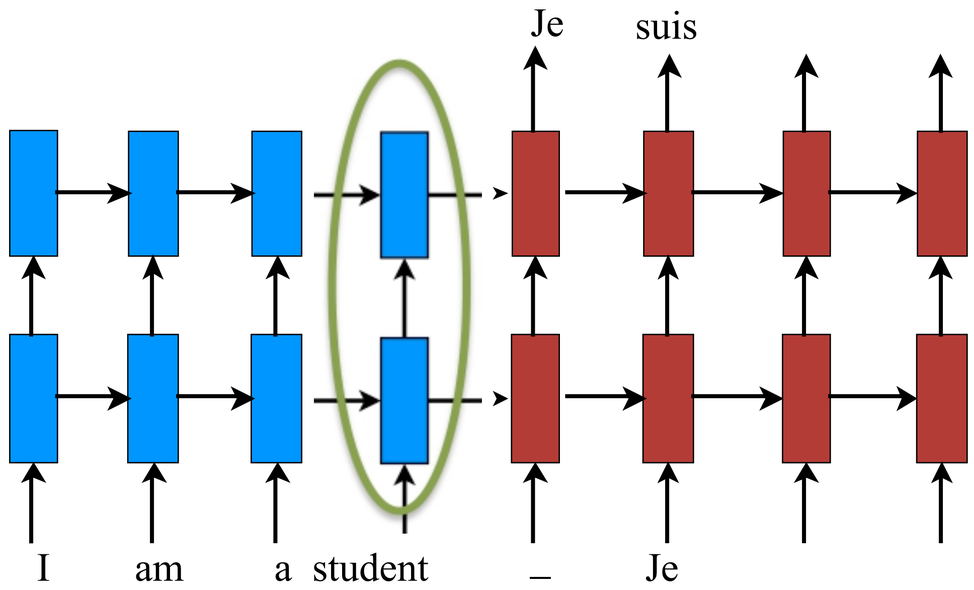
- #2 : the ***sentence length*** problem
  - *Sol* : attention mechanism.



- #3 : the ***language complexity*** problem
  - *Sol* : character-level translation.



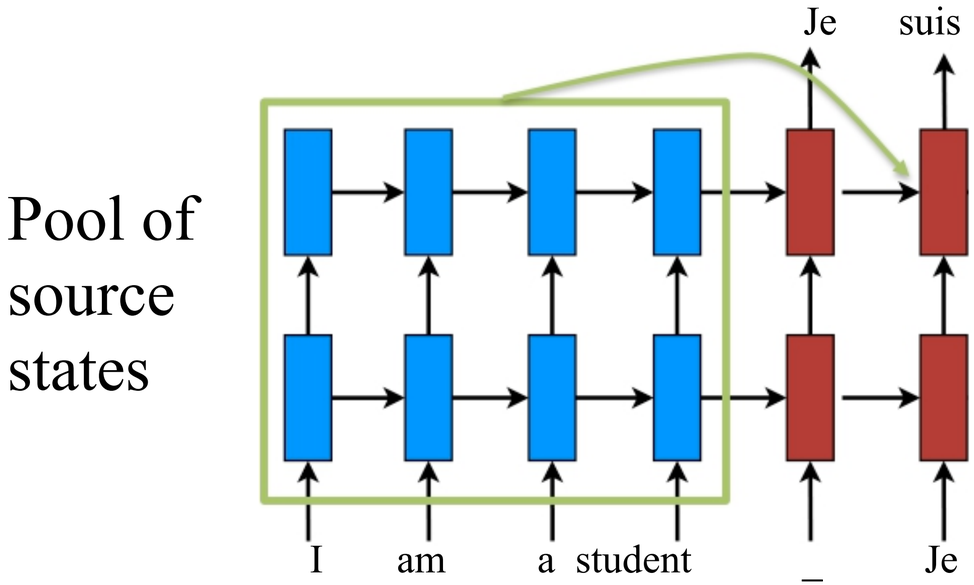
#2 The Sentence Length Problem



- Translation quality degrades with long sentences.

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

Attention Mechanism

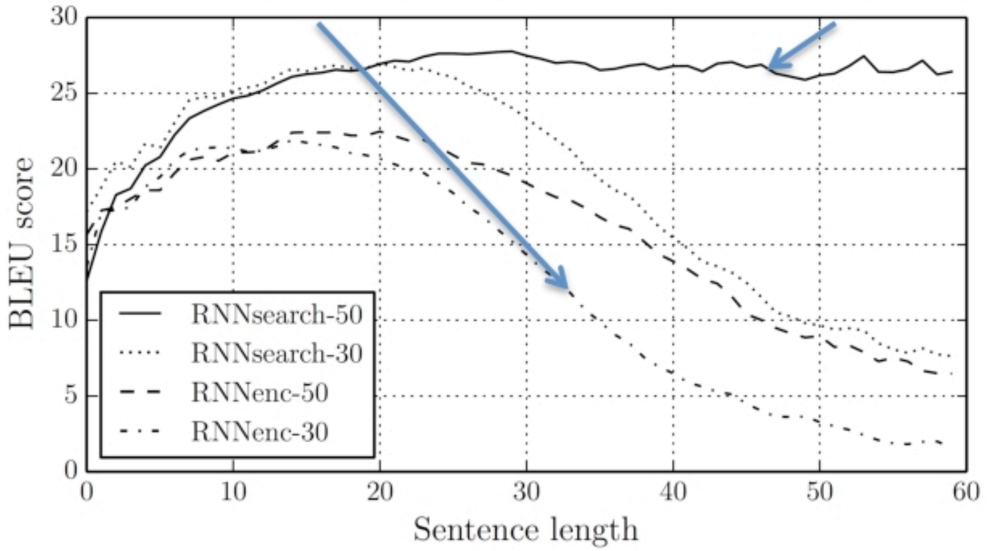


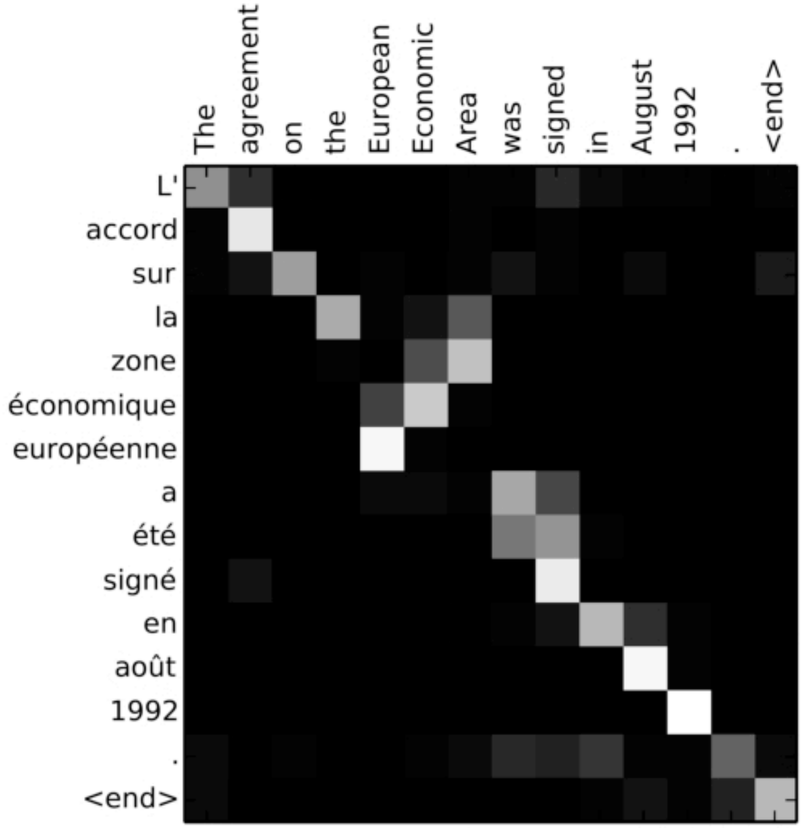
- Solution: random access memory
  - Retrieve as needed.

Université de Montréal

Without attention

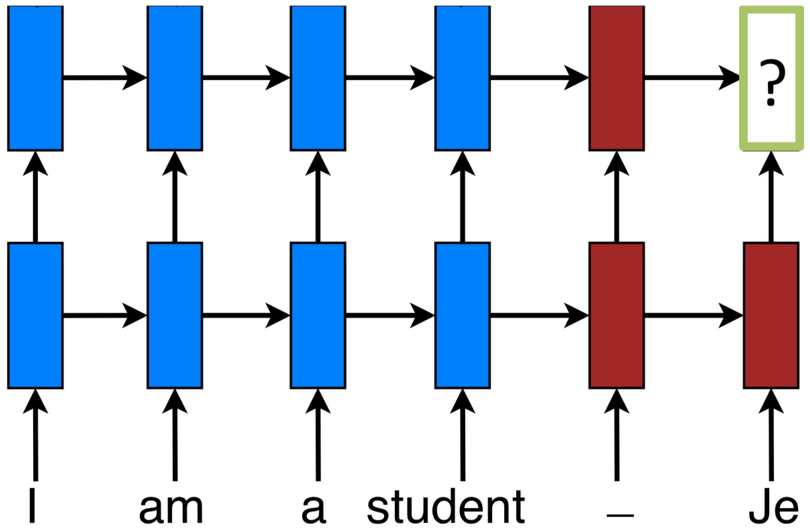
With attention





Dzmitry Bahdanau , KyungHuyn Cho, and Yoshua Bengio. ***Neural Machine Translation by Jointly Learning to Translate and Align*** *. ICLR 2015.*

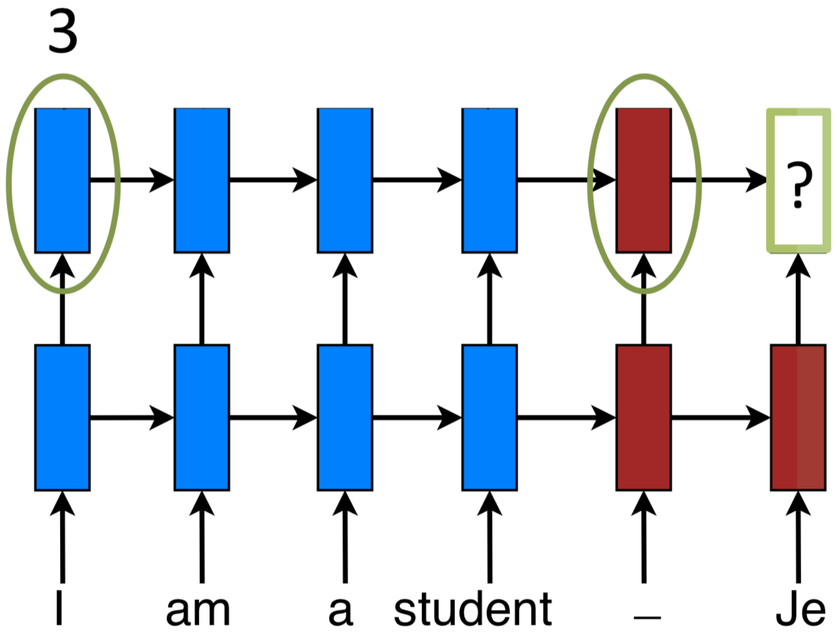
Attention Mechanism



A simplified version of (Bahdanau et al., 2015)

Attention Mechanism – *Scoring*

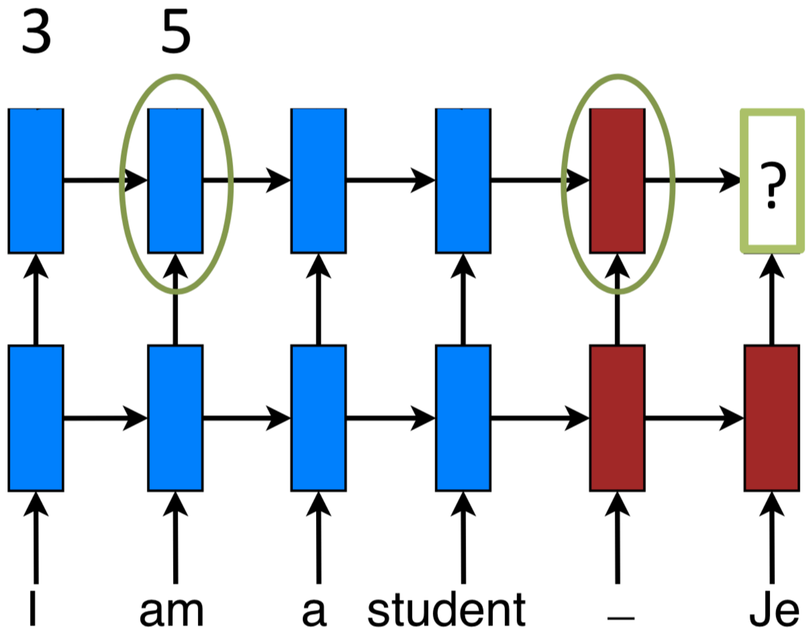
$$\text{score}(\boldsymbol{h}_t, \bar{\boldsymbol{h}}_s)$$



- Compare target and source hidden states.

Attention Mechanism – *Scoring*

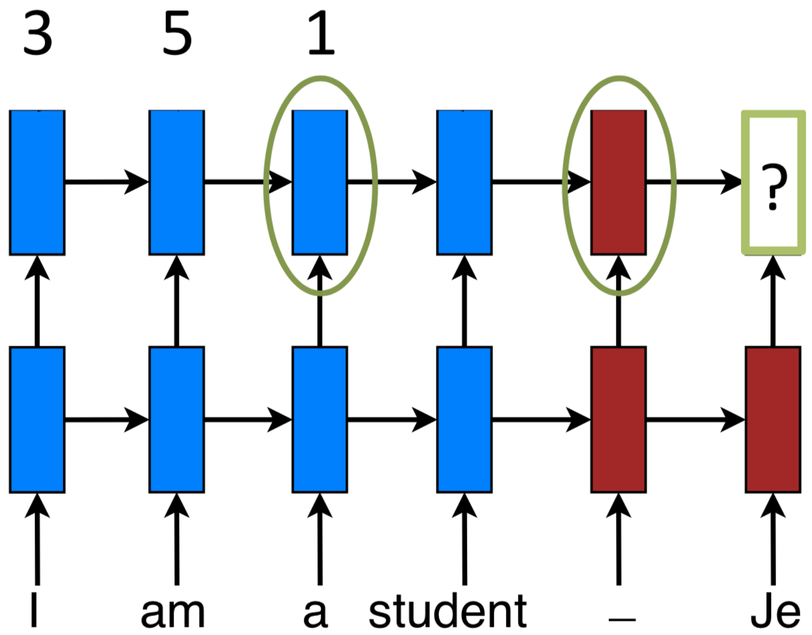
$$\text{score}(\boldsymbol{h}_t, \bar{\boldsymbol{h}}_s)$$



- Compare target and source hidden states.

Attention Mechanism – *Scoring*

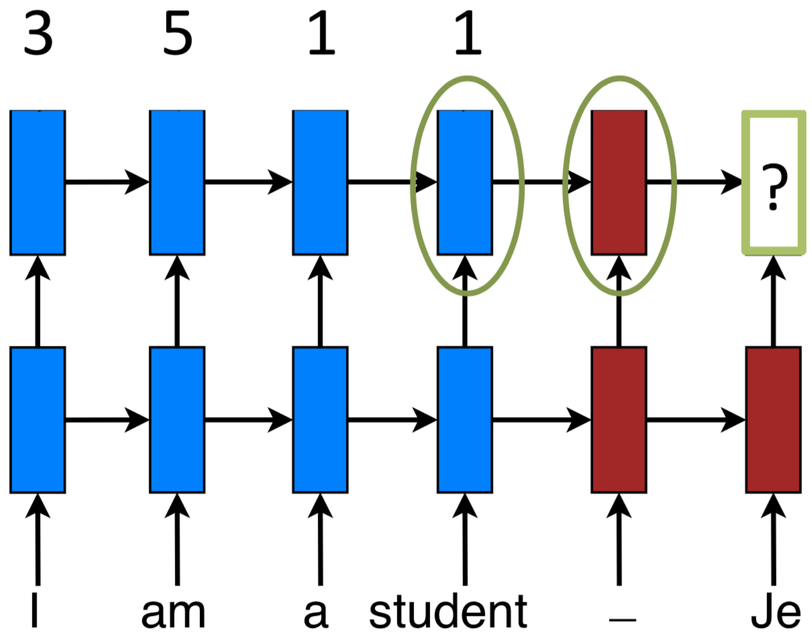
$$\text{score}(\boldsymbol{h}_t, \bar{\boldsymbol{h}}_s)$$



- Compare target and source hidden states.

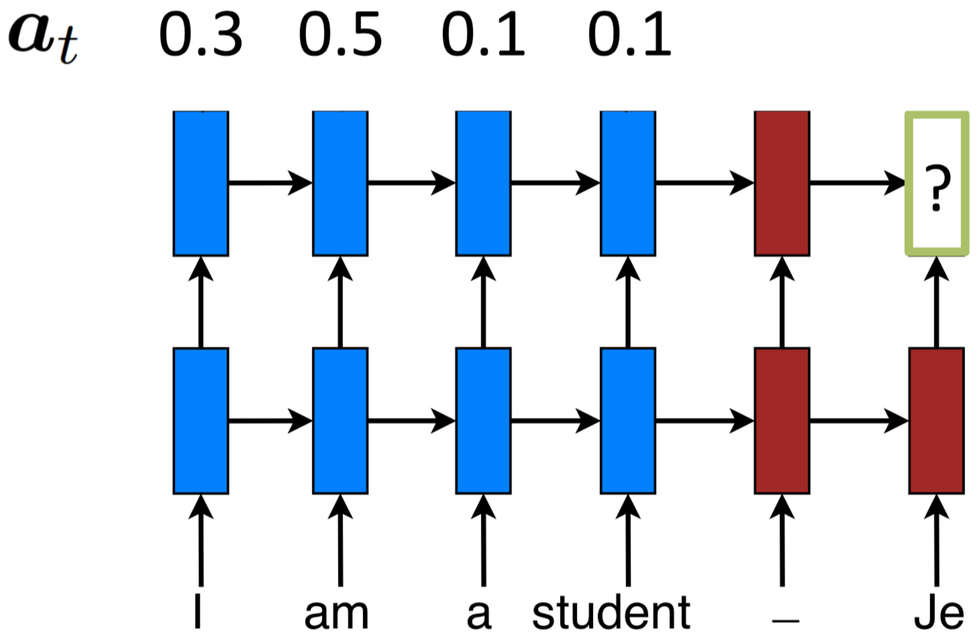
Attention Mechanism – *Scoring*

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



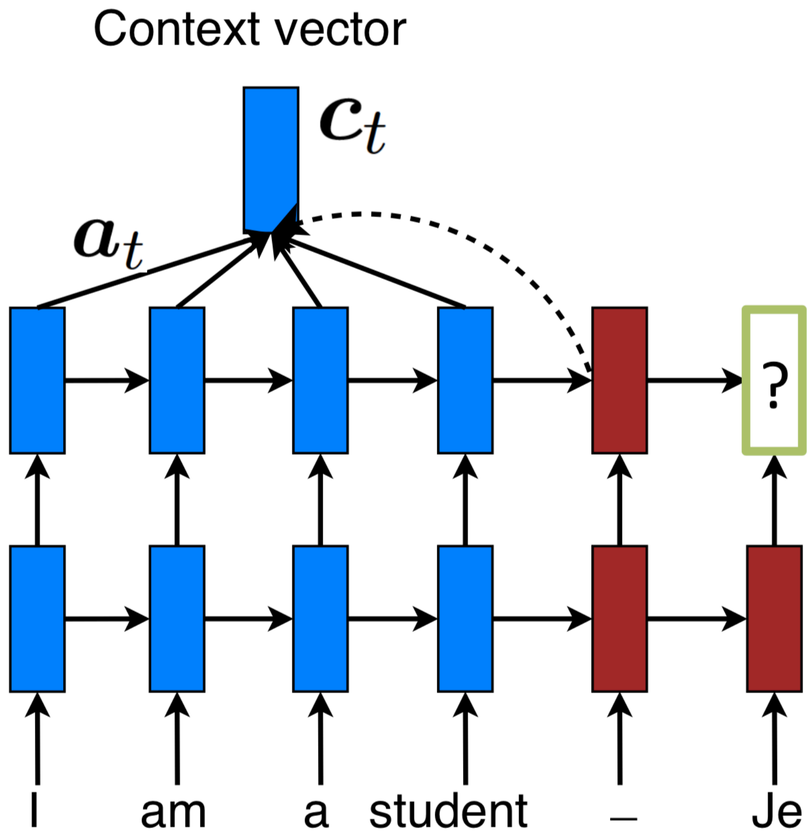
- Compare target and source hidden states.

Attention Mechanism – *Normalization*



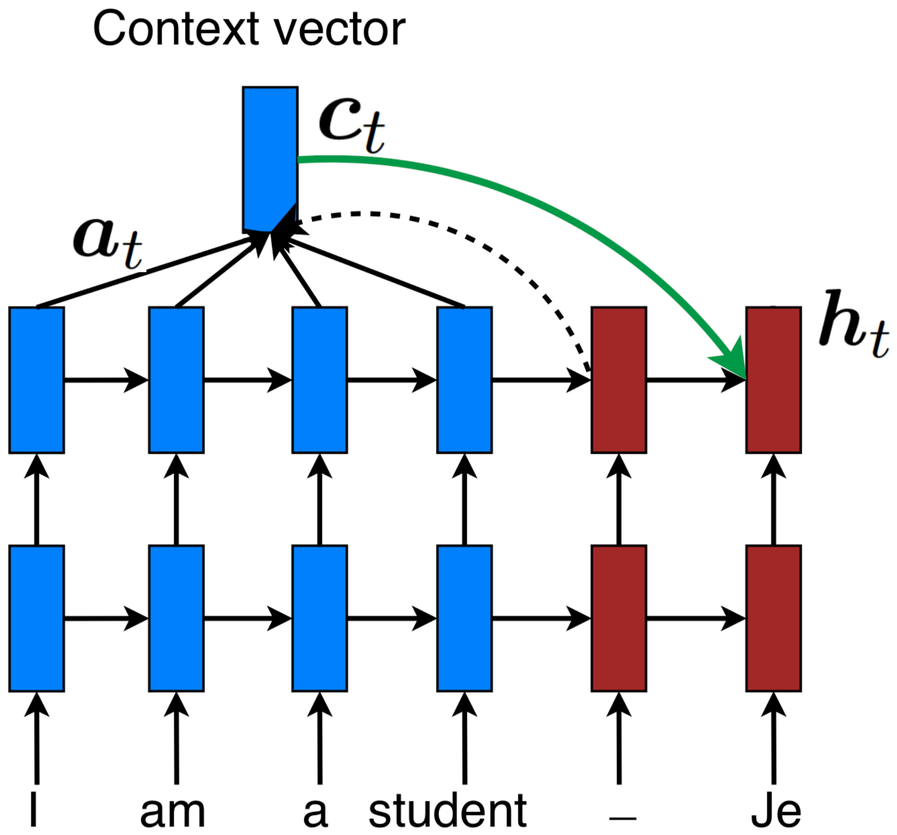
- Convert into alignment weights.

Attention Mechanism – *Context vector*



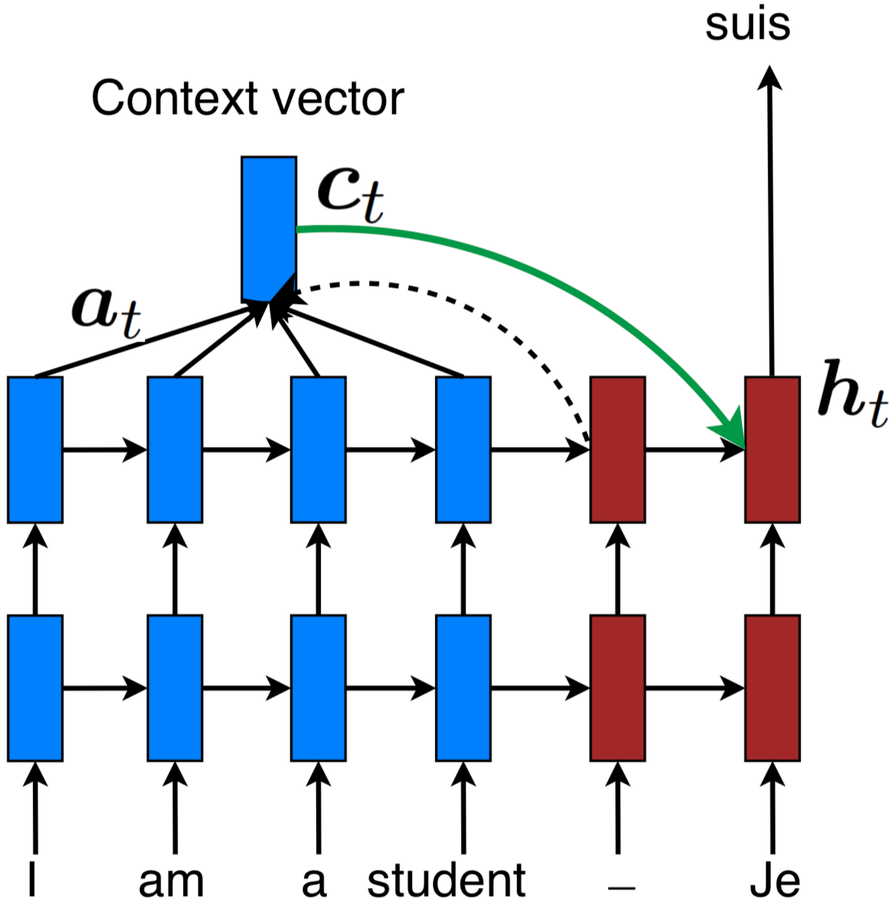
- Build context vector: weighted average.

Attention Mechanism – *Hidden state*



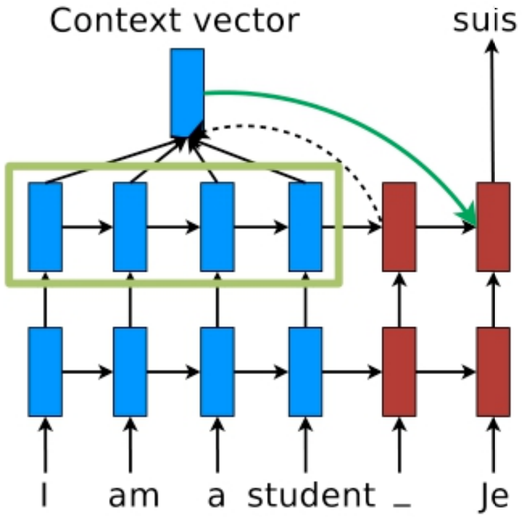
- Compute the next hidden state.

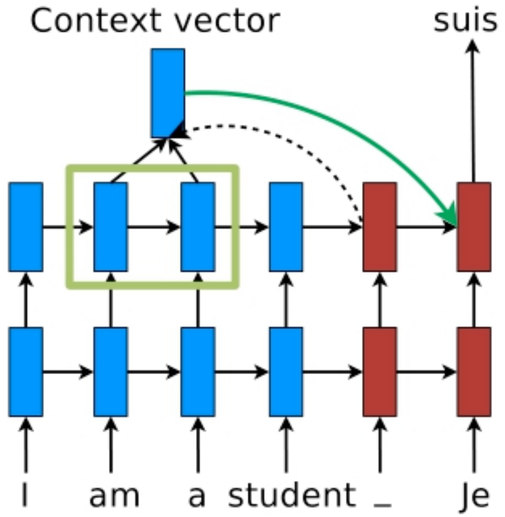
Attention Mechanism – *Predict*



- Predict the next word .

- Examine various attention mechanisms:





Global : ***all*** source states.

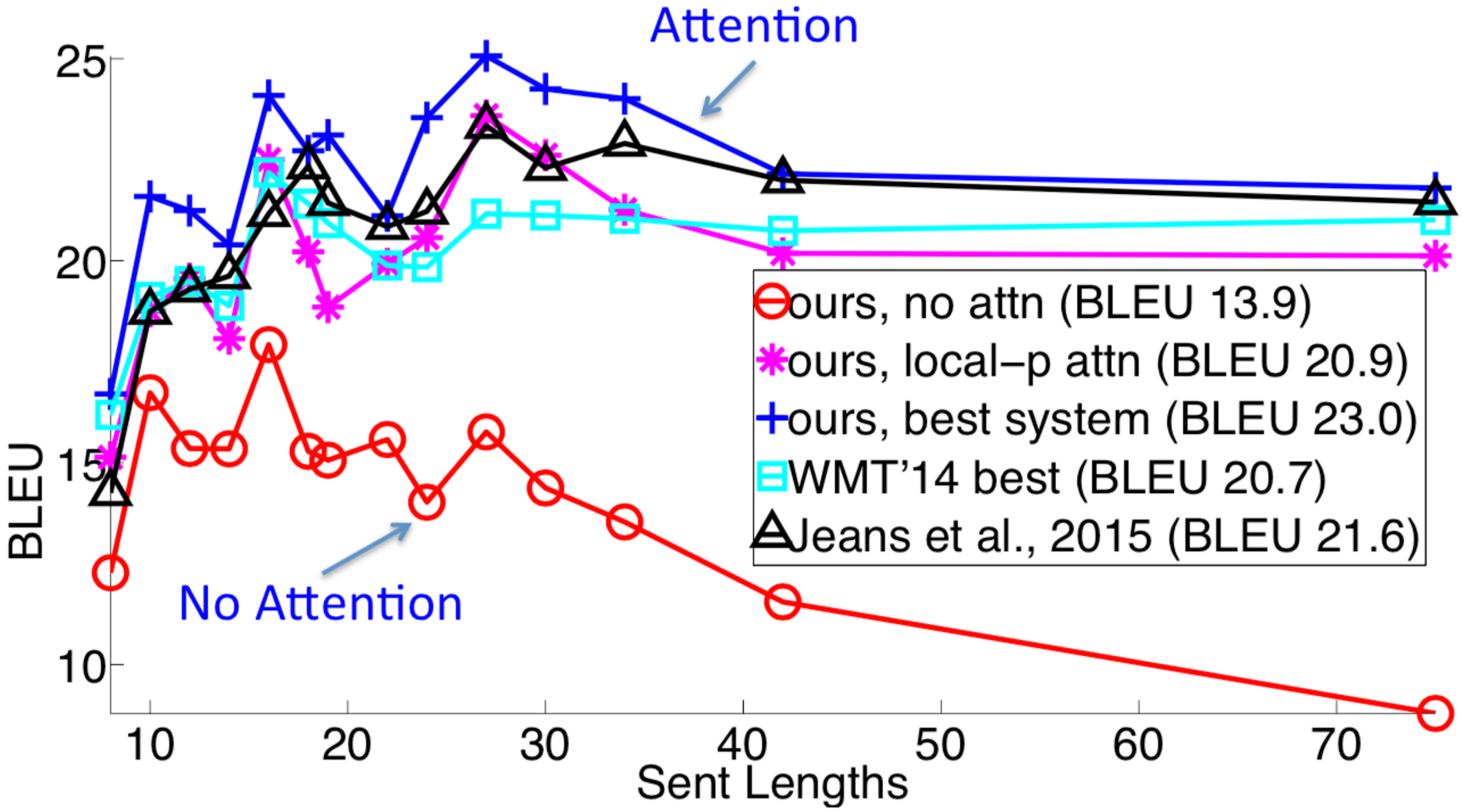
Local: ***subset*** of source states.

SOTA for English-German translation.

Thang Luong, Hieu Pham, and Chris Manning. ***Effective Approaches to Attention-based Neural Machine Translation*** *. EMNLP 2015.*

Used with permission from Thong Luong

Translate Long Sentences



Sample English-German translations

source	Orlando Bloom and *Miranda Kerr* still love each other
human	Orlando Bloom und **Miranda Kerr** lieben sich noch immer
best	Orlando Bloom und **Miranda Kerr** lieben einander noch immer .
base	Orlando Bloom und **Lucas Miranda** lieben einander noch immer .

- Translate names correctly.

Sample English-German translations

source	We ′ re pleased the FAA recognizes that an enjoyable passenger experience is ***not incompatible*** with safety and security , said Roger Dow , CEO of the U.S. Travel Association .
human	Wir freuen uns , dass die FAA erkennt , dass ein angenehmes Passagiererlebnis nicht **im Wider- spruch zur Sicherheit steht** , sagte Roger Dow , CEO der U.S. Travel Association .
best	Wir freuen uns , dass die FAA anerkennt , dass ein angenehmes ist nicht mit Sicherheit und Sicherheit **unvereinbar** ist , sagte Roger Dow , CEO der US - die .
base	Wir freuen uns u ber die < unk> , dass ein <unk> < unk> mit Sicherheit nicht **vereinbar** ist mit Sicherheit und Sicherheit , sagte Roger Cameron , CEO der US - <unk> .

- Translate a doubly-negated phrase correctly

Fail to ensure the passenger experience is

Sample English-German translations

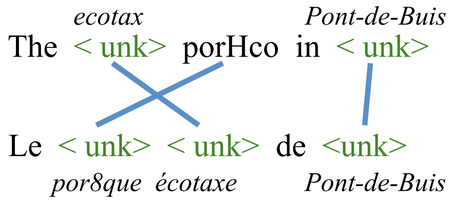
source	We ′ re pleased the FAA recognizes that an enjoyable passenger experience is ***not incompatible*** with safety and security , said Roger Dow , CEO of the U.S. Travel Association .
human	Wir freuen uns , dass die FAA erkennt , dass ein angenehmes Passagiererlebnis nicht **im Wider- spruch zur Sicherheit steht** , sagte Roger Dow , CEO der U.S. Travel Association .
best	Wir freuen uns , dass die FAA anerkennt , dass ein angenehmes ist nicht mit Sicherheit und Sicherheit **unvereinbar** ist , sagte Roger Dow , CEO der US - die .
base	Wir freuen uns u ber die < unk> , dass ein <unk> < unk> mit Sicherheit nicht **vereinbar** ist mit Sicherheit und Sicherheit , sagte Roger Cameron , CEO der US - <unk> .

- Translate a doubly-negated phrase correctly

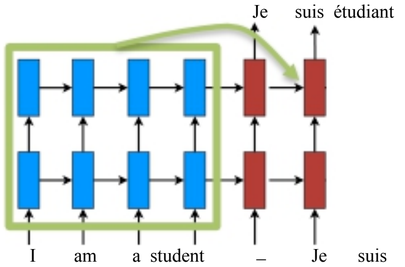
Fail to ensure that passenger experience is

Advancing NMT

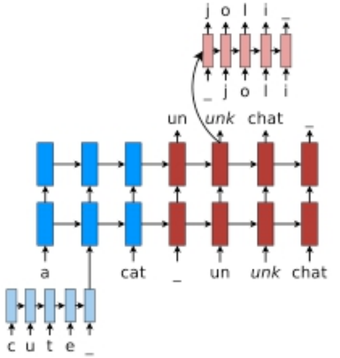
- #1: the *vocabulary size* problem
  - *Sol* : "copy" mechanism.



- #2: the *sentence length* problem
  - *Sol* : avenHon mechanism.



- #3 : the ***language complexity*** problem
  - *Sol* : character-level translation.



#3 The rare word problem

- "Copying" mechanisms are **not sufficient**.
  - Different alphabets: *Christopher* ↔ *Kryštof*
  - Multi-word alignment: *Solar system* ↔ *Sonnensystem*

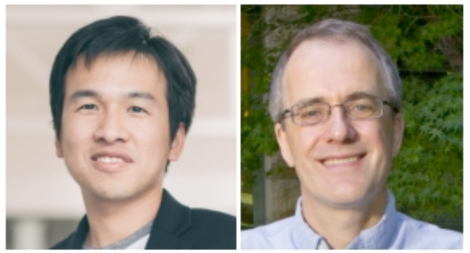
- Need to handle **large, open vocabulary**
  - *Rich morphology*: *nejneobhospodařovávatelnějšímu*
    (*"to the worst farmable one"*)
  - *Informal spelling*: *goooooood morning !!!!!*

Be able to generate at the character level.

Previous character-level NMT

- Unsatisfactory performance
  - (Wang Ling, Isabel Trancoso , Chris Dyer, Alan Black, arXiv 2015)

- Incomplete solution
  - Decoder only (Junyoung Chung, Kyunghyun Cho, Yoshua Bengio. arXiv 2016).



- A *best-of-both-worlds* architecture:
  - Translate mostly at the word level
  - Only go the character level when needed.

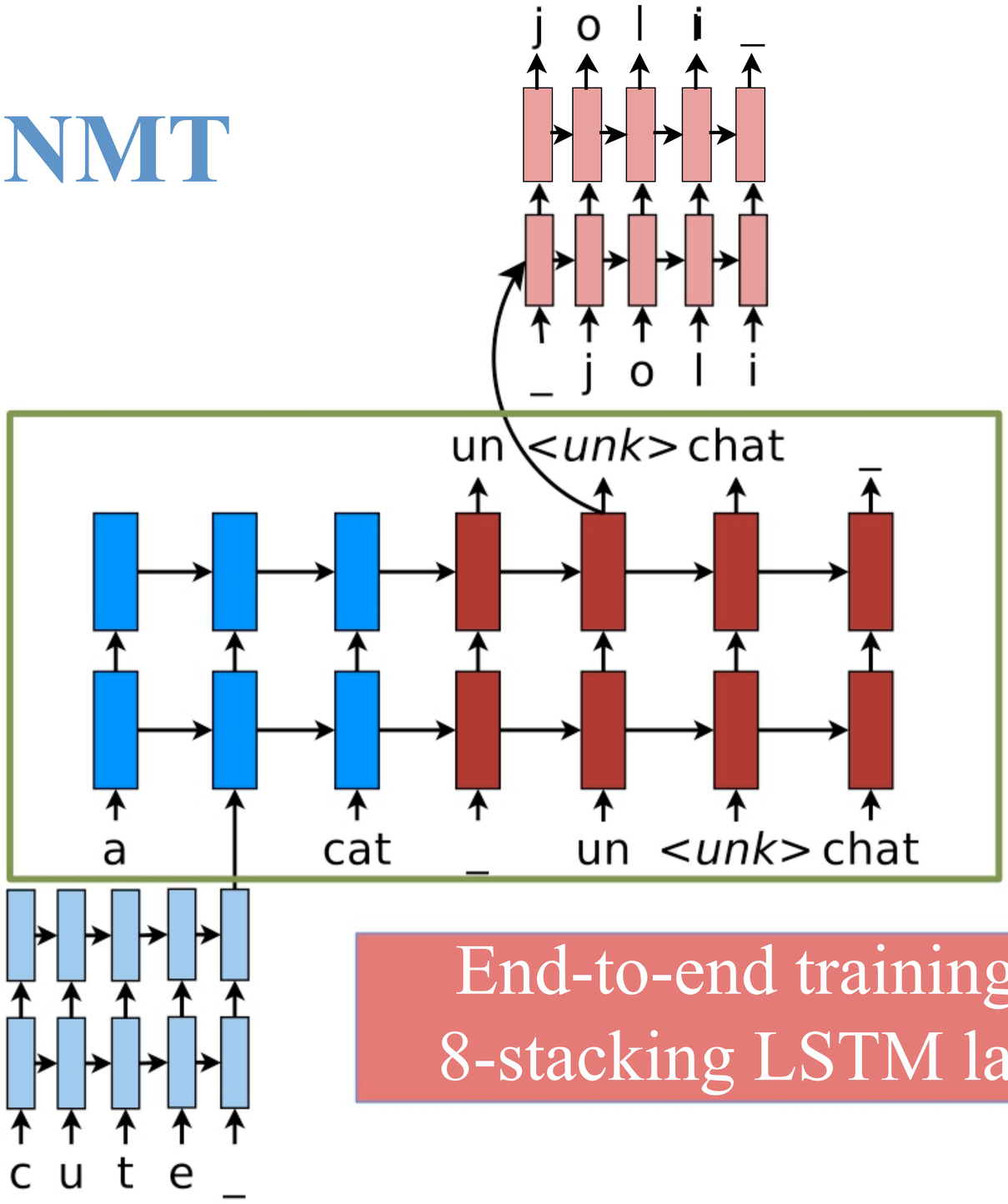
- Additional +11.4 BLEU improvement.

SOTA for English-Czech translaHon.

Thang Luong and Chris Manning. ***Achieving Open Vocabulary Neural Machine Translation with Hybrid Word-Character Models****. ACL 2016.*

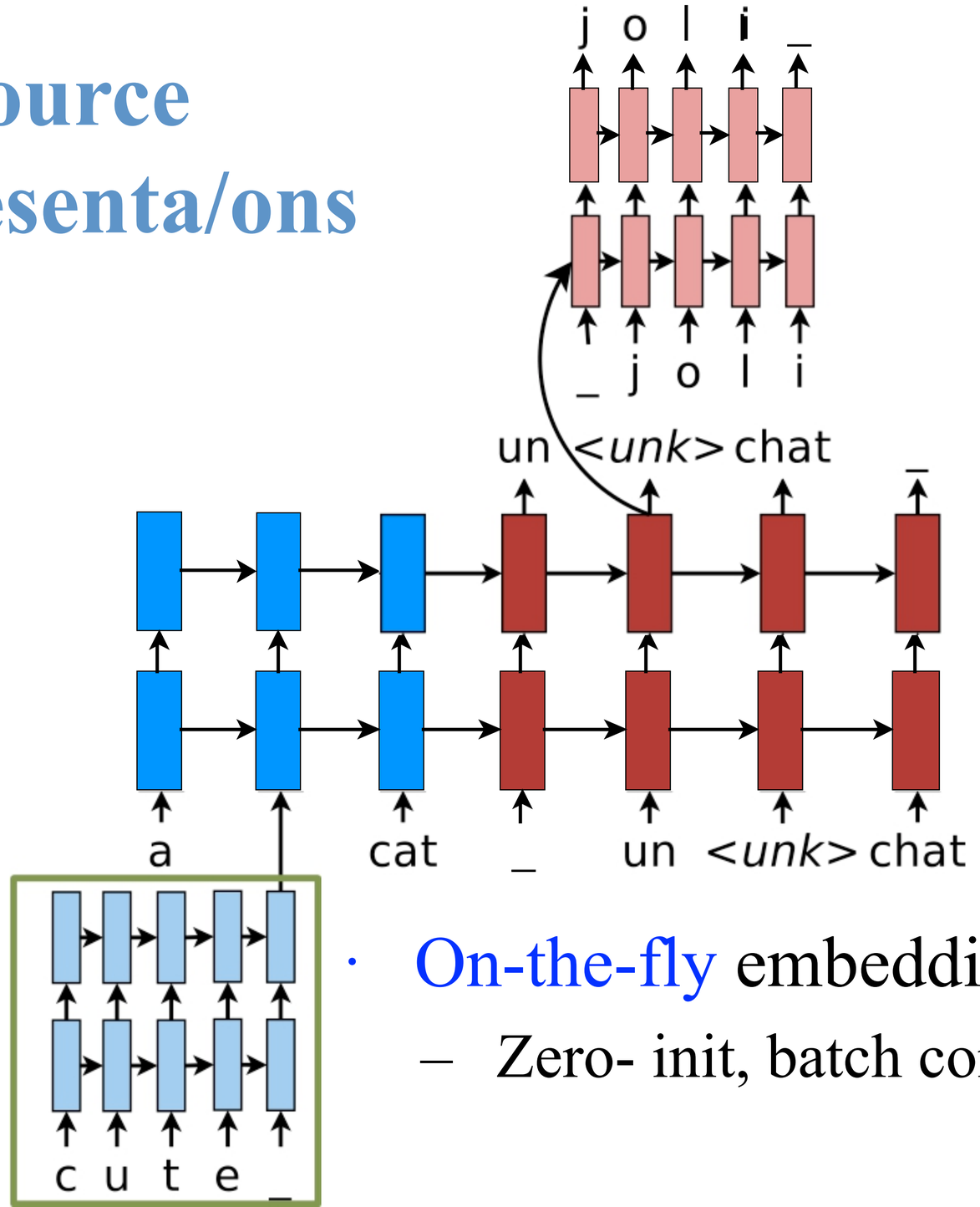
Hybrid NMT

Word-level
(4 layers)



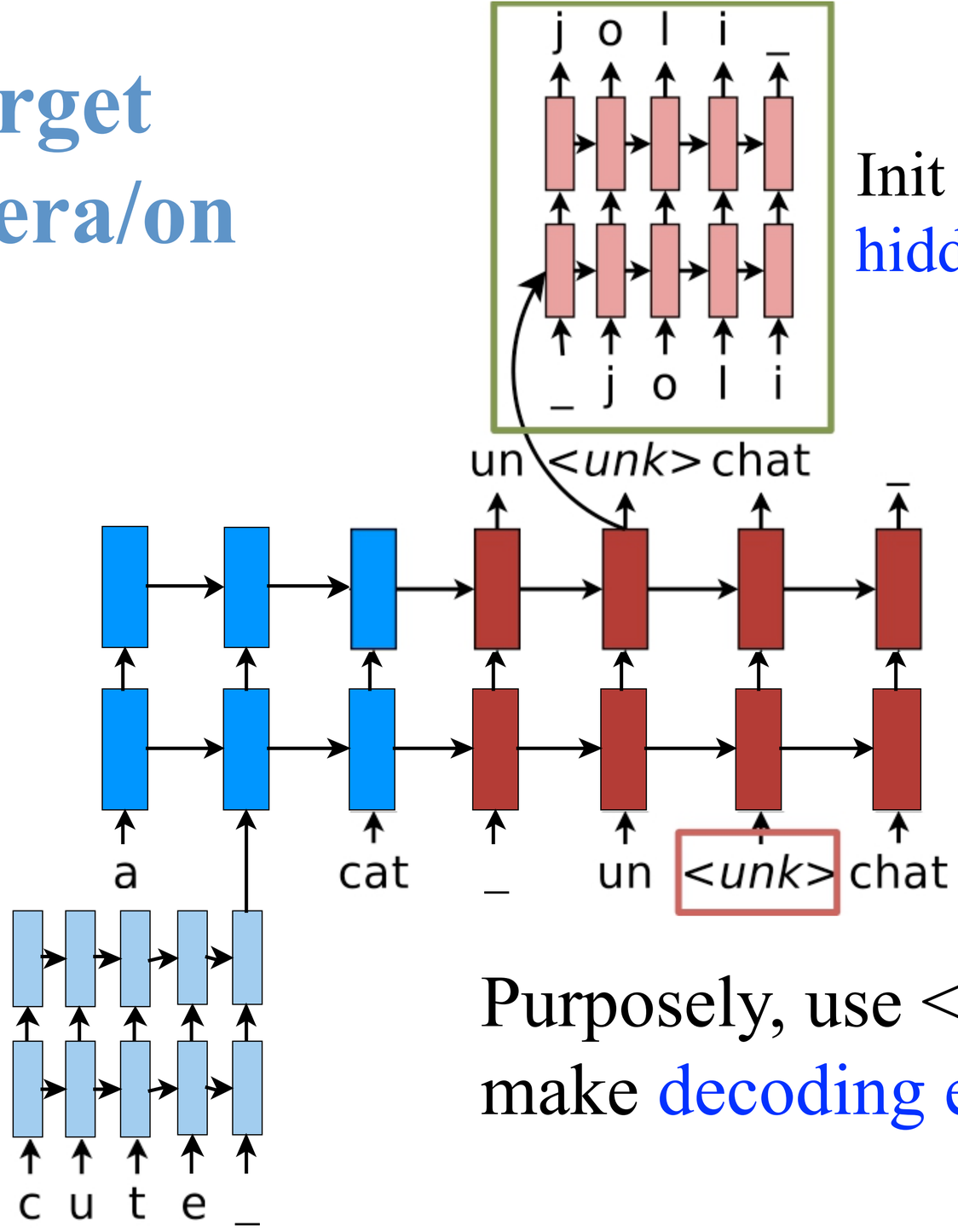
End-to-end training
8-stacking LSTM layers.

Source Representa/ons



- On-the-fly embeddings.
  - Zero- init, batch computation.

Target Genera/on



Init with word hidden states.

Purposely, use <unk> to make decoding easier.

English-Czech WMT'15 Results

Systems	BLEU
Winning entry (Bojar & Tamchyna , 2015)	***18.8***
Existing ***word-level*** *NMT* (Jean et al., 2015)	
Single model	15.7
Ensemble 4 models	18.3

30x data
3 systems

Large vocab
+ unk replace

English-Czech WMT'15 Results

Systems	BLEU
Winning entry (Bojar & Tamchyna , 2015)	***18.8***
Existing ***word-level*** *NMT* (Jean et al., 2015)	
Single model	15.7
Ensemble 4 models	18.3
Character-based *NMT*	
Single model (600-step backprop)	15.9

30x data
3 systems

Large vocab
+ unk replace

- Purely character-based: **slow but promising**!

English-Czech WMT'15 Results

Systems	BLEU
Winning entry (Bojar & Tamchyna , 2015)	***18.8***
Existing ***word-level*** *NMT* (Jean et al., 2015)	
Single model	15.7
Ensemble 4 models	18.3
Character-based *NMT*	
Single model (600-step backprop)	15.9
Hybrid *NMT*	
Single model	**19.6**

30x data
3 systems

Large vocab
+ unk replace

New SOTA!

English-Czech WMT'15 Results

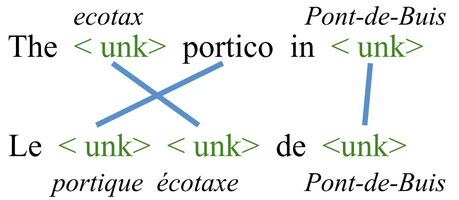
Systems	BLEU
Winning entry (Bojar & Tamchyna , 2015)	***18.8***
Existing ***word-level*** *NMT* (Jean et al., 2015)	
Single model	15.7
Ensemble 4 models	18.3
Character-based *NMT*	
Single model (600-step backprop)	15.9
Hybrid *NMT*	
Single model	19.6
Ensemble 4 models	**20.7**

30x data
3 systems

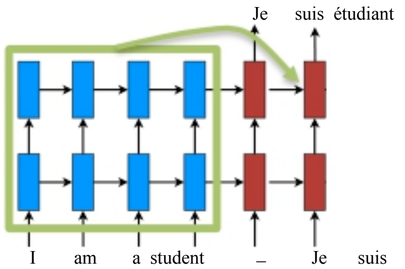
Large vocab
+ unk replace

Advanced NMT

- #1 : the ***vocabulary size*** problem
  - *Sol* : "copy" mechanism.



- #2 : the ***sentence length*** problem
  - *Sol* : attention mechanism.



- #3 : the ***language complexity*** problem
  - *Sol* : character-level translation.