

Natural Language Understanding

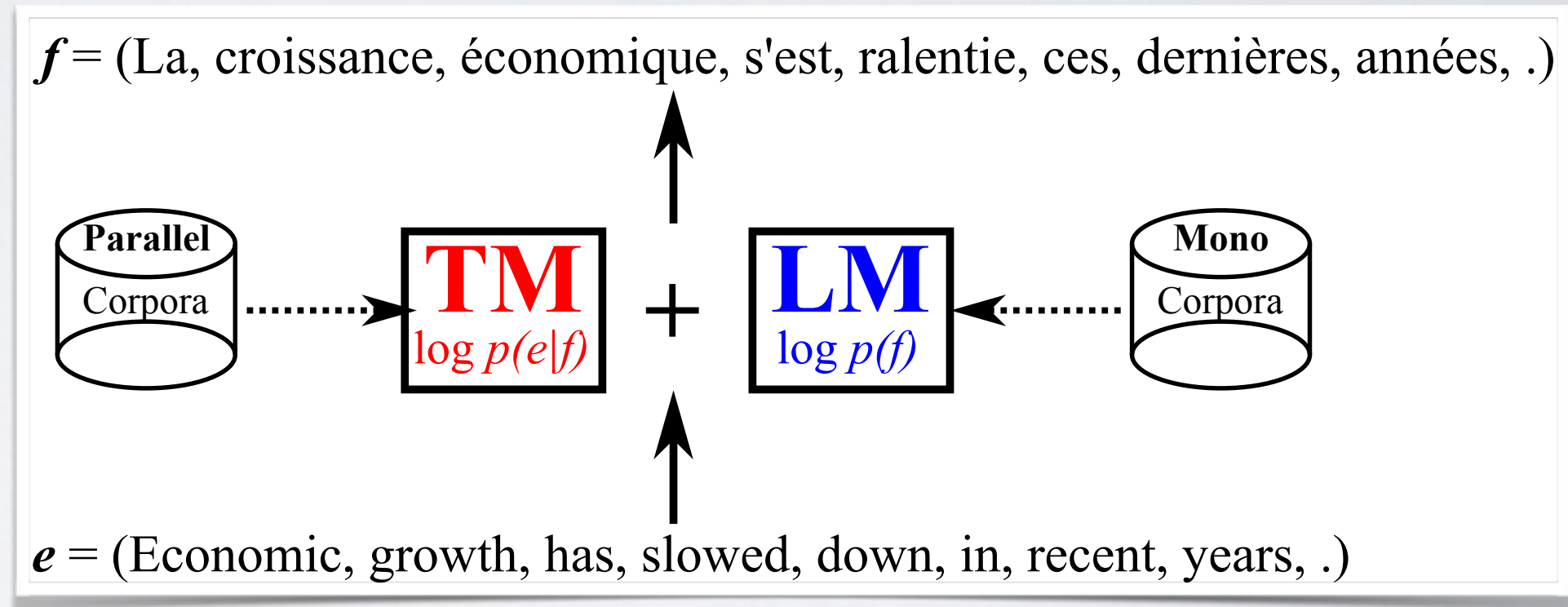
Kyunghyun Cho, NYU & U. Montreal

Machine Translation

NEURAL MACHINE TRANSLATION

Topics: Statistical Machine Translation

- $\log p(f|e) = \log p(e|f) + \log p(f)$
 - Translation model: $\log p(e|f)$
 - Fit it with parallel corpora
 - Language model: $\log p(f)$
 - Fit it with monolingual corpora

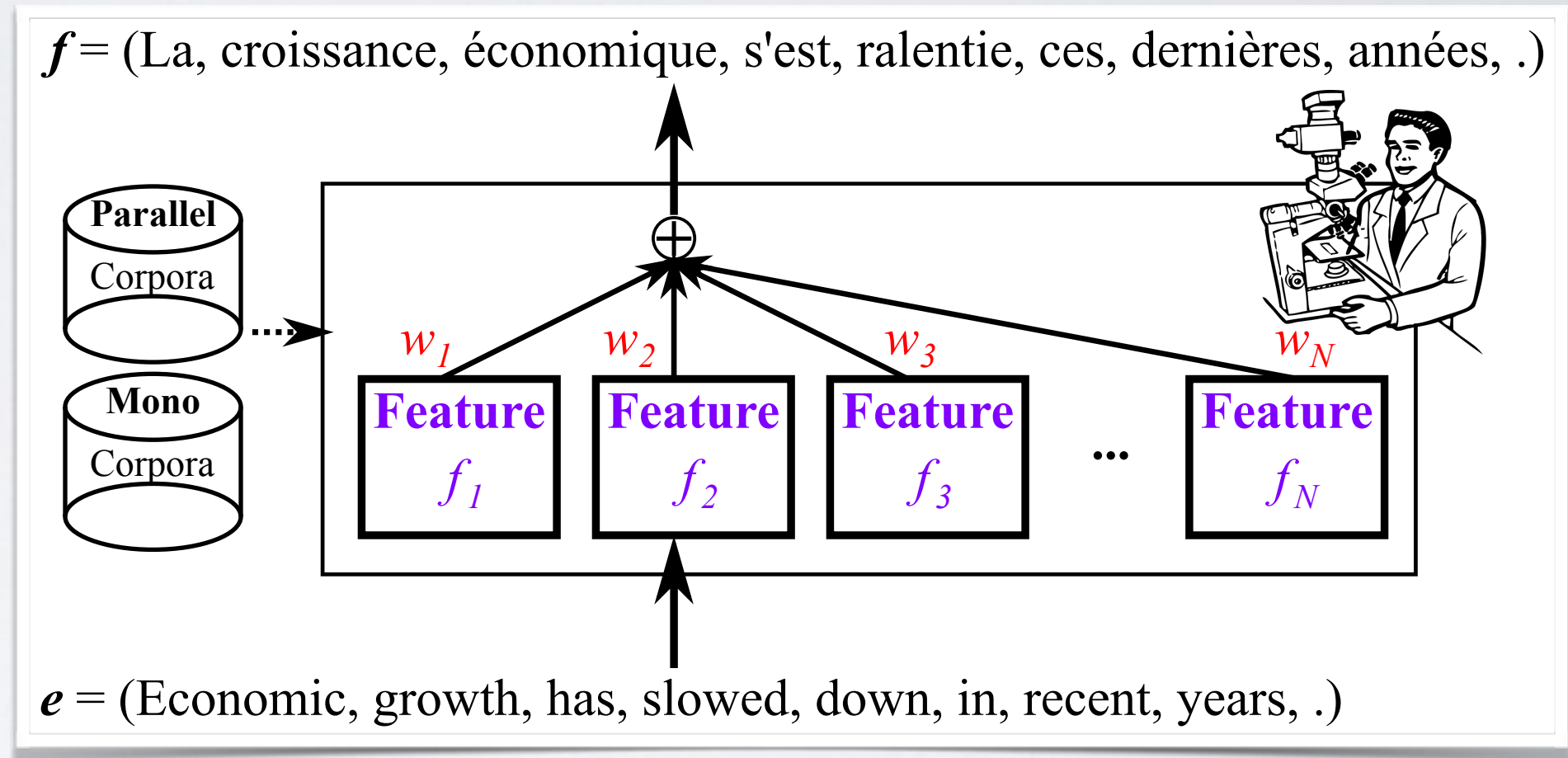


- The whole task $\log p(f|e)$ is **conditional language modelling**.

NEURAL MACHINE TRANSLATION

Topics: Statistical Machine Translation - In Reality

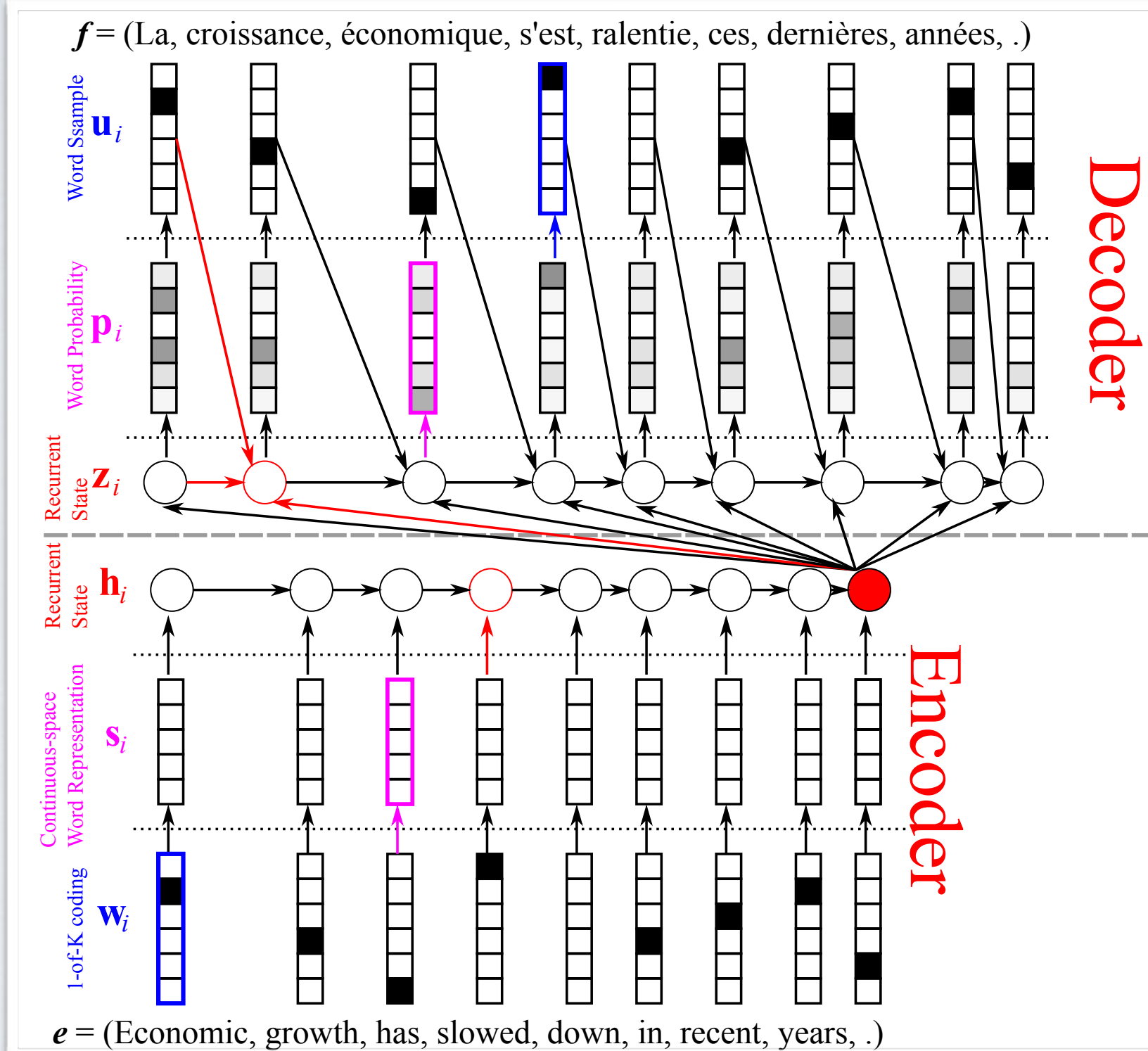
- $\log p(f|e) \approx \sum_{n=1}^N f_n(e, f) + C$
 - Log-linear model
 - Feature function $f_n(e, f)$
 - Count-based or linguistics-based
 - Learned from corpora
- Steps:
 - (1) Experts engineer *useful* features
 - (2) Use a simple log-linear model
 - (3) Use a strong, external language model



NEURAL MACHINE TRANSLATION

Topics: Sequence-to-Sequence Learning

(Forcada&Ñeco, 1997;
Kalchbrenner&Blunsom, 2013;
Sutskever et al., 2014;
Cho et al., 2014)



NEURAL MACHINE TRANSLATION

Topics: Sequence-to-Sequence Learning — Encoder

- Encoder

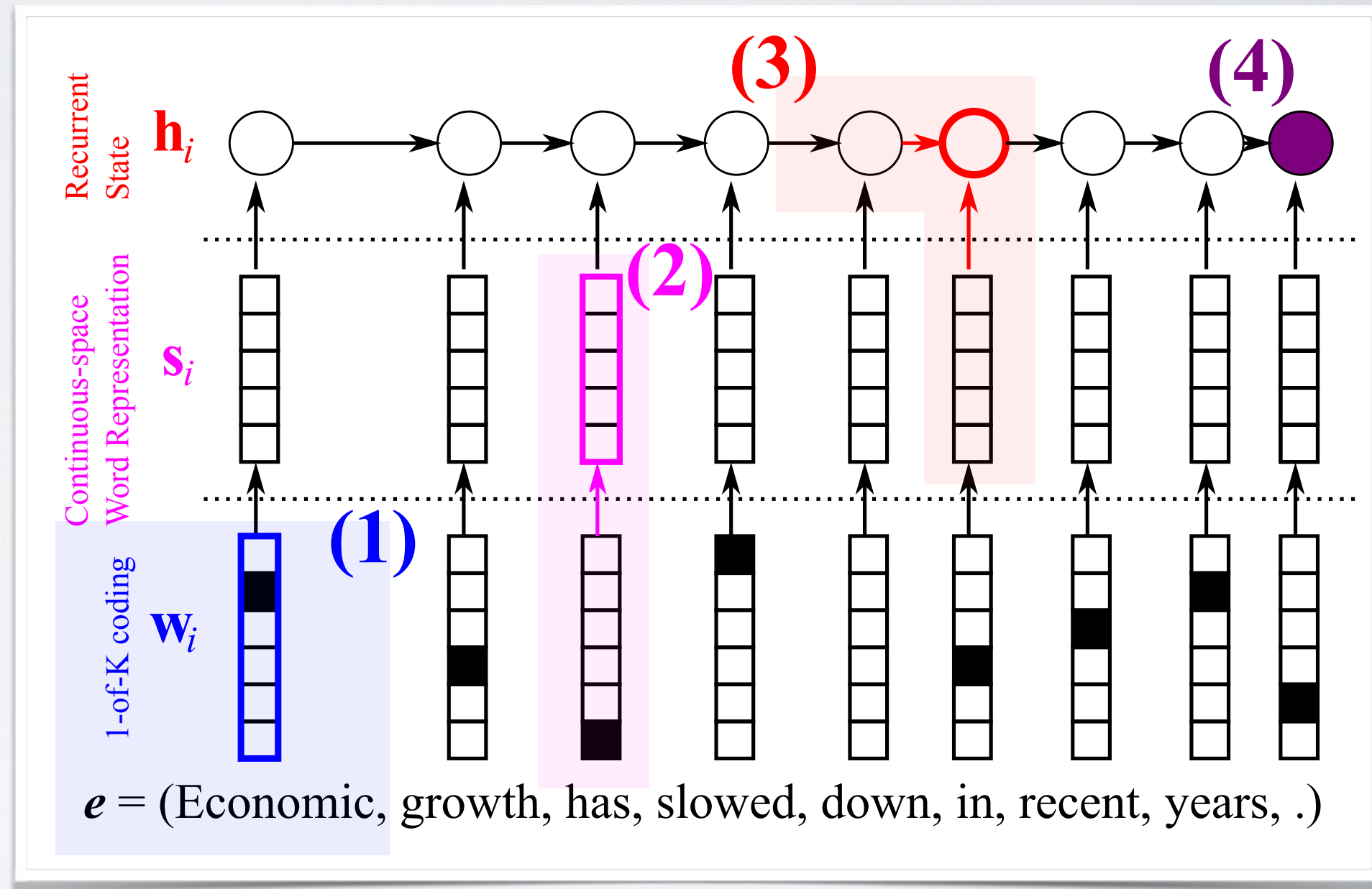
(1) 1-of-K coding of *source* words

(2) Continuous-space representation

$$s_{t'} = W^T x_{t'}, \text{ where } W \in \mathbb{R}^{|V| \times d}$$

(3) Recursively read words

$$h_t = f(h_{t-1}, s_t), \text{ for } t = 1, \dots, T$$



NEURAL MACHINE TRANSLATION

Topics: Sequence-to-Sequence Learning — Decoder

- Decoder

(1) Recursively update the memory

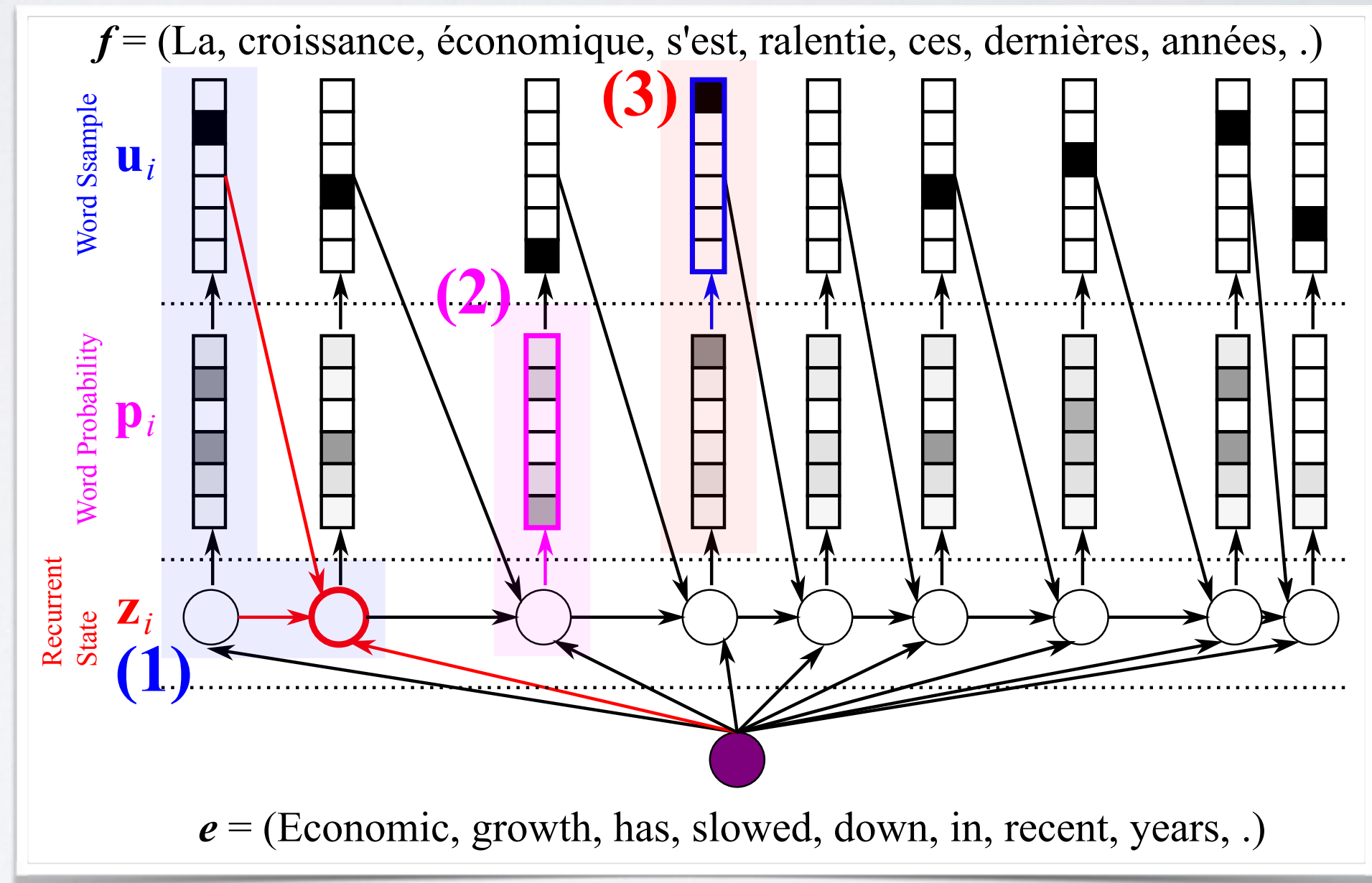
$$z_{t'} = f(z_{t'-1}, u_{t'-1}, h_T)$$

(2) Compute the next word prob.

$$p(u_{t'} | u_{<t'}) \propto \exp(R_{u_{t'}}^\top z_{t'} + b_{u_{t'}})$$

(3) Sample a next word

- *Beam search is a good idea*

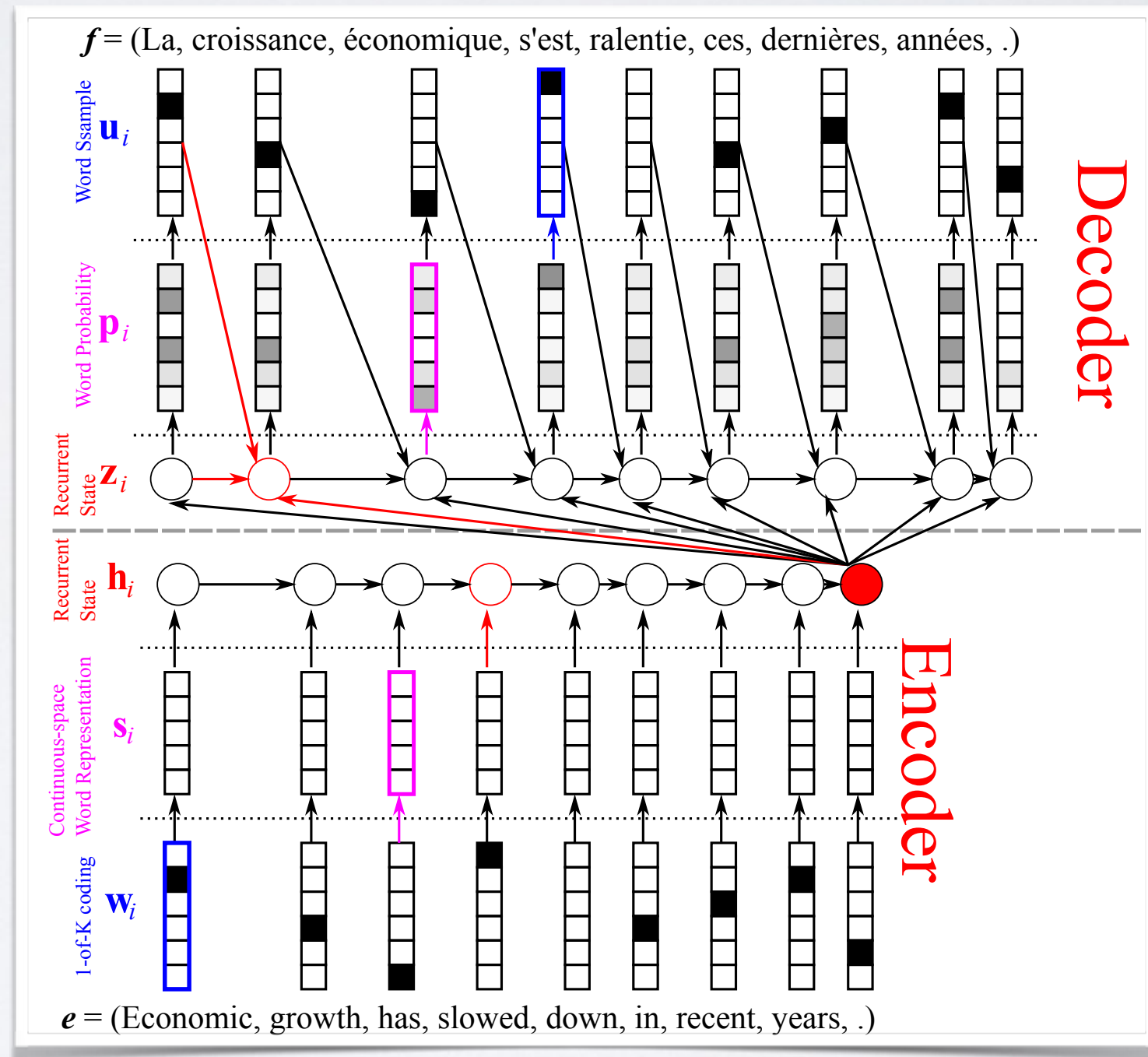


NEURAL MACHINE TRANSLATION

Topics: Sequence-to-Sequence Learning — Issue

- This is quite an unrealistic model.
- *Why?*

“You can’t cram the meaning of a whole sentence into a single vector!” Ray Mooney



NEURAL MACHINE TRANSLATION

Topics: Attention-based Model

- Encoder: Bidirectional RNN

- A set of *annotation* vectors

$$\{h_1, h_2, \dots, h_T\}$$

- Attention-based Decoder

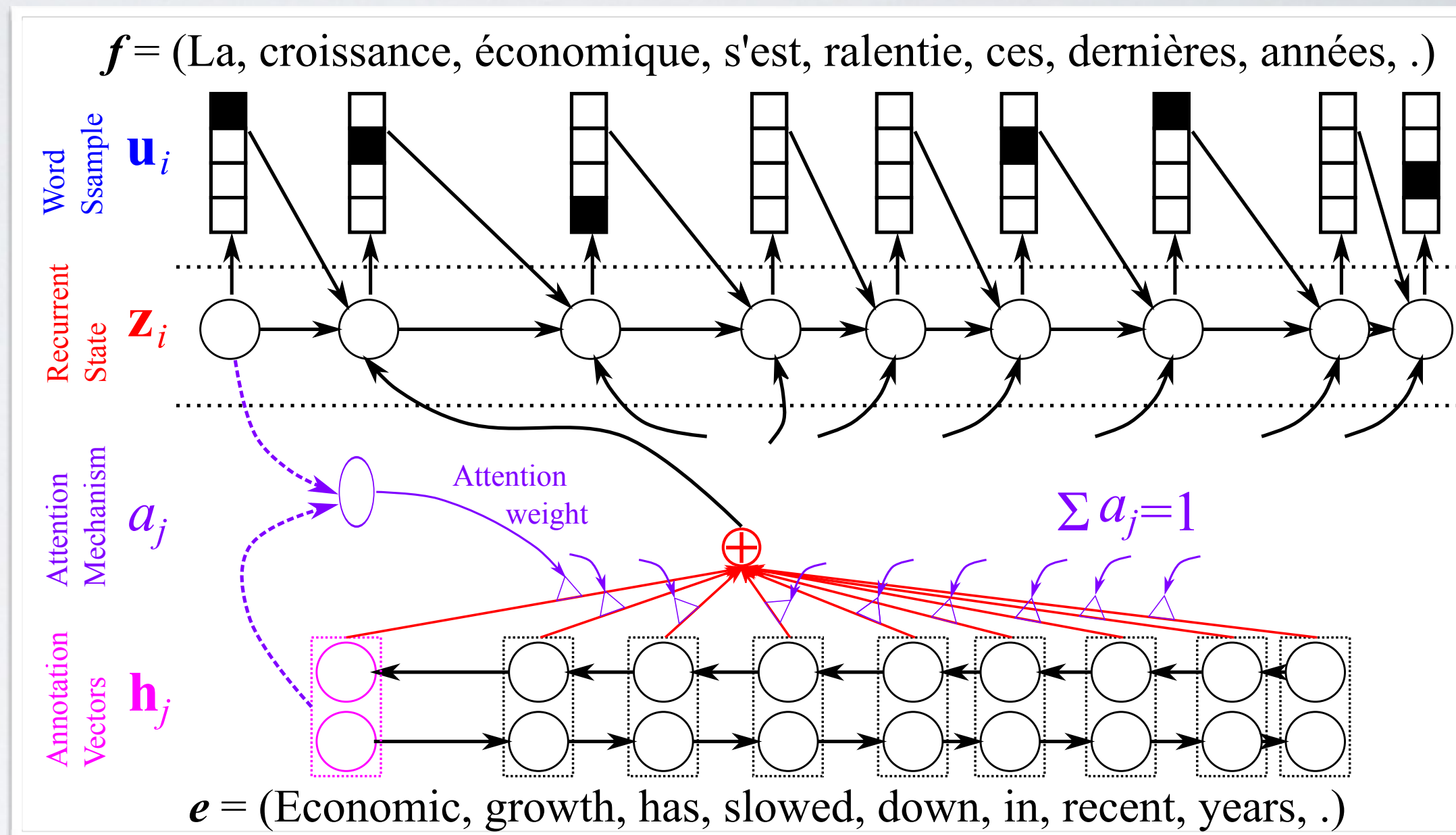
- Compute attention weights

$$\alpha_{t',t} \propto \exp(e(z_{t'-1}, u_{t'-1}, h_t))$$

- Weighted-sum of the annotation vectors

$$c_{t'} = \sum_{t=1}^T \alpha_{t',t} h_t$$

- Use $c_{t'}$ instead of h_T



NEURAL MACHINE TRANSLATION

Topics: Attention-based Model

- Encoder: Bidirectional RNN

- A set of *annotation* vectors

$$\{h_1, h_2, \dots, h_T\}$$

- Attention-based Decoder

- Compute attention weights

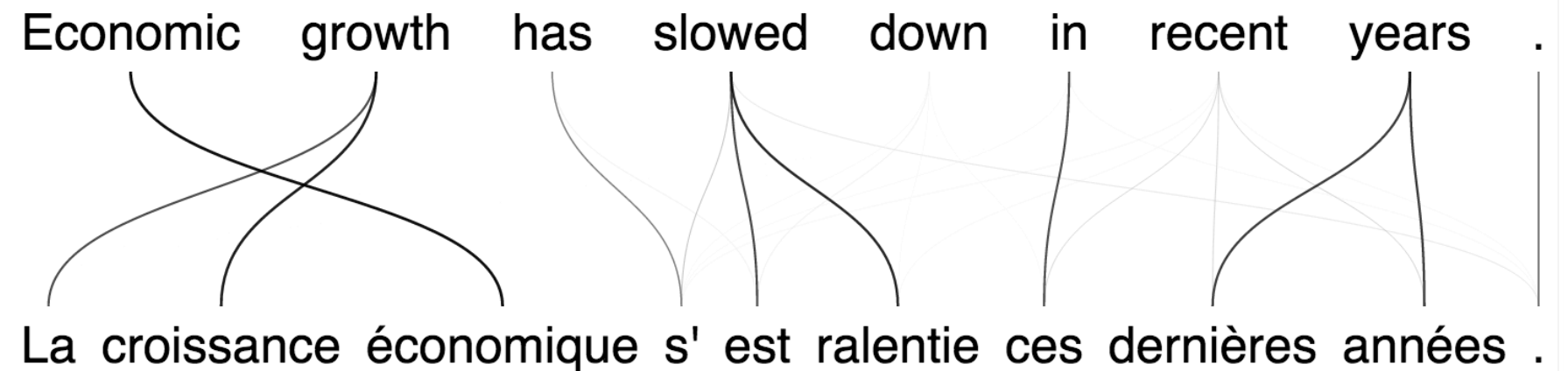
$$\alpha_{t',t} \propto \exp(e(z_{t'-1}, u_{t'-1}, h_t))$$

- Weighted-sum of the annotation vectors

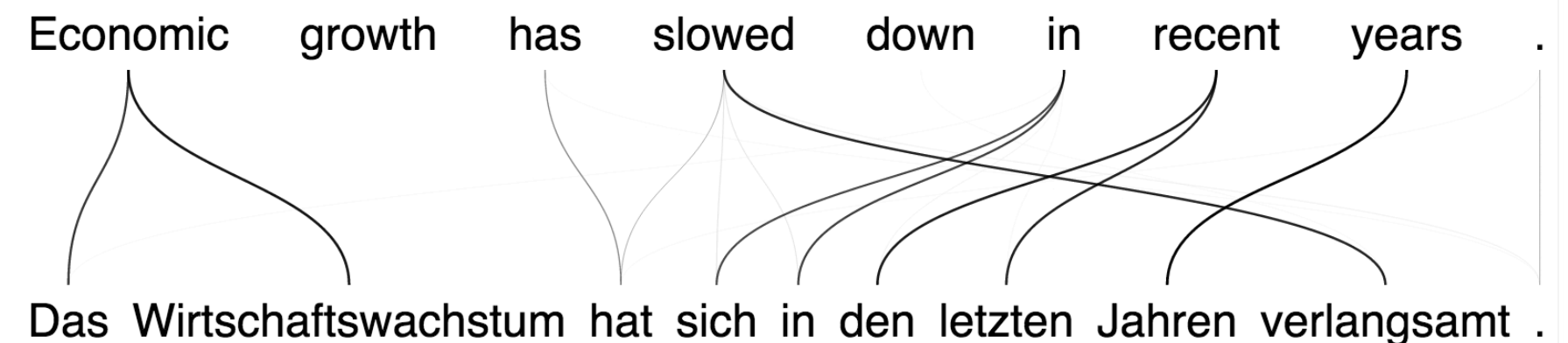
$$c_{t'} = \sum_{t=1}^T \alpha_{t',t} h_t$$

- Use $c_{t'}$ instead of h_T

English-French



English-German



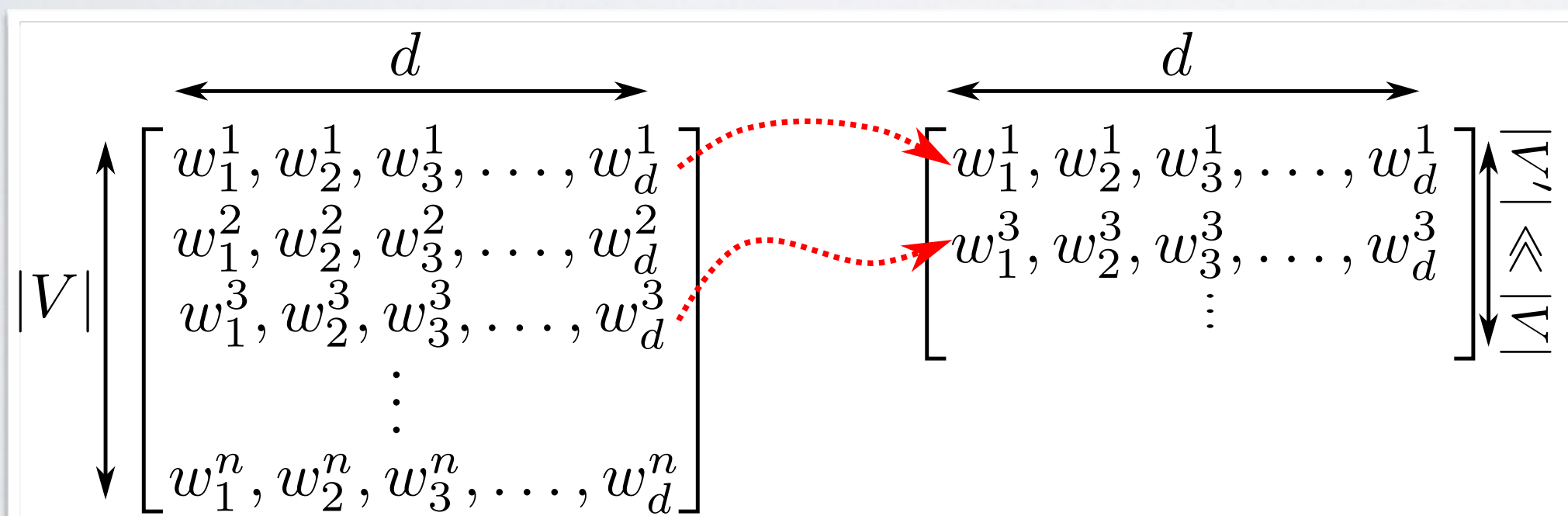
NEURAL MACHINE TRANSLATION

Topics: Few tricks for neural machine translation

- Very large target vocabulary (Jean et al., 2015)

$$p(y_t \mid y_{<t}, x) = \frac{\exp \{w_t^\top \phi(y_{t-1}, z_t, c_t)\}}{\sum_{k: \mathbf{y}_k \in \mathbf{V}} \exp \{w_k^\top \phi(y_{t-1}, z_t, c_t)\}}$$

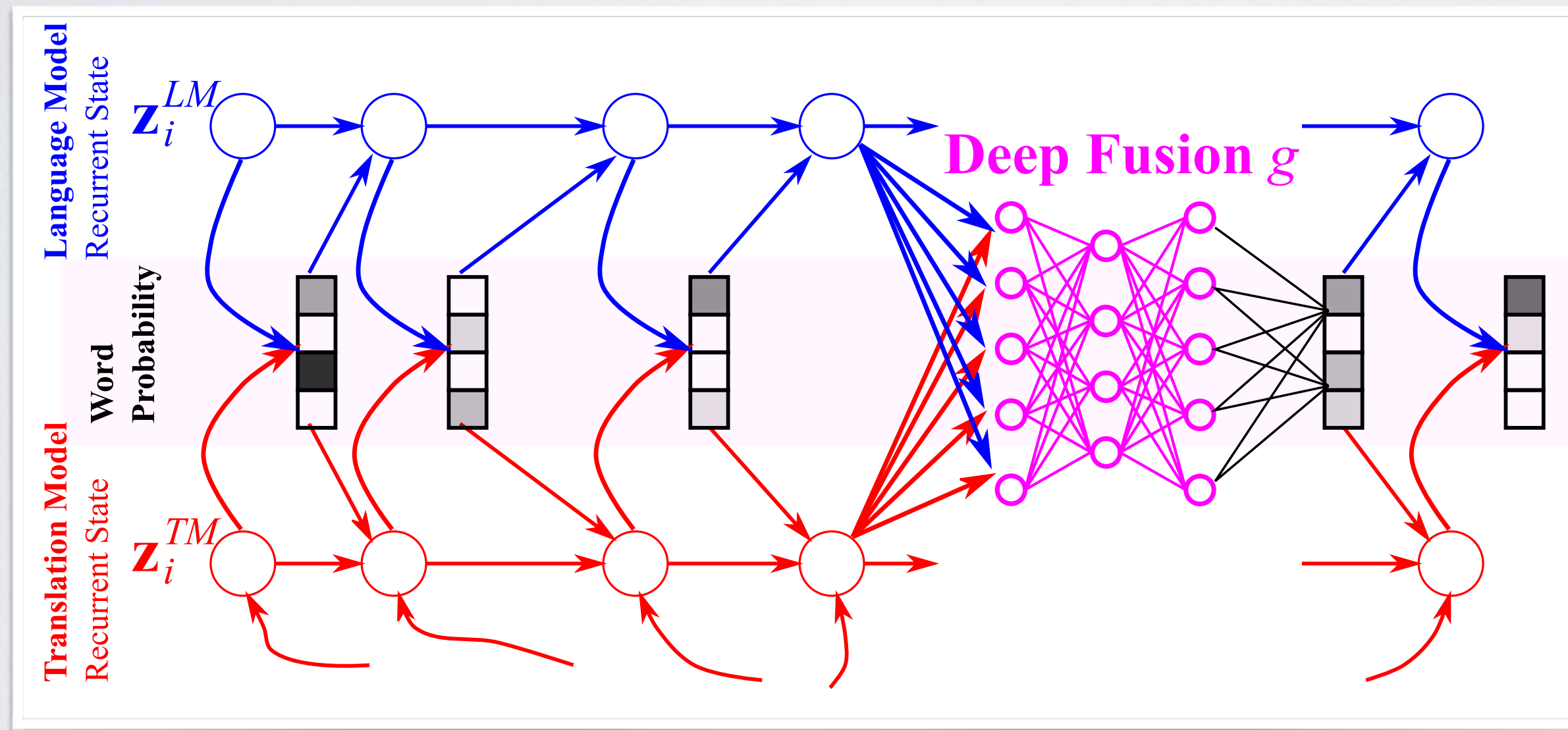
$$\approx \frac{\exp \{w_t^\top \phi(y_{t-1}, z_t, c_t)\}}{\sum_{k: \mathbf{y}_k \in \mathbf{V}'} \exp \{w_k^\top \phi(y_{t-1}, z_t, c_t)\}}$$



NEURAL MACHINE TRANSLATION

Topics: Few tricks for neural machine translation

- Deep Fusion of Target Language Model (Gulcehre&Firat et al., 2015)



$$p(y_t | y_{<t}, x) \propto \exp(y_t^\top (W_o f_{o,\theta}(z_t^{LM}, g_t \cdot z_t^{TM}, y_{t-1}, c_t) + b_o))$$

*Attention-based **neural machine translation***
*is **comparable to***
*phrase-based **statistical machine translation***

Teaching Machines to Read, Comprehend and Answer

Based on (Hermann et al., 2015; Blunsom, 2015)

READING COMPREHENSION

Topics: Teaching machines to read and comprehend

CNN article:

Document The BBC producer allegedly struck by Jeremy Clarkson will not press charges against the “Top Gear” host, his lawyer said Friday. Clarkson, who hosted one of the most-watched television shows in the world, was dropped by the BBC Wednesday after an internal investigation by the British broadcaster found he had subjected producer Oisin Tymon “to an unprovoked physical and verbal attack.” ...

Query Producer **X** will not press charges against Jeremy Clarkson, his lawyer says.

Answer Oisin Tymon

READING COMPREHENSION

Topics: Teaching machines to read and comprehend

— *Deep LSTM Reader*

- Document Reader

$$h_t = f(h_{t-1}, w_t), \text{ for all } t = 1, \dots, T$$

- Summary of the document: h_T

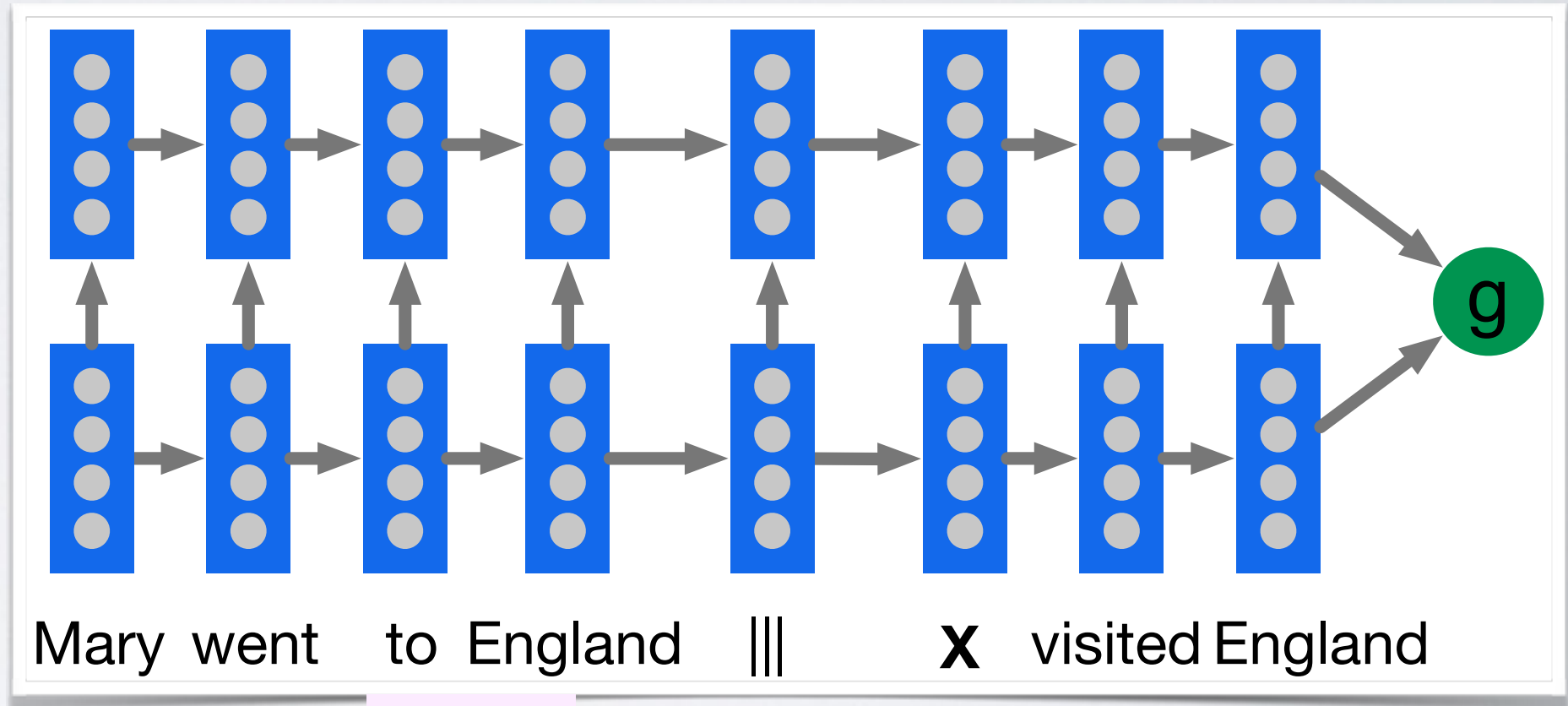
- Query Reader

$$z_t = f(z_{t-1}, w'_t), \text{ for all } t = 1, \dots, T'$$

- Summary of the query: $z_{T'}$

- Answer selection

$$p(a | \{w_t\}_{t=1}^T, \{w_{t'}\}_{t'=1}^{T'}) = g_a(h_T, z_{T'})$$



No!!!

READING COMPREHENSION

Topics: Teaching machines to read and comprehend

— *Attentive Reader*

- Document Reader: BiRNN

- Annotation vectors: $\{h_1, h_2, \dots, h_T\}$

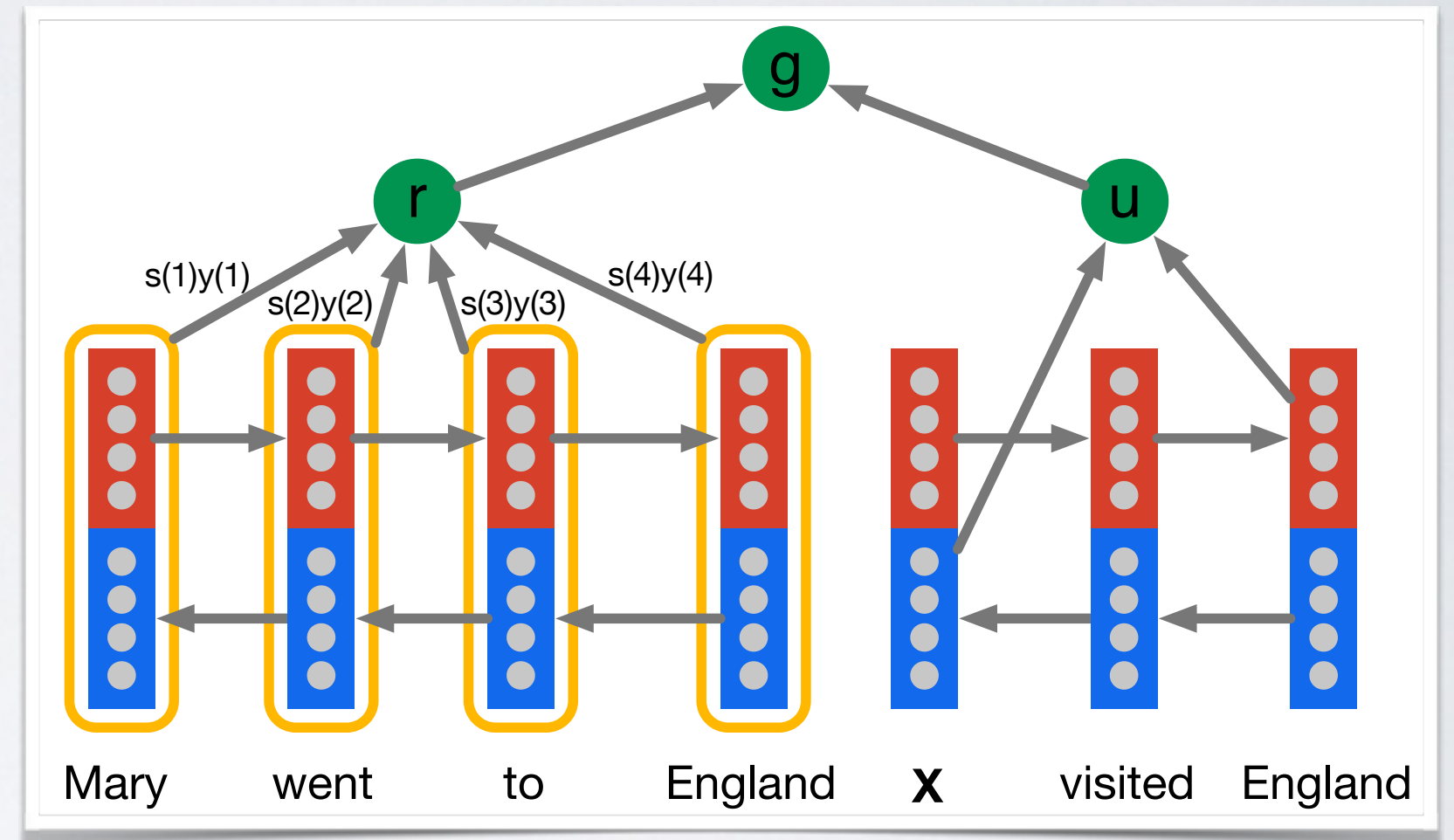
- Query Reader: $z_{T'}$

- Answer selection

- Attention mechanism $\alpha_t \propto e(h_t, z_{T'})$

- Query-dependent document summary $c = \sum_{t=1}^T \alpha_t h_t$

- Answer selection: $p(a | \{w_t\}_{t=1}^T, \{w_{t'}\}_{t'=1}^{T'}) = g_a(z_{T'}, c)$



READING COMPREHENSION

Topics: Teaching machines to read and comprehend

— *Attentive Reader* (Examples)

- Visualize the attention

by *ent40* , *ent62* correspondent updated 9:49 pm et , thu march 19 , 2015 (*ent62*) a *ent88* was killed in a parachute accident in *ent87* , *ent28* , near *ent66* , a *ent47* official told *ent62* on wednesday . he was identified thursday as special warfare operator 3rd class *ent49* , 29 , of *ent44* , *ent13* . `` *ent49* distinguished himself consistently throughout his career . he was the epitome of the quiet professional in all facets of his life , and he leaves an inspiring legacy of natural tenacity and focused commitment for posterity , " the *ent47* said in a news release . *ent49* joined the seals in september after enlisting in the *ent47* two years earlier . he was married , the *ent47* said . initial indications are the parachute failed to open during a jump as part of a training exercise . *ent49* was part of a *ent57* - based *ent88* team .

ent47 identifies deceased sailor as **X** , who leaves behind a wife

Going beyond Natural Languages

Is a human language special?

BEYOND NATURAL LANGUAGES

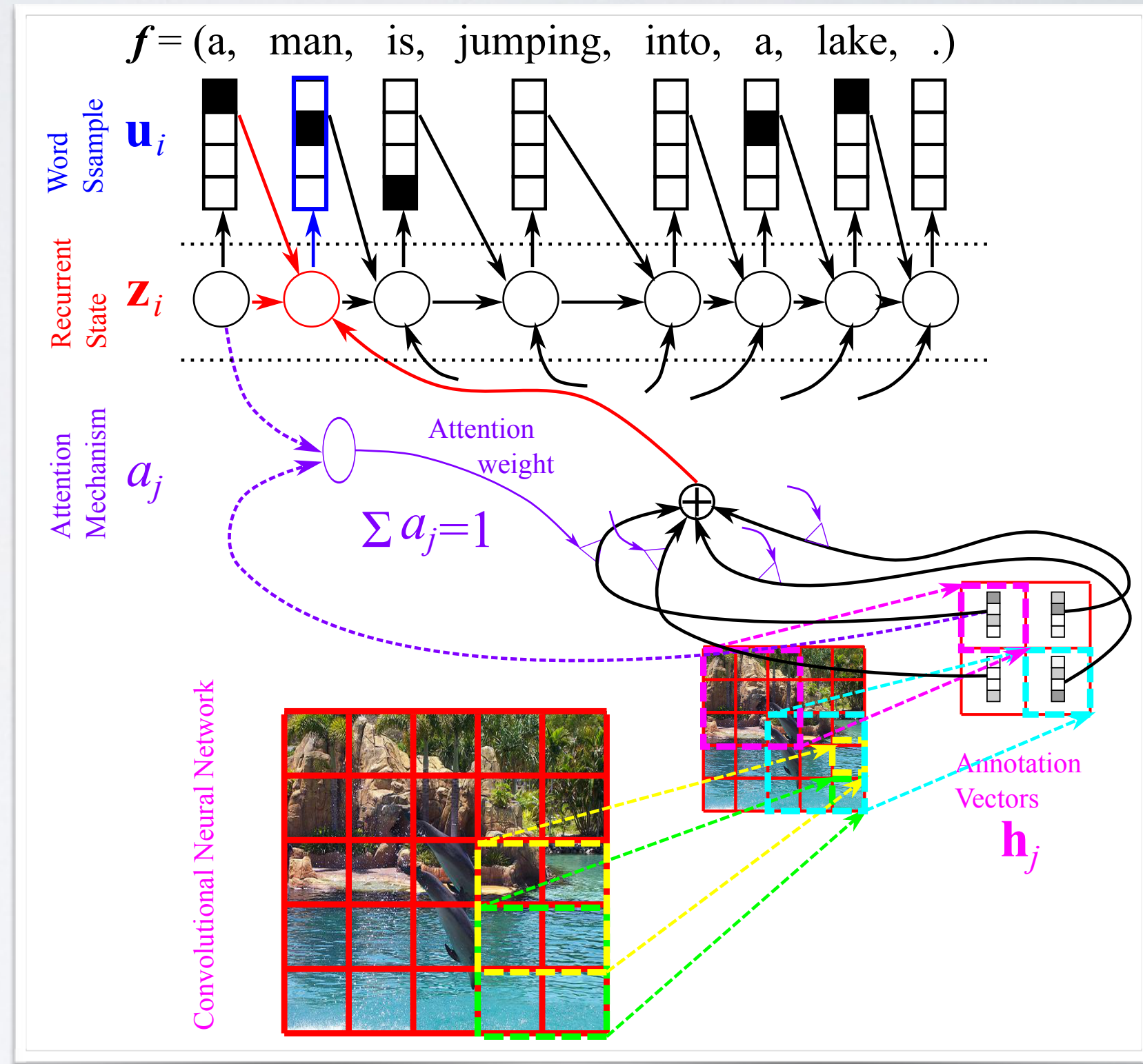
Topics: Beyond Natural Languages — Image Caption Generation

- Task: *conditional* language modelling

$p(\text{Two, dolphins, are, diving} | \text{Image}) = ?$

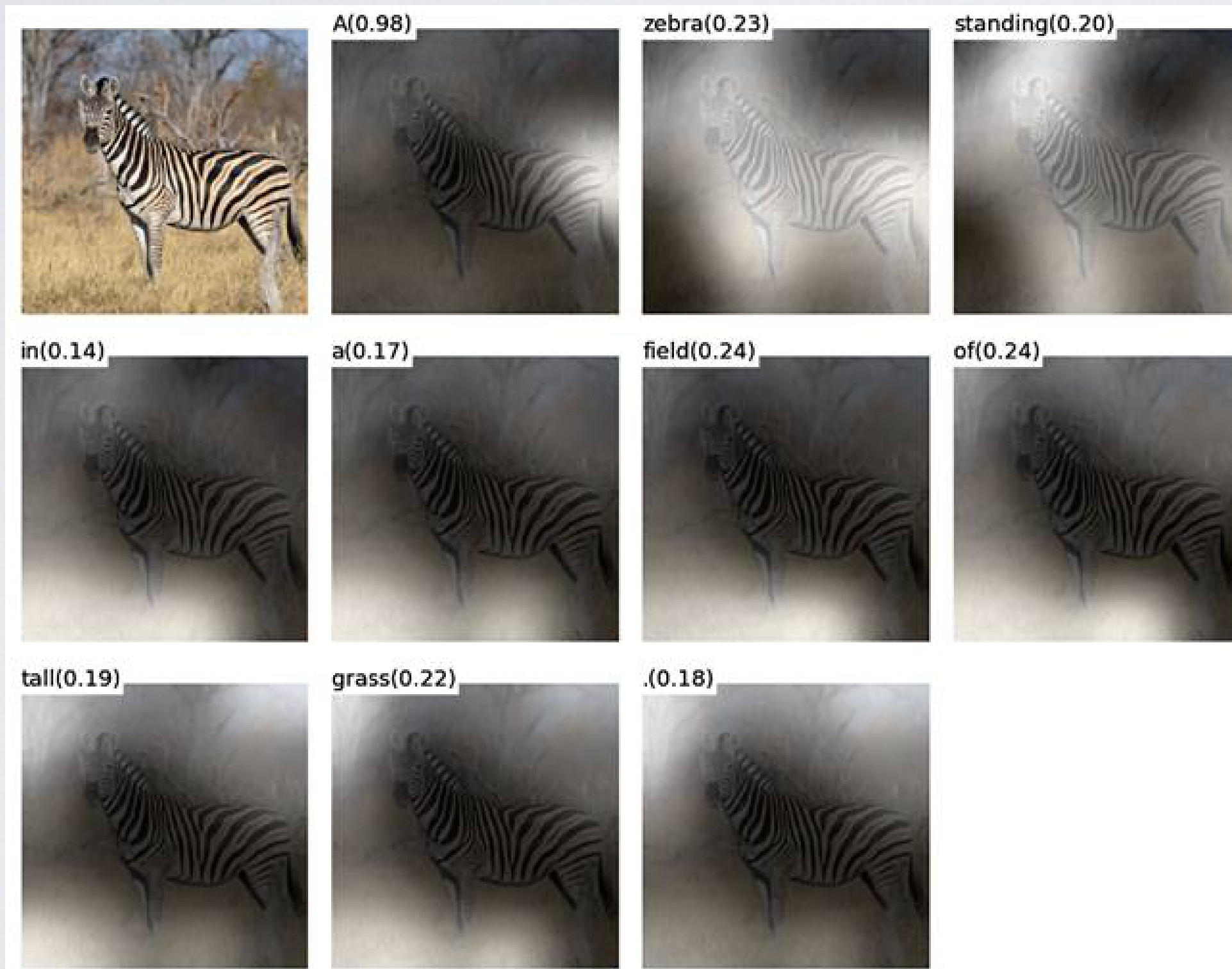


- Encoder: convolutional network
 - Pretrained as a classifier or autoencoder
- Decoder: recurrent neural network
 - RNN Language model
- With attention mechanism (Xu et al., 2015)



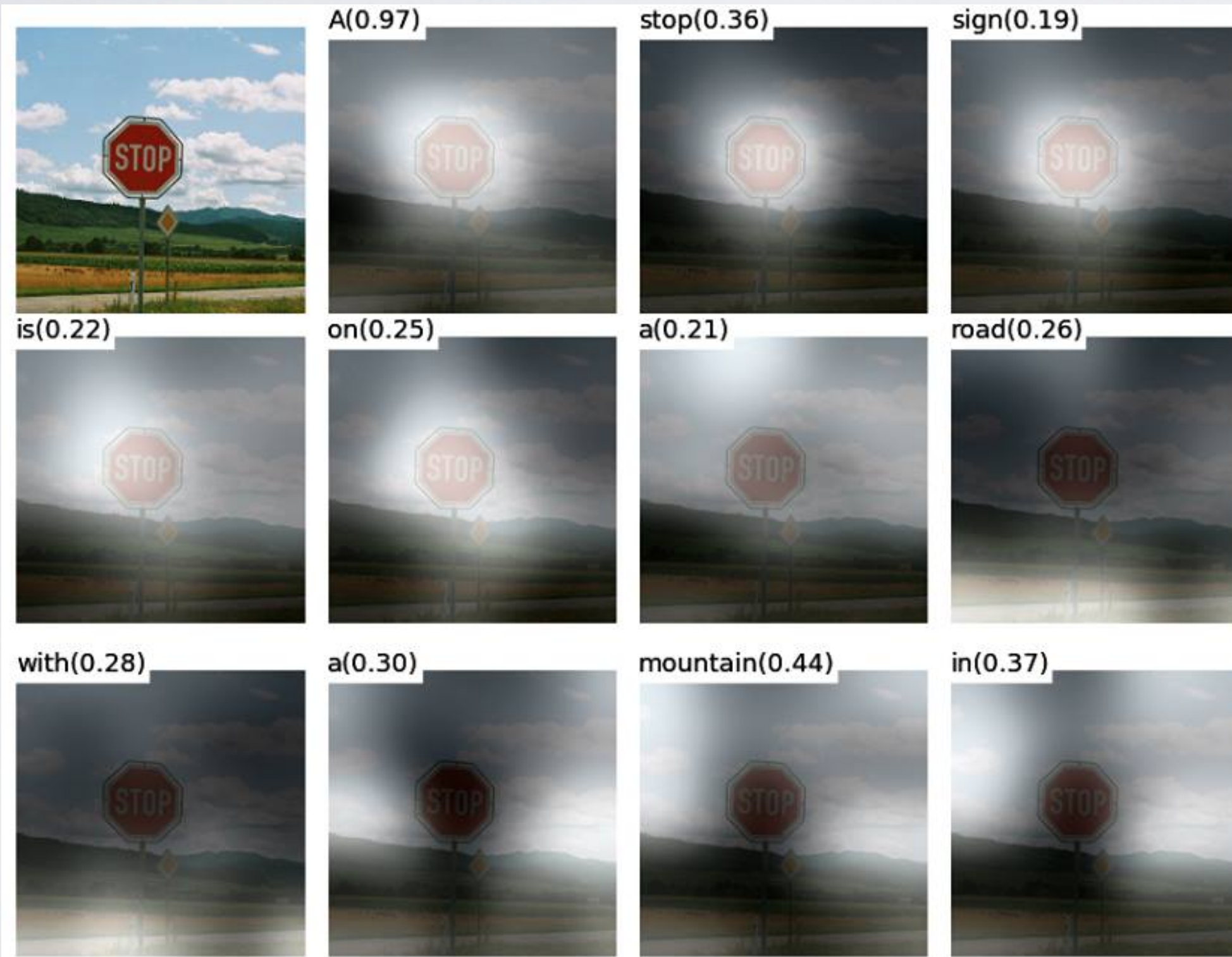
BEYOND NATURAL LANGUAGES

Topics: Beyond Natural Languages — Image Caption Generation (Examples)



BEYOND NATURAL LANGUAGES

Topics: Beyond Natural Languages — Image Caption Generation (Examples)



BEYOND NATURAL LANGUAGES

Topics: Beyond Natural Languages — Attention Models

- End-to-End Speech Recognition (Chorowski et al., 2015; Chan et al., 2015)

- Video Description Generation (Yao et al., 2015)

- Discrete Optimization (Vinyals et al., 2015)

- and many more...

(Cho et al., 2015) and references therein

+Local+Global: A **man** and a **woman** are **talking** on the **road**

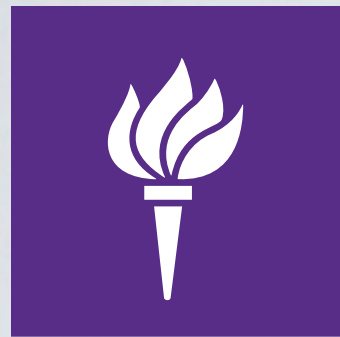
Ref: A man and a woman ride a motorcycle

+Local+Global: **Someone** is **frying** a **fish** in a **pot**

Ref: A woman is frying food

Connectionist Approach to Natural Language Understanding

(see the slides 33-40 of my talk at CVSC 2015)



NEW YORK UNIVERSITY



NYU

**GRADUATE SCHOOL
OF ARTS & SCIENCE**



NYU

**COURANT INSTITUTE OF
MATHEMATICAL SCIENCES**

- **Department of Computer Science**

- Ph.D. Programme: Application dl. 12th December

- **Center for Data Science**

- M.Sc. Programme in Data Science: Application dl. 4th February



Aalto University



Aalto University
School of Science

- **Department of Computer Science**

- M.Sc. Programme in Machine Learning and Data Mining (Macadamia)
- Ph.D. Programme: Prof. Tapani Raiko