Sequence to Sequence Models

Deep Learning

Woohwan Jung

한양대학교
HANYANG UNIVERSITY

DSLAB
데이터사이언스 연

Outline

- Basic models
- Beam search
- Attention mechanism

Basic models

Sequence to Sequence Model

입력/출력 길이가 다르

$x^{<1>}$ $x^{<2>}$ $x^{<3>}$ $x^{<4>}$ $x^{<5>}$

Jane visite l'Afrique en septembre

Jane is visiting Africa in September

$y^{<1>}$ $y^{<2>}$ $y^{<3>}$ $y^{<4>}$ $y^{<5>}$ $y^{<6>}$

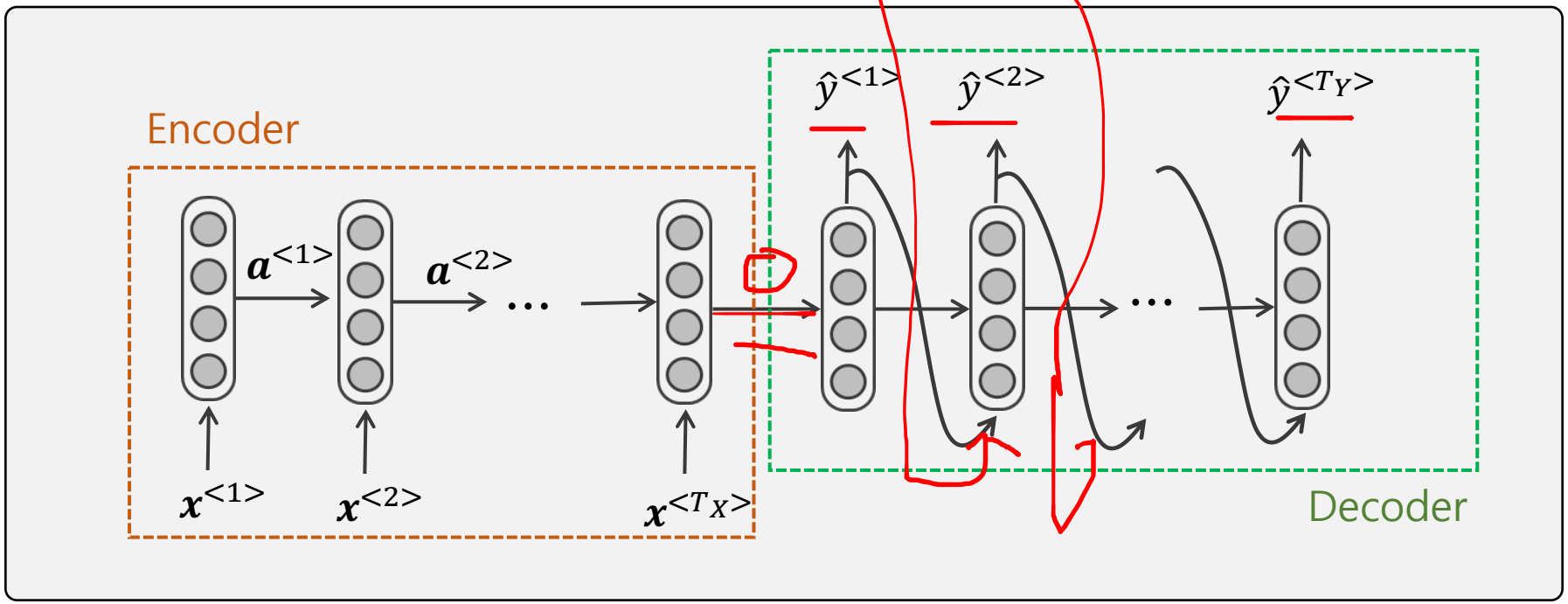
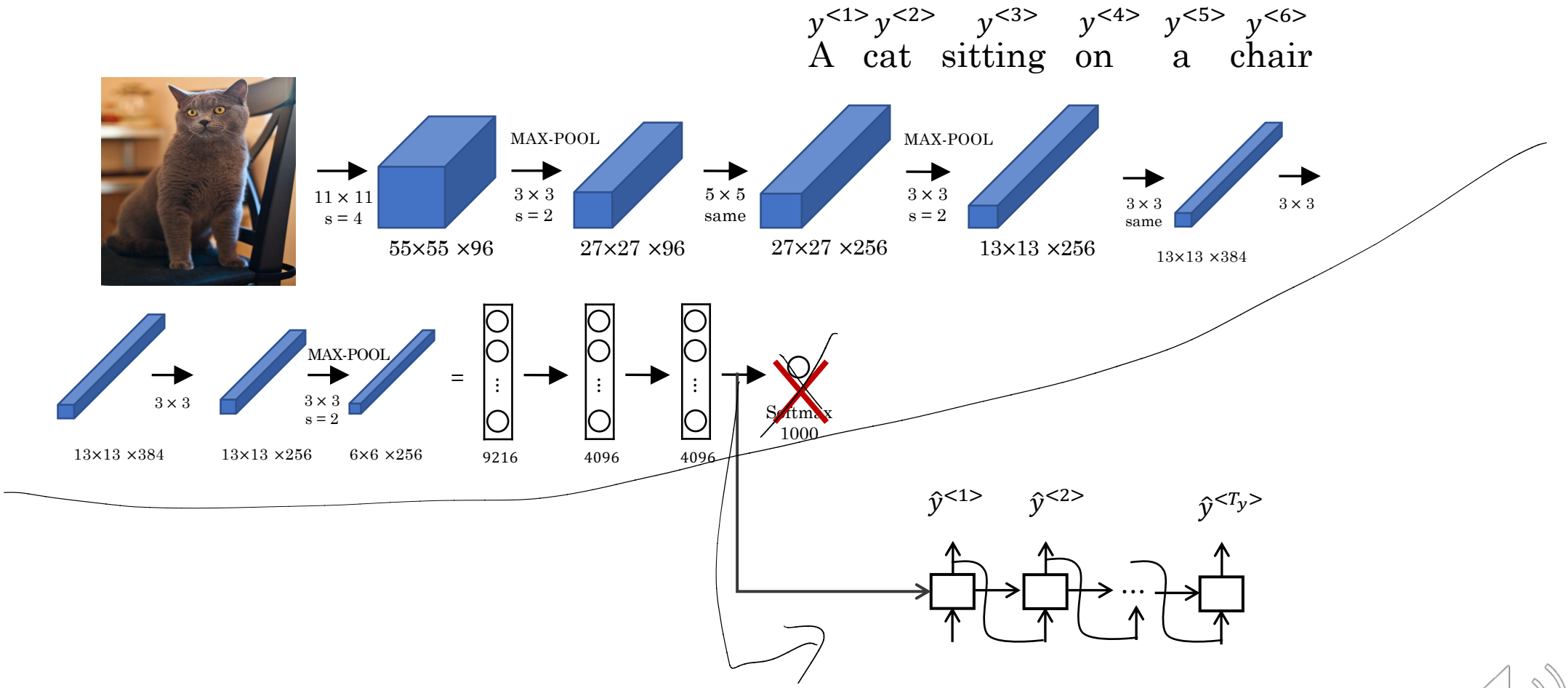


Image Captioning

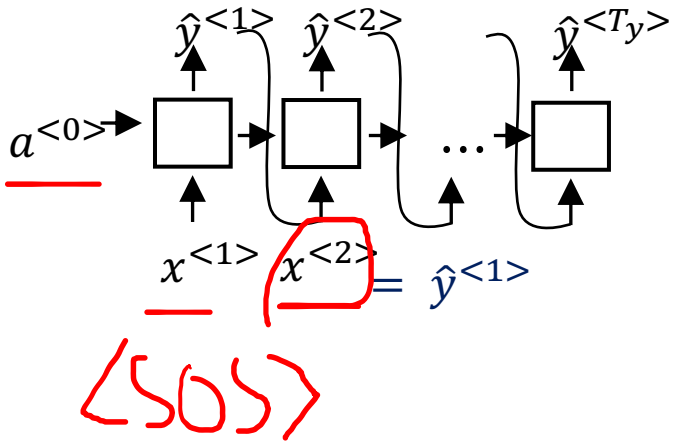
CNN으로 이미지 인식

RNN으로 문장 생성



Machine Translation as Building a Conditional Language Model

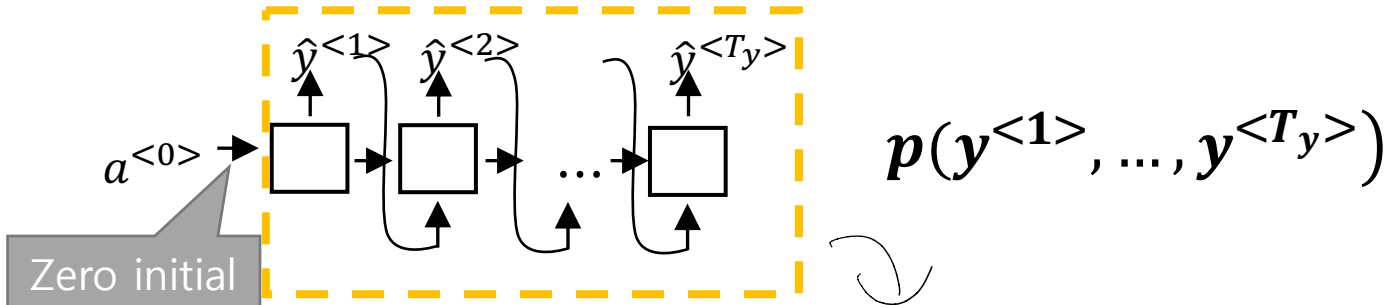
Language model:



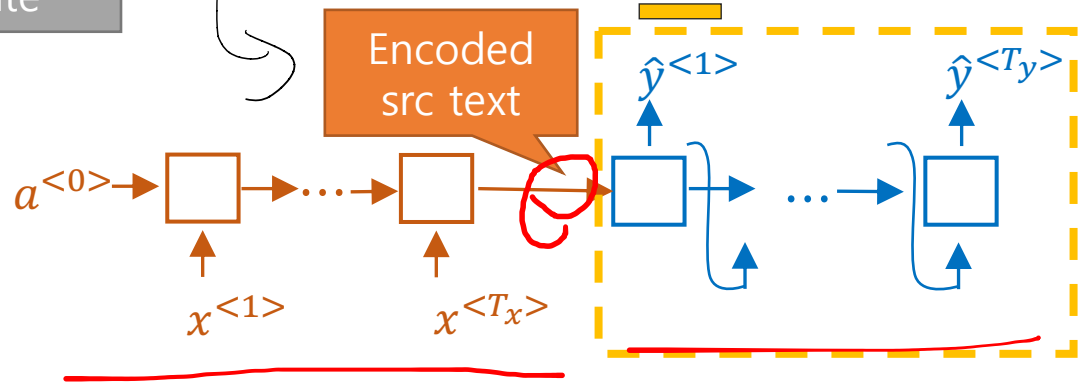
$$p(y^{<1>}, \ldots, y^{<T_y>})$$

Machine Translation as Building a Conditional Language Model

Language model:



Machine translation:



Conditional language model $p(y^{<1>}, \ldots, y^{<T_y>} \mid x^{<1>}, \ldots, x^{<T_x>})$

Finding the Most likely Translation

Jane visite l'Afrique en septembre.

$$P(y^{<1>}, \dots, y^{<T_y>} | x)$$

English

French

→ Jane is visiting Africa in September.

→ Jane is going to be visiting Africa in September.

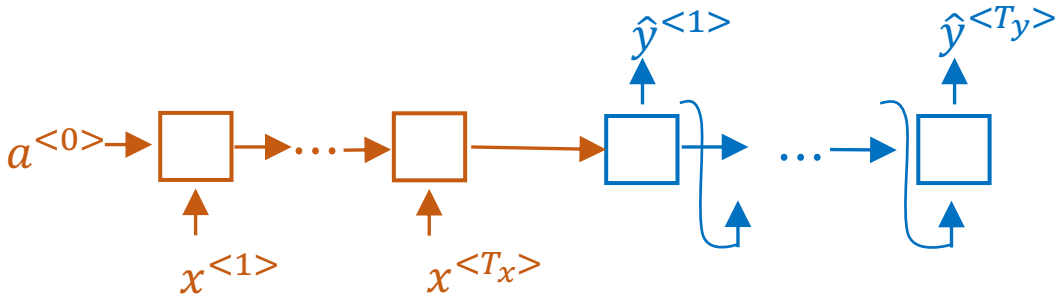
→ In September, Jane will visit Africa.

→ Her African friend welcomed Jane in September.

$$\arg\max_{y^{<1>},\dots,y^{<T_y>}} P(\hat{y}^{<1>}, \hat{y}^{<2>}, \dots, y^{<T_y>} | x)$$

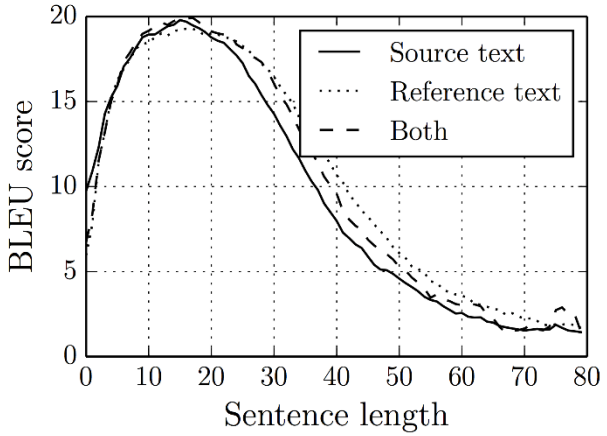
Attention mechanism

The Problems of Long Sequences



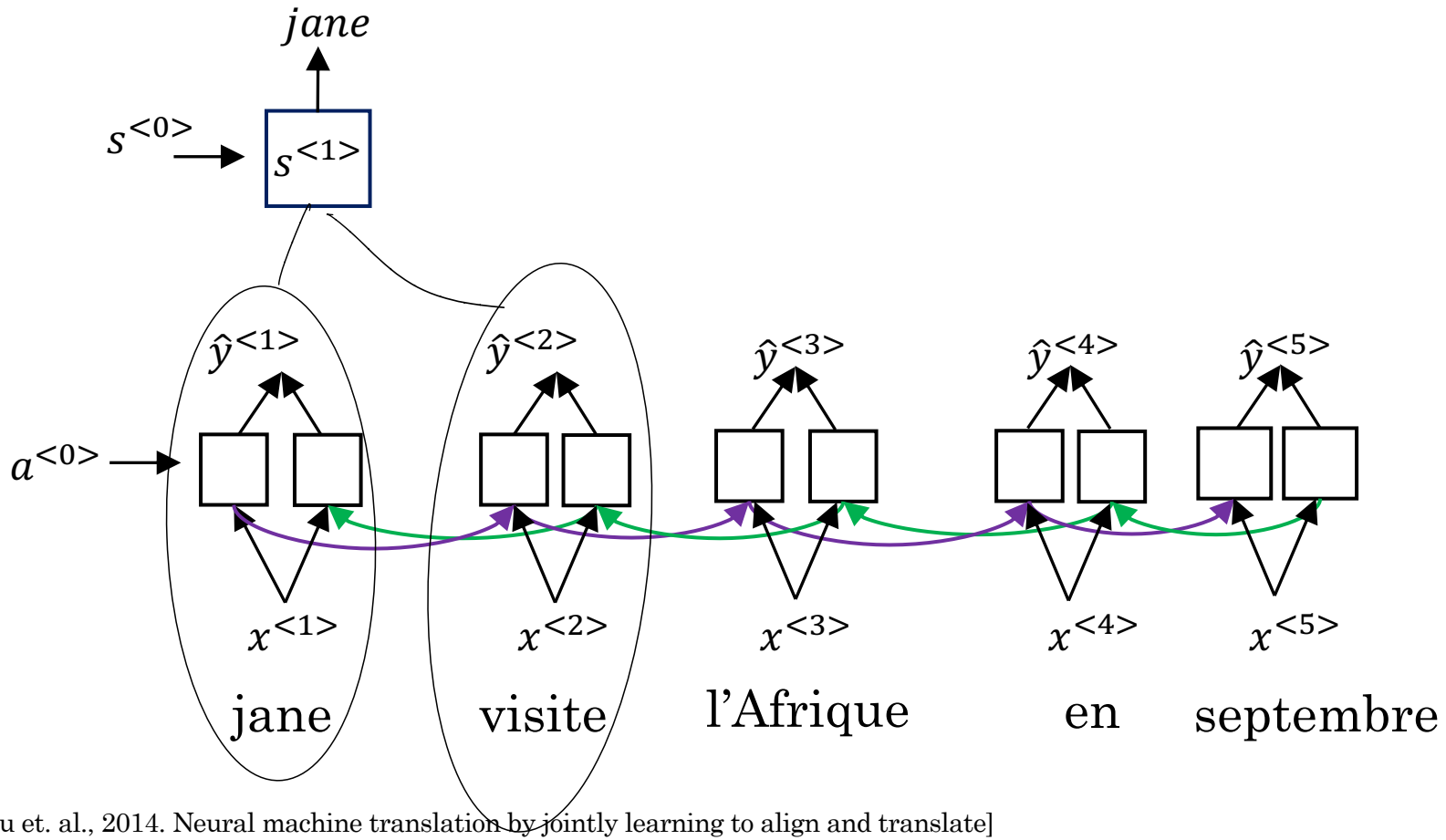
Jane s'est rendue en Afrique en septembre dernier, a apprécié la culture et a rencontré beaucoup de gens merveilleux; elle est revenue en parlant comment son voyage était merveilleux, et elle me tent e d'y aller aussi.

Jane went to Africa last September, and enjoyed the culture and met many wonderful people; she came back raving about how wonderful her trip was, and is tempting me to go too.



https://devblogs.nvidia.com/introduction-neural-machine-translation-gpus-part-3/figure1_bleuscore_vs_sentencelength/

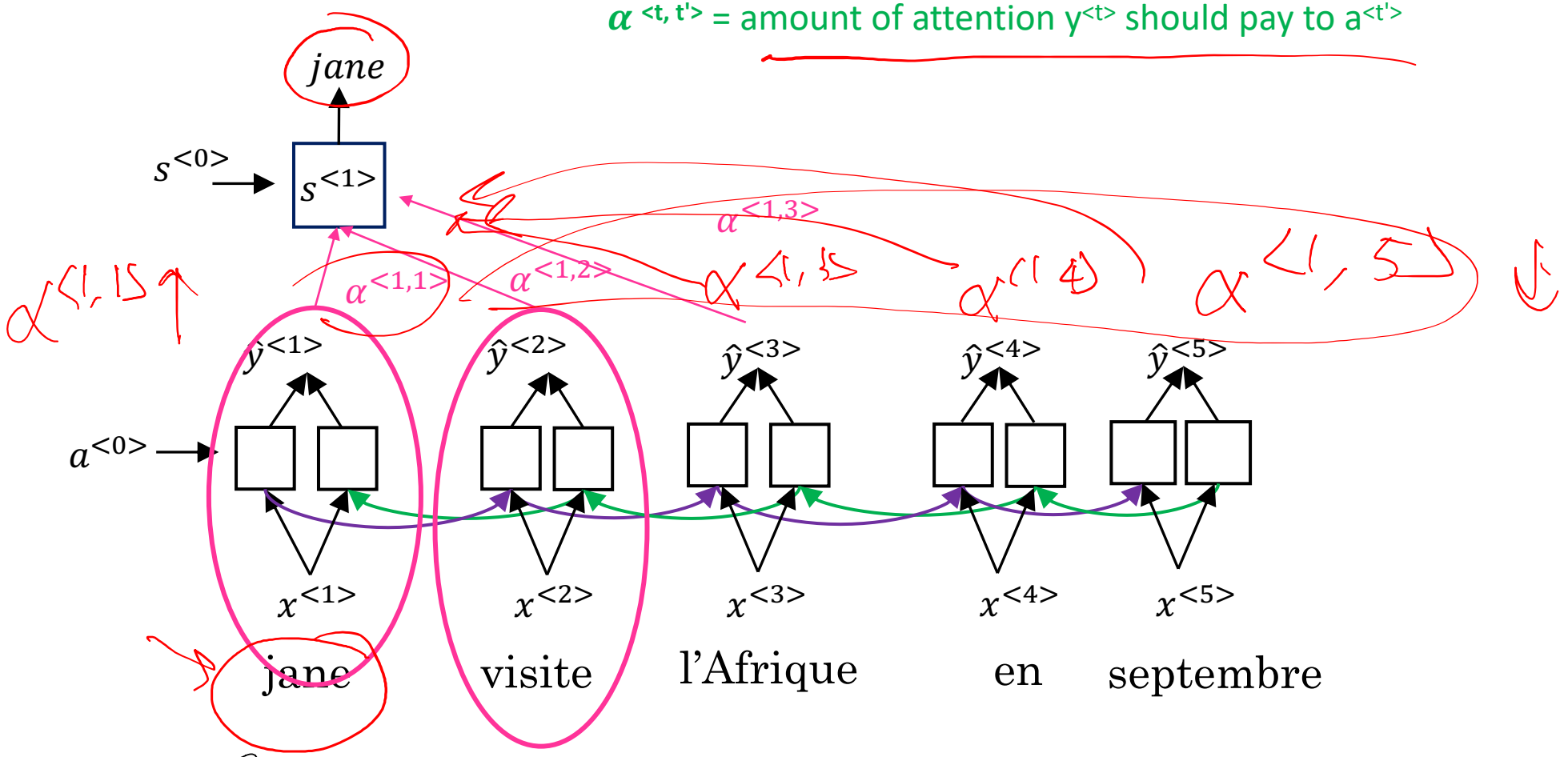
Attention Model intuition



[Bahdanau et. al., 2014. Neural machine translation by jointly learning to align and translate]

Attention Model intuition

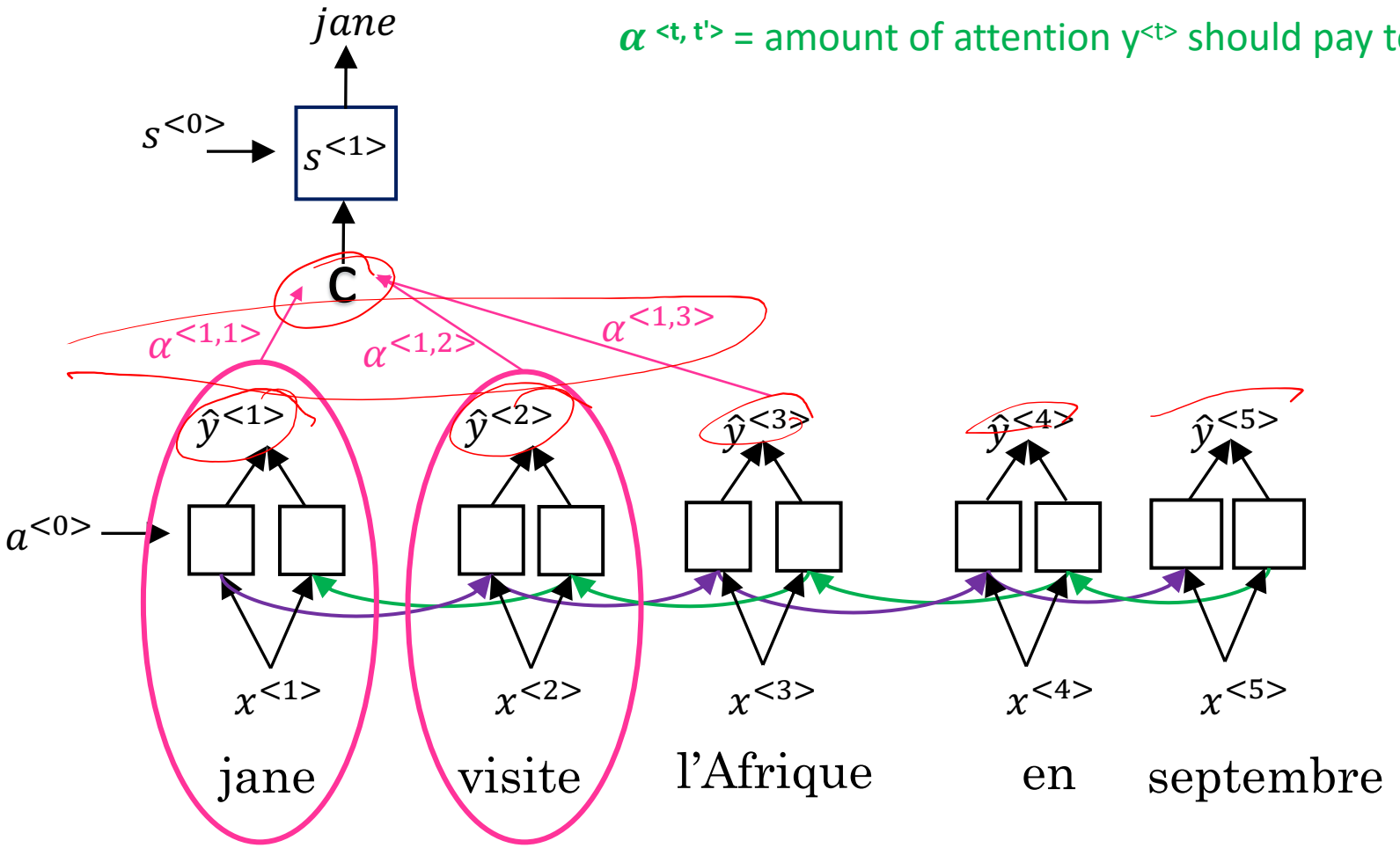
$\alpha^{<t,t'>}$ = amount of attention $y^{<t>}$ should pay to $a^{<t'>}$



[Bahdanau et. al., 2014. Neural machine translation by jointly learning to align and translate]

가중치 올라감 (영향 단어의 개수)

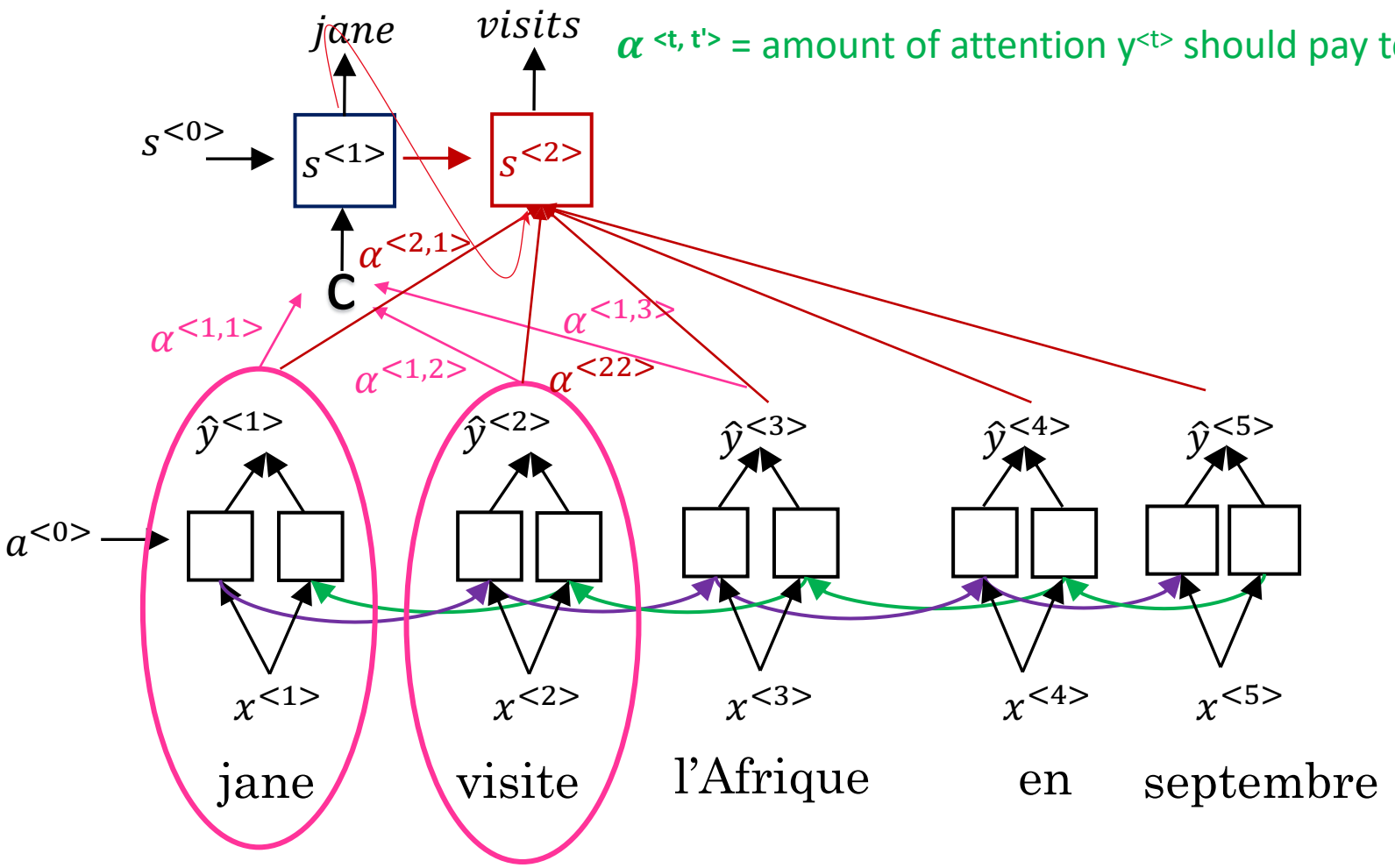
Attention Model intuition



$\alpha^{<t, t'>}$ = amount of attention $y^{<t>}$ should pay to $a^{<t'>}$

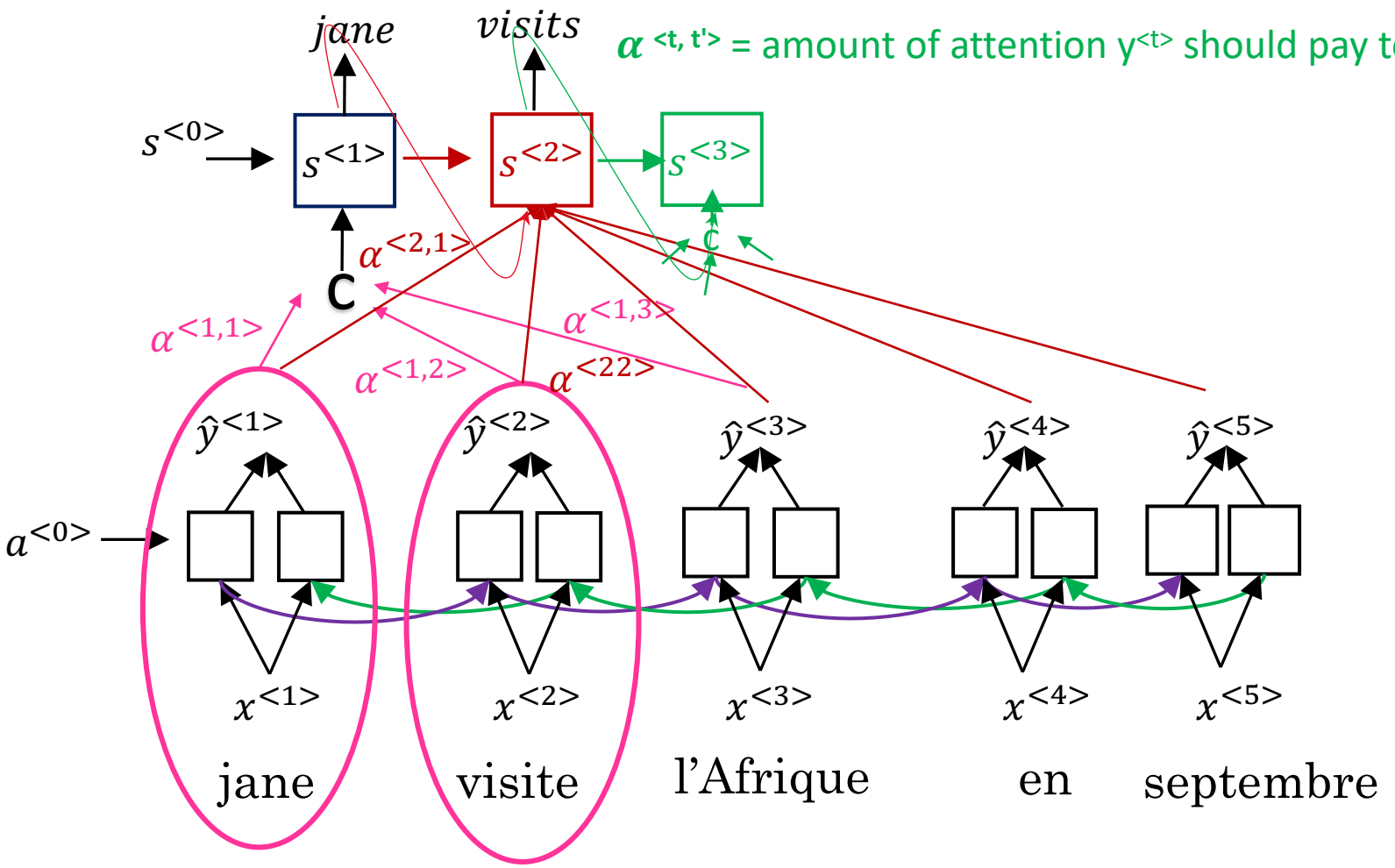
[Bahdanau et. al., 2014. Neural machine translation by jointly learning to align and translate]

Attention Model intuition



[Bahdanau et. al., 2014. Neural machine translation by jointly learning to align and translate]

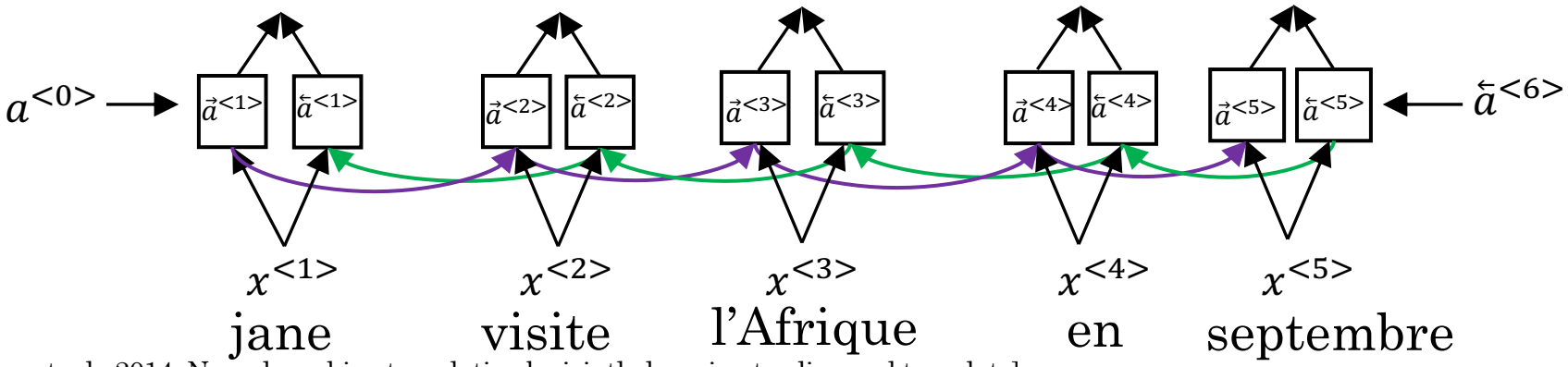
Attention Model intuition



[Bahdanau et. al., 2014. Neural machine translation by jointly learning to align and translate]

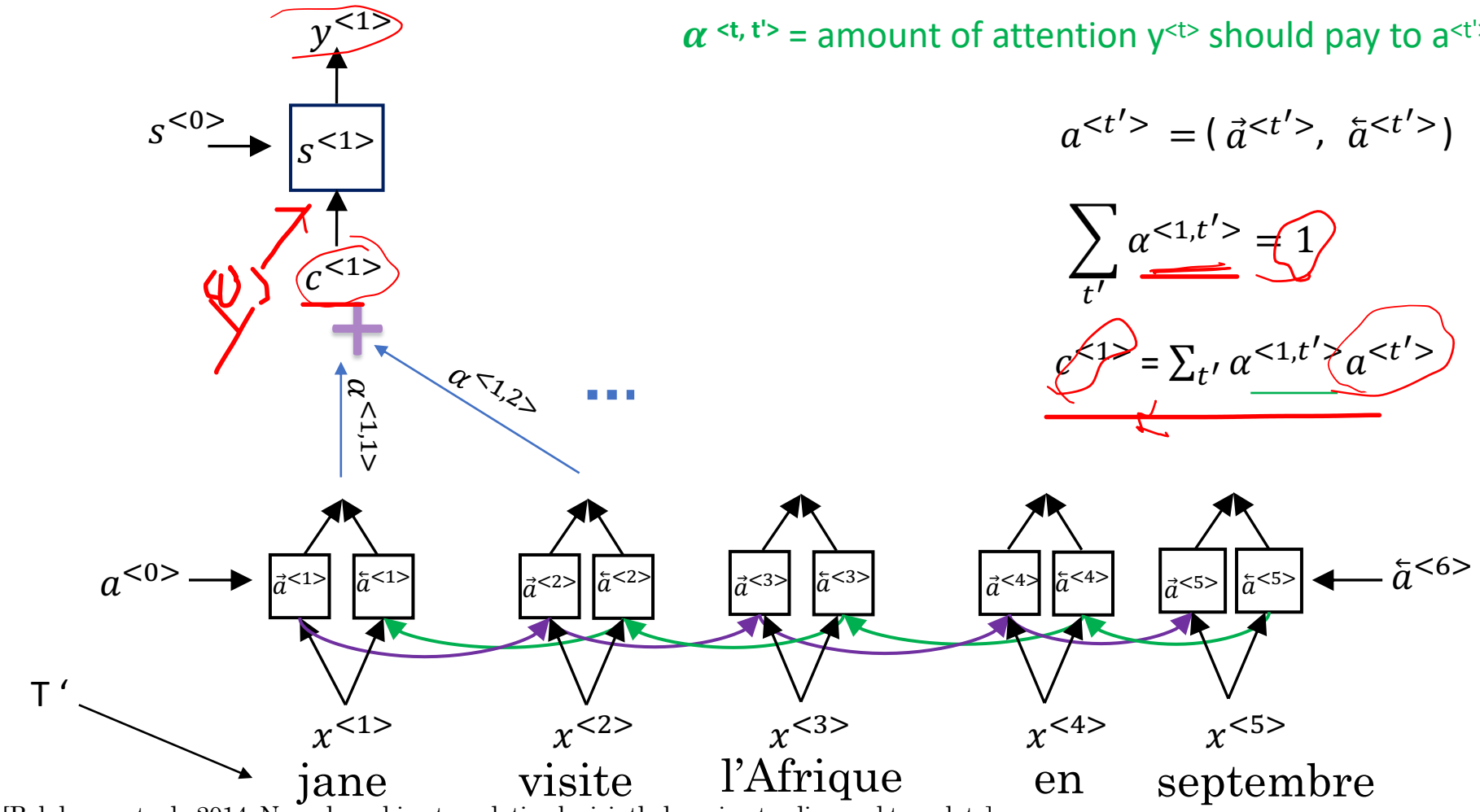
Attention Model

$$a^{<t>} = (\vec{a}^{<t>}, \overleftarrow{a}^{<t>})$$



[Bahdanau et. al., 2014. Neural machine translation by jointly learning to align and translate]

Attention Model



$\alpha^{<t,\,t'>}$ = amount of attention $y^{<t>}$ should pay to $a^{<t'>}$

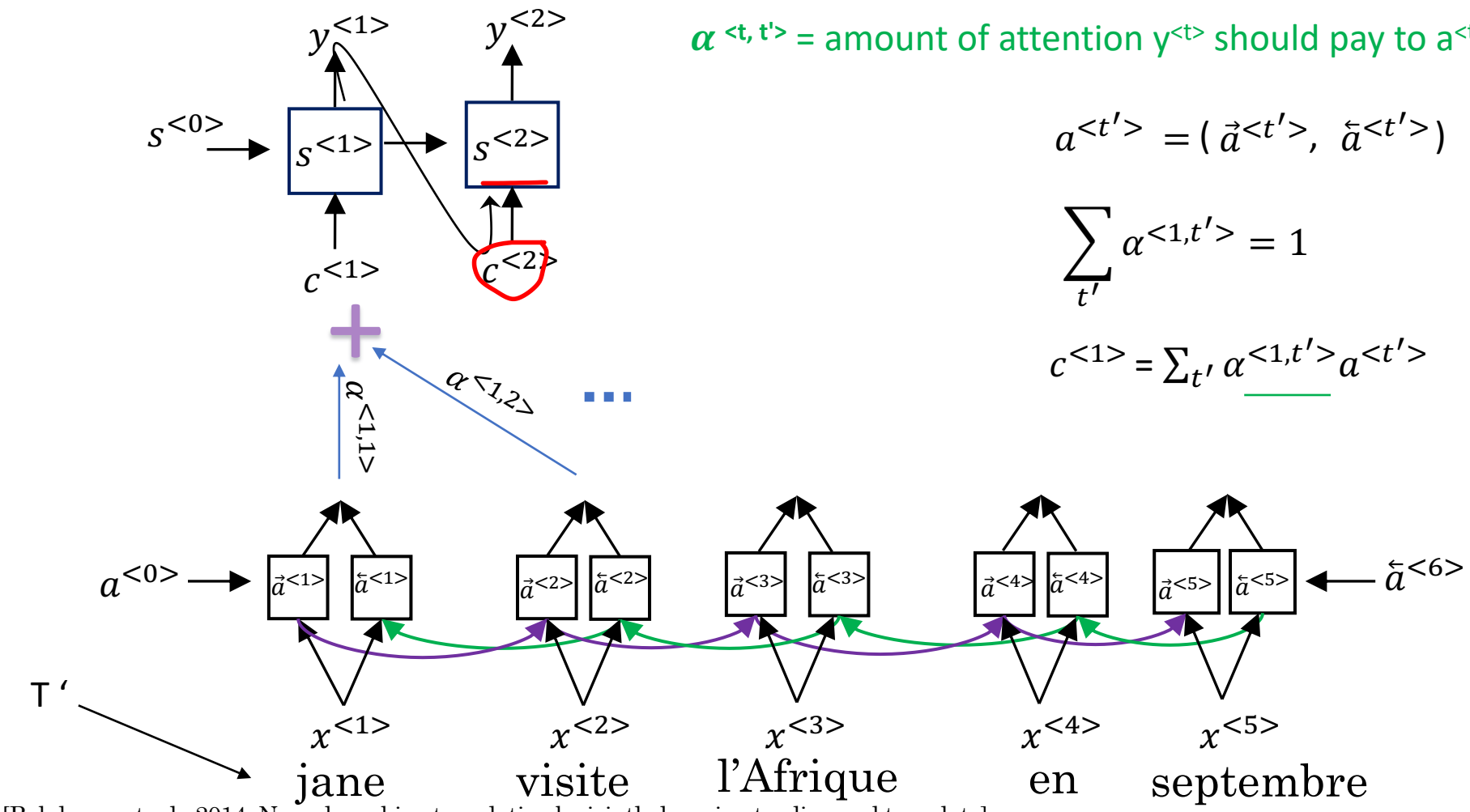
$$a^{<t'>} = (\vec{a}^{<t'>}, \overleftarrow{a}^{<t'>})$$

$$\sum_{t'} \alpha^{<1,t'>} = 1$$

$$c^{<1>} = \sum_{t'} \alpha^{<1,t'>} a^{<t'>}$$

[Bahdanau et. al., 2014. Neural machine translation by jointly learning to align and translate]

Attention Model



$\alpha^{<t,t'>}$ = amount of attention $y^{<t>}$ should pay to $a^{<t'>}$

$$a^{<t'>} = (\vec{a}^{<t'>}, \overleftarrow{a}^{<t'>})$$

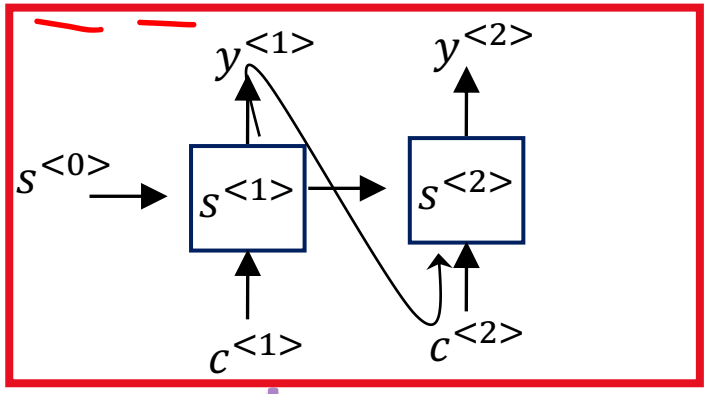
$$\sum_{t'} \alpha^{<1,t'>} = 1$$

$$c^{<1>} = \sum_{t'} \alpha^{<1,t'>} a^{<t'>}$$

[Bahdanau et. al., 2014. Neural machine translation by jointly learning to align and translate]

Attention Model

$s_t = RNNCell(s_{t-1}, y_{t-1}, c_t)$



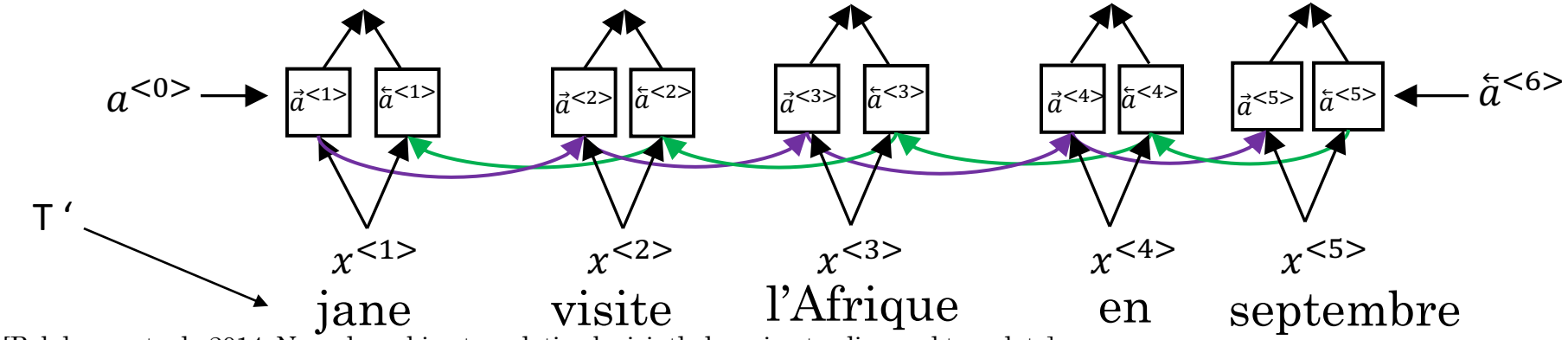
Attention decoder

$\alpha^{<t,t'>}$ = amount of attention $y^{<t>}$ should pay to $a^{<t'>}$

$$a^{<t'>} = (\vec{a}^{<t'>}, \overleftarrow{a}^{<t'>})$$

$$\sum_{t'} \alpha^{<1,t'>} = 1$$

$$c^{<1>} = \sum_{t'} \alpha^{<1,t'>} a^{<t'>}$$

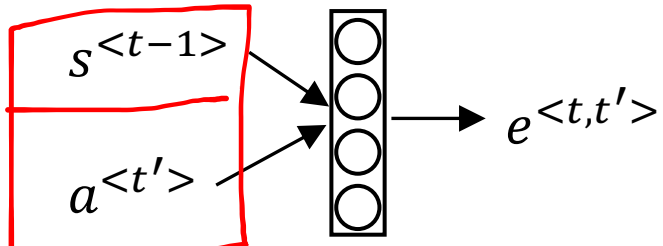


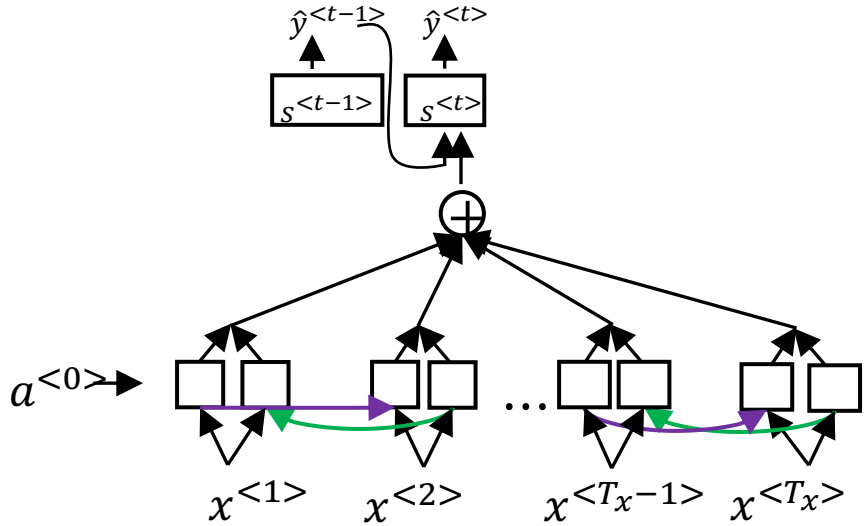
[Bahdanau et. al., 2014. Neural machine translation by jointly learning to align and translate]

Computing Attention $\alpha^{<t,t'>}$

$\alpha^{<t,t'>}$ = amount of attention $y^{<t>}$ should pay to $a^{<t'>}$

$$\alpha^{<t,t'>} = \frac{\exp(e^{<t,t'>})}{\sum_{t'=1}^{T_x} \exp(e^{<t,t'>})}$$





$$e^{<t,t'>} = v_a^T \tanh\left(W_{as} s^{<t-1>} + U_{aa} a^{<t'>}\right)$$

$$= v_a^T \tanh\left(W_a \left[s^{<t-1>}; a^{<t'>}\right]\right)$$

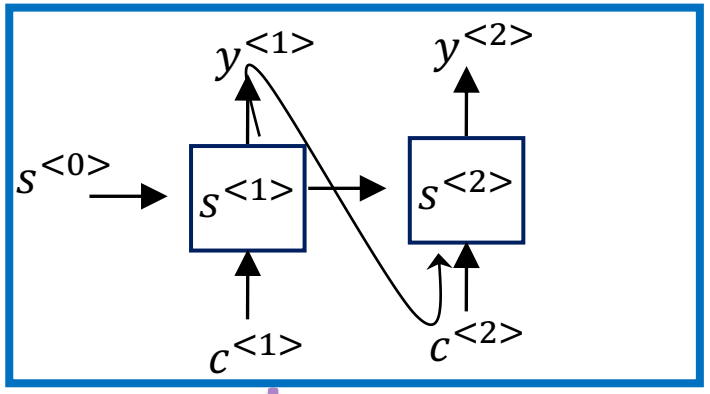
scala

레이어 1층

[Bahdanau et. al., 2014. Neural machine translation by jointly learning to align and translate]

Attention Model

Attention decoder



$$e^{<t,t'>} = v_a^T \tanh(W_a[s^{<t-1>}; a^{<t'>}])$$

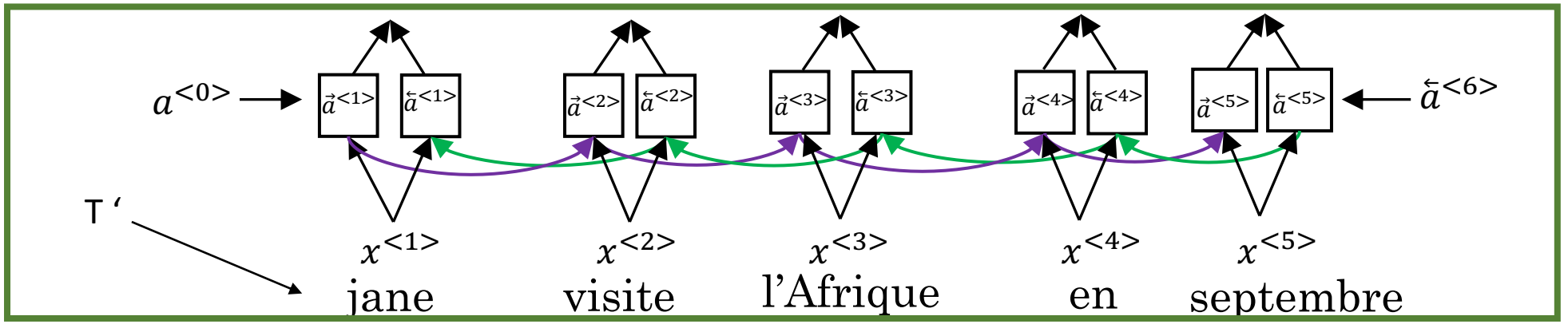
$$\alpha^{<t,t'>} = \frac{\exp(e^{<t,t'>})}{\sum_{t'=1}^{T_x} \exp(e^{<t,t'>})}$$

$$c^{<t>} = \sum_{t'} \alpha^{<t,t'>} a^{<t'>}$$

$$s_t = RNNCell(s_{t-1}, y_{t-1}, c_t)$$

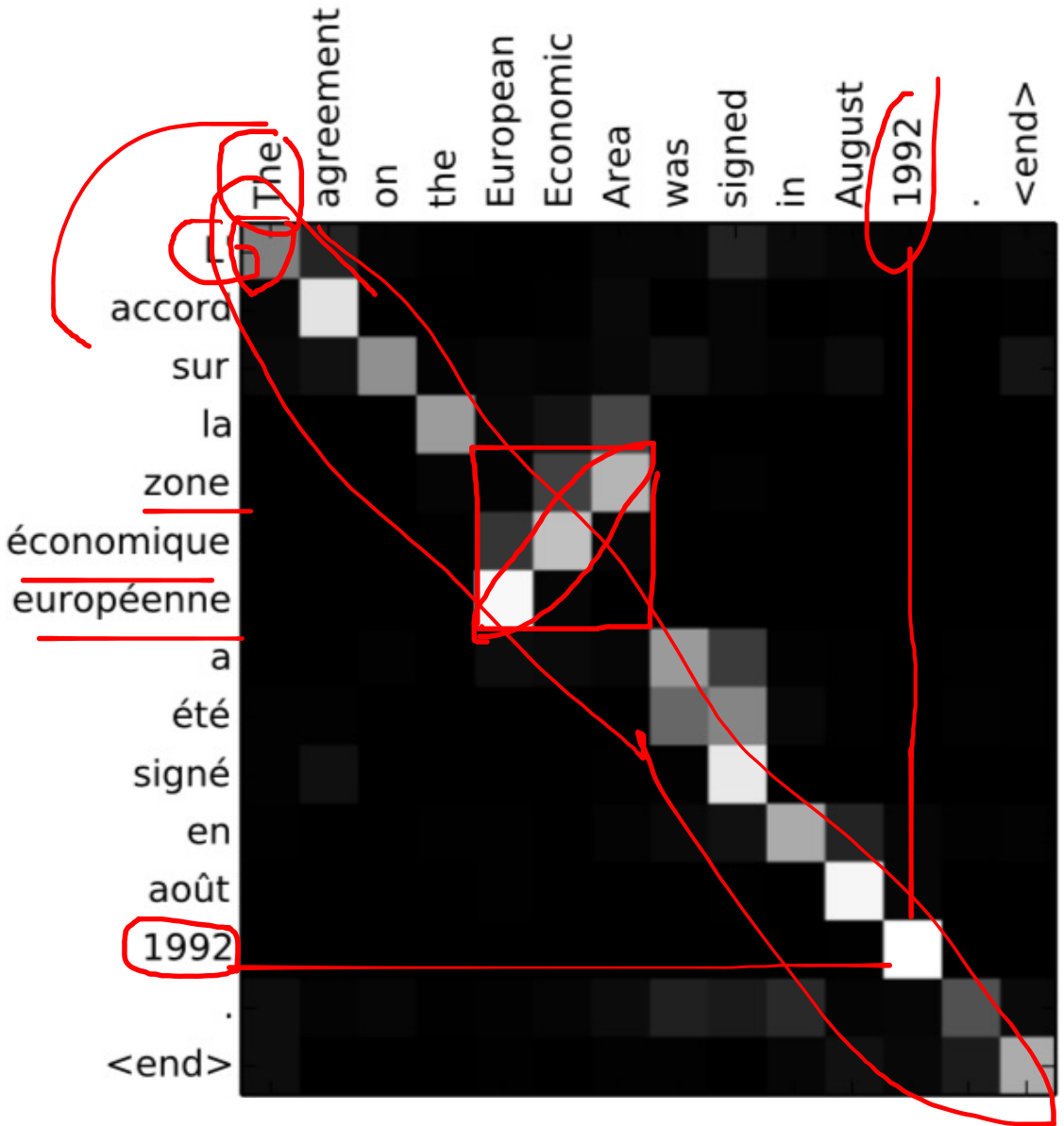
$\alpha^{<1,1>}$ $\alpha^{<1,2>}$ ...

(Bidirectional) encoder $a^{<t'>} = (\vec{a}^{<t'>}, \overleftarrow{a}^{<t'>})$



[Bahdanau et. al., 2014. Neural machine translation by jointly learning to align and translate]

Attention


The
agreement
on
the
European
Economic
Area
was
signed
in
August
1992
.
<end>
L'
accord
sur
la
zone
économique
européenne
a
été
signé
en
août
1992
.
<end>

References

- Sequence to sequence models, Alireza Akhavan Pour
- https://d2l.ai/chapter_recurrent-modern/beam-search.html