Insight
Centre for Data Analytics

DEEP
LEARNING
WORKSHOP

Dublin City University
21-22 May 2018

DCU

#InsightDL2018

Day 1 Lecture 6

Neural Machine Translation

Xavier Giro-i-Nieto
xavier.giro@upc.edu

Associate Professor
Intelligent Data Science and Artificial Intelligence Center
Universitat Politecnica de Catalunya (UPC)

Acknowledgements

Marta R. Costa-jussà

TALP Research Center

Acknowledgements

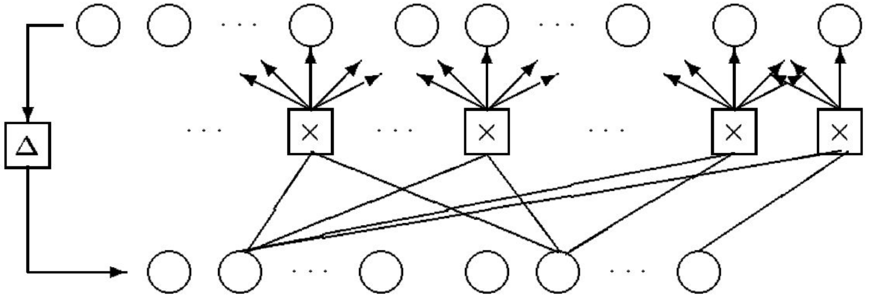
Kyunghyun Cho

NYU | COURANT INSTITUTE OF MATHEMATICAL SCIENCES

Precedents

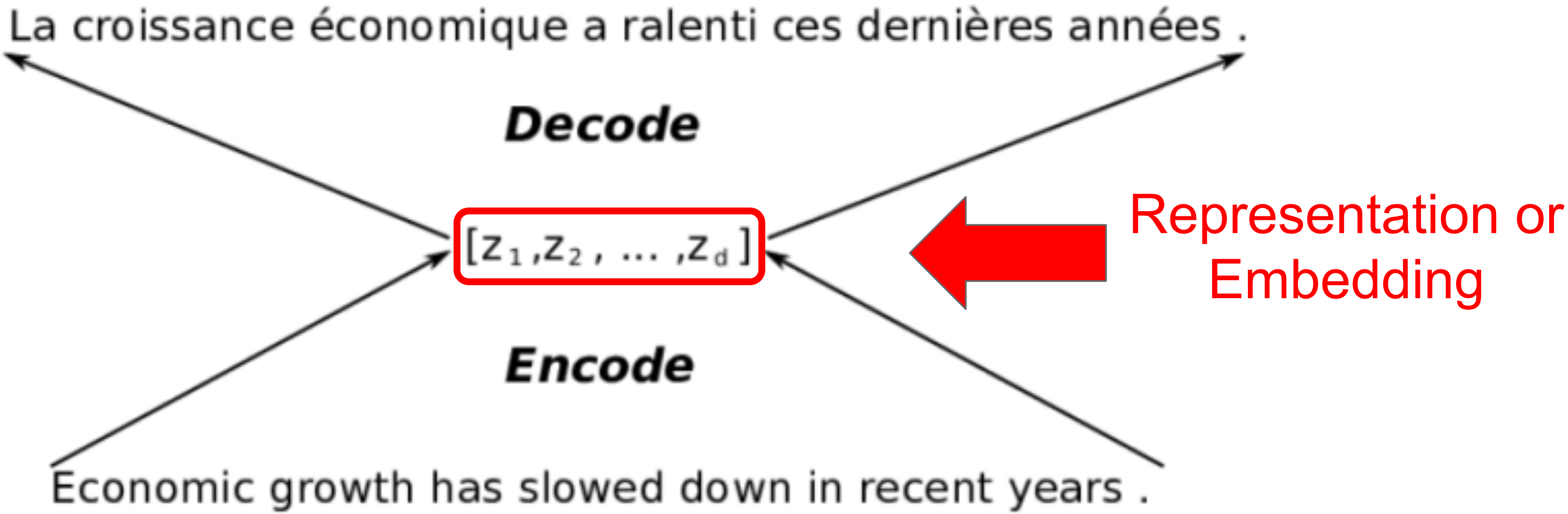
Asynchronous translations with recurrent neural nets

Ramón P. Ñeco, Mikel L. Forcada
Departament de Llenguatges i Sistemes Informàtics,
Universitat d'Alacant, E-03071 Alacant, Spain.
E-mail: {neco,mlf}@dlsi.ua.es



Neco, R.P. and Forcada, M.L., 1997, June. Asynchronous translations with recurrent neural nets. In Neural Networks, 1997., International Conference on (Vol. 4, pp. 2535-2540). IEEE.

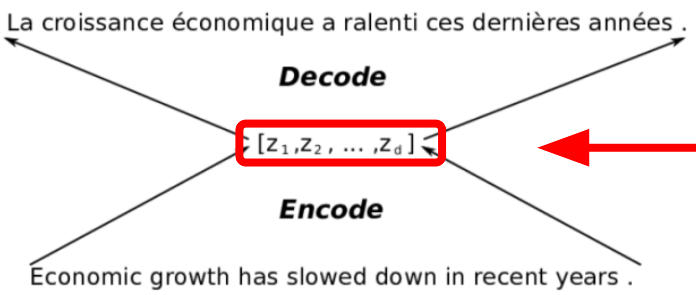
Encoder-Decoder



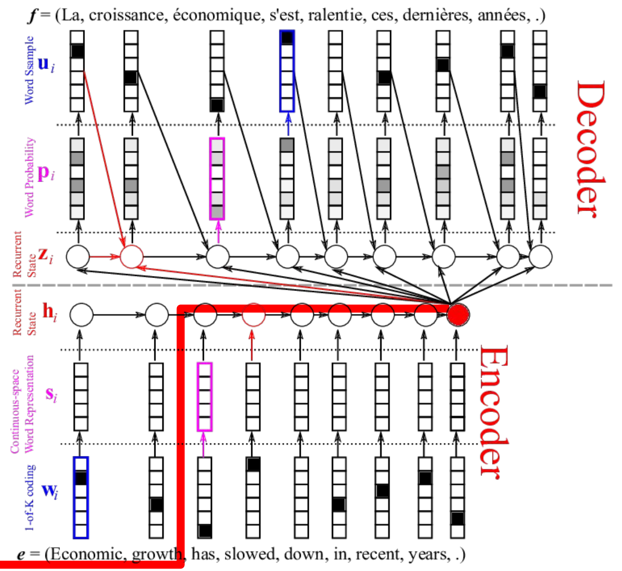
Cho, Kyunghyun, Bart Van Merriënboer, Dzmitry Bahdanau, and Yoshua Bengio. "On the properties of neural machine translation: Encoder-decoder approaches." SSST-8 (2014).

Encoder-Decoder

Front View



Side View

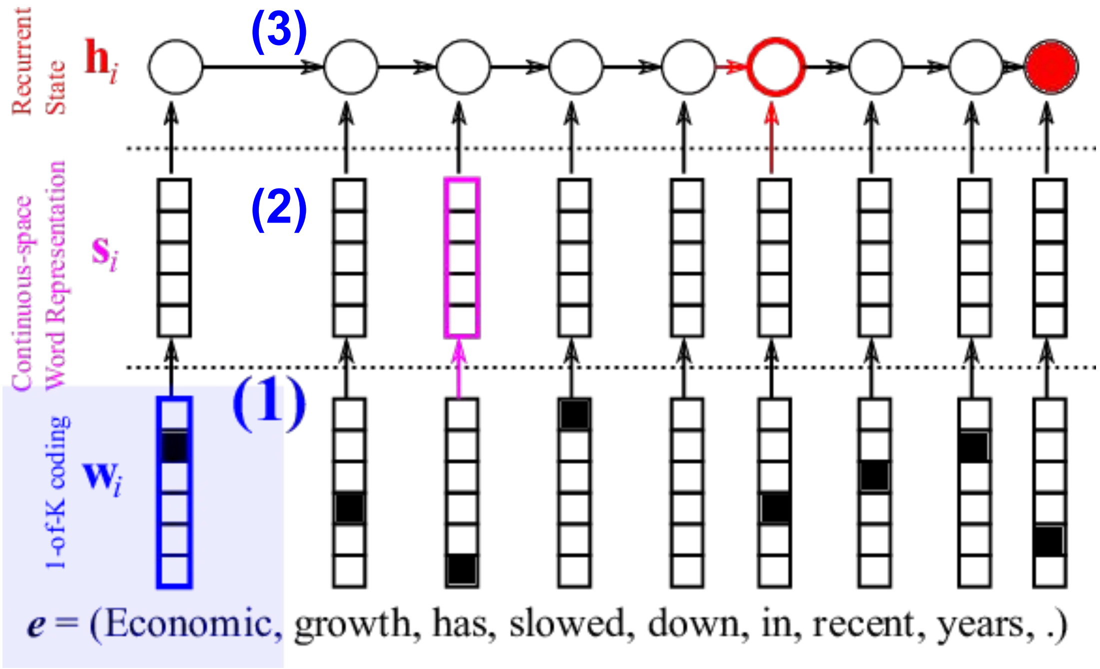


Representation of the sentence

Kyunghyun Cho, [“Introduction to Neural Machine Translation with GPUs”](#) (2015)
Cho, Kyunghyun, Bart Van Merriënboer, Caglar Gulcehre, Dzmitry Bahdanau, Fethi Bougares, Holger Schwenk, and Yoshua Bengio. ["Learning phrase representations using RNN encoder-decoder for statistical machine translation."](#) arXiv preprint arXiv:1406.1078 (2014).

Encoder

Encoder in three steps



(1) One hot encoding

(2) Word embedding

(3) Sequence summarization

Kyunghyun Cho, [“Introduction to Neural Machine Translation with GPUs”]() (2015)

(1) One hot encoding

cat: x^T = [1,0,0, ..., 0]

dog: x^T = [0,1,0, ..., 0]

.

.

house: x^T = [0,0,0, …,0,1,0,...,0]

.

.

.

(2) Word embeddings

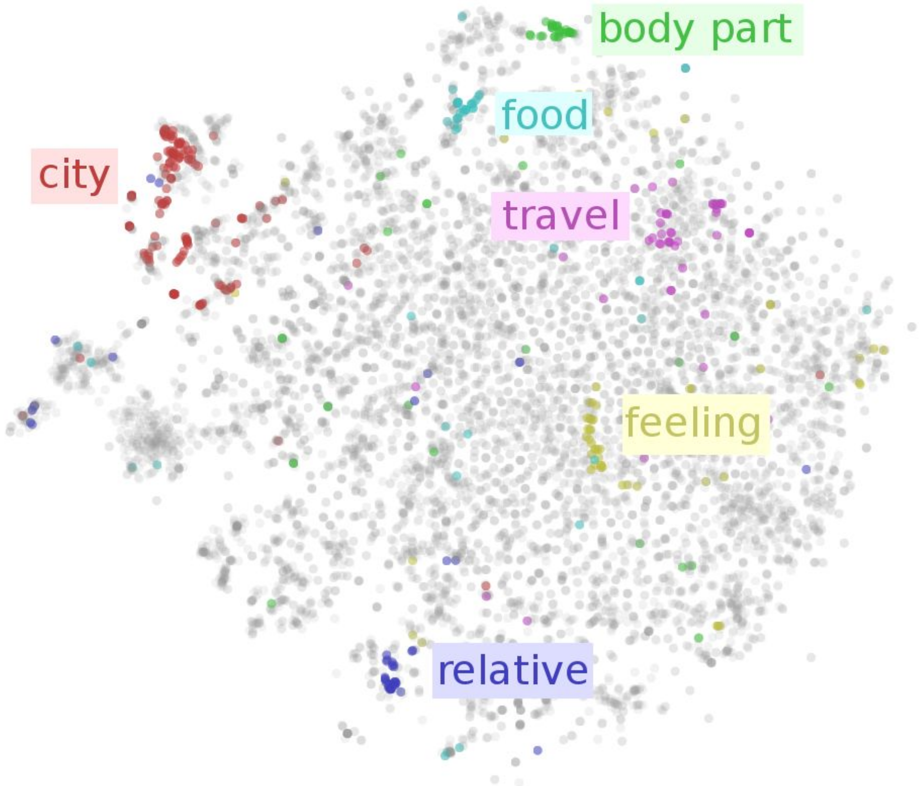
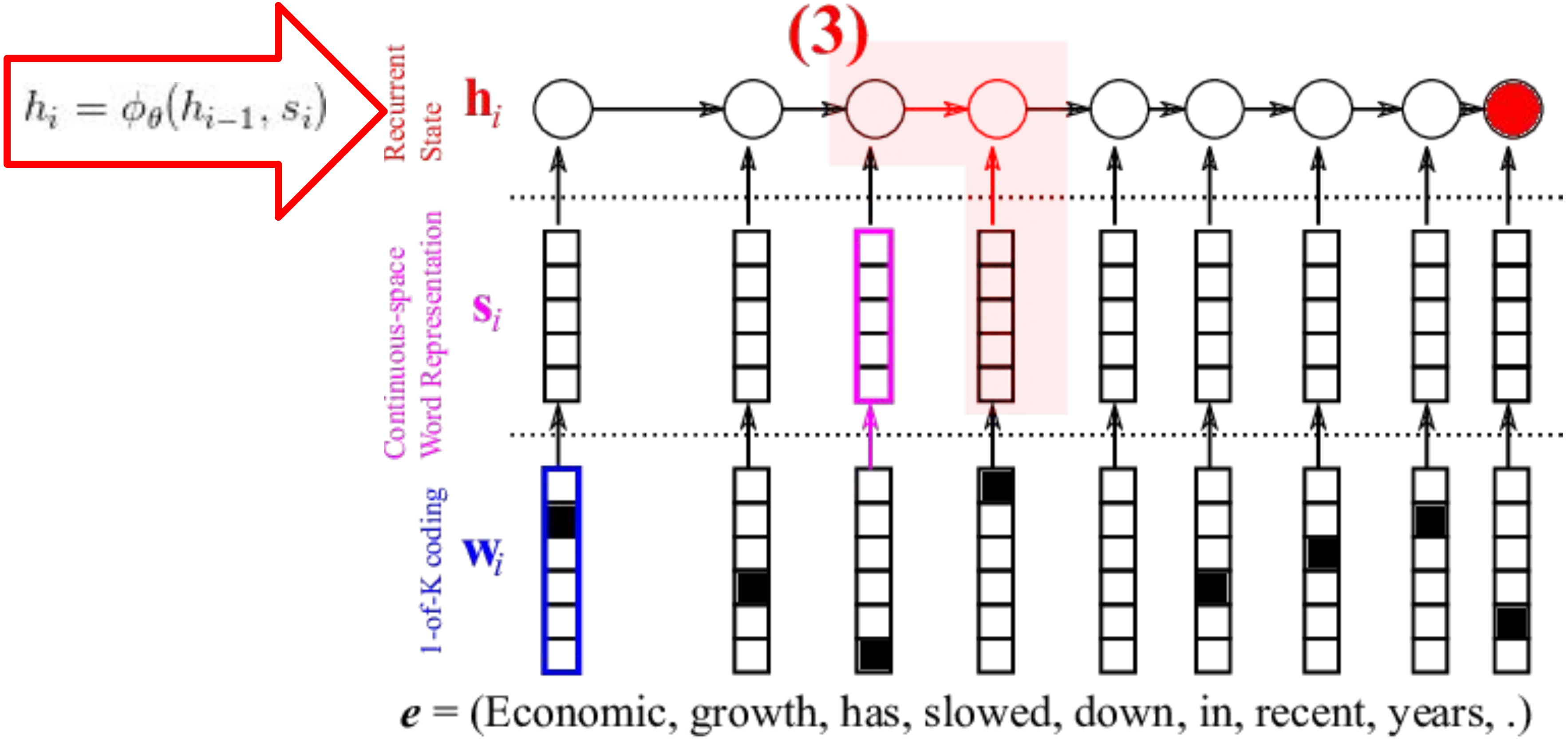


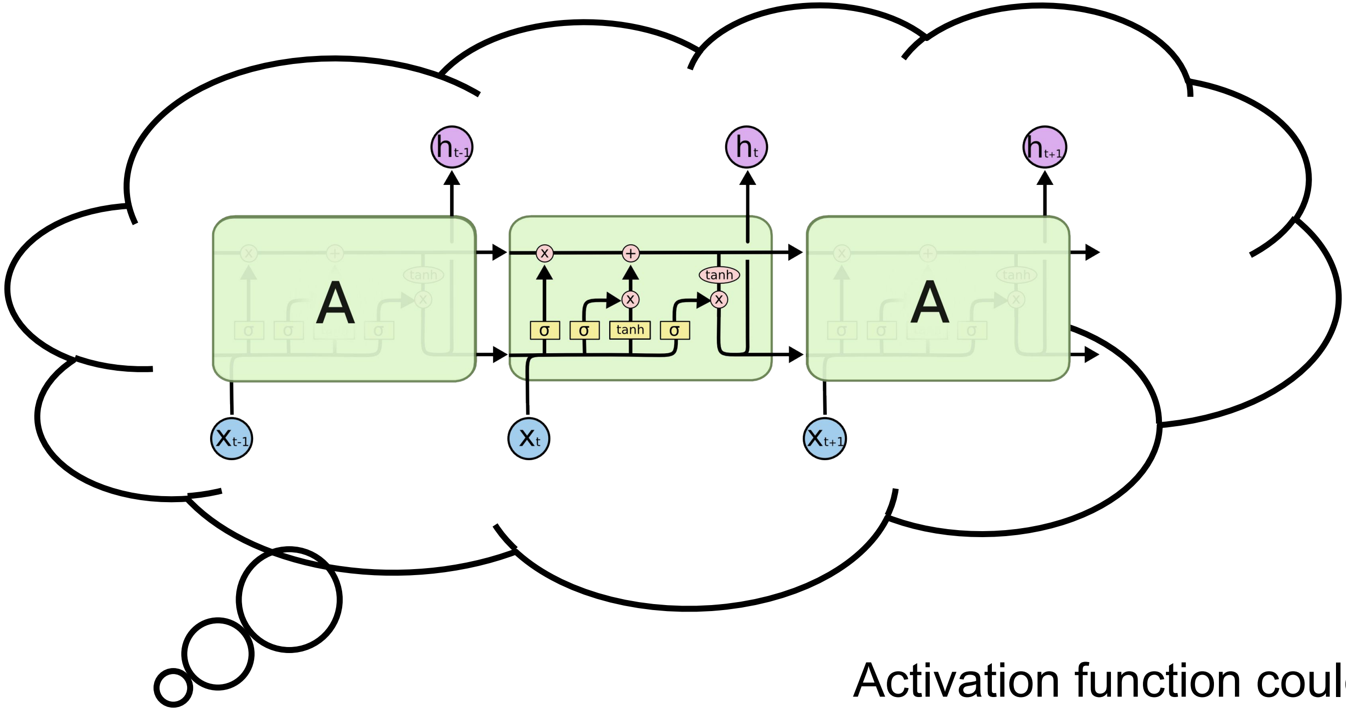
Figure: Christopher Olah, [Visualizing Representations]

(3) Sequence summarization



Kyunghyun Cho, [“Introduction to Neural Machine Translation with GPUs”]() (2015)

(3) Sequence summarization



Sequence

Activation function could be LSTM, GRU, QRNN, pLSTM...

Figure: Cristopher Olah, "Understanding LSTM Networks" (2015)

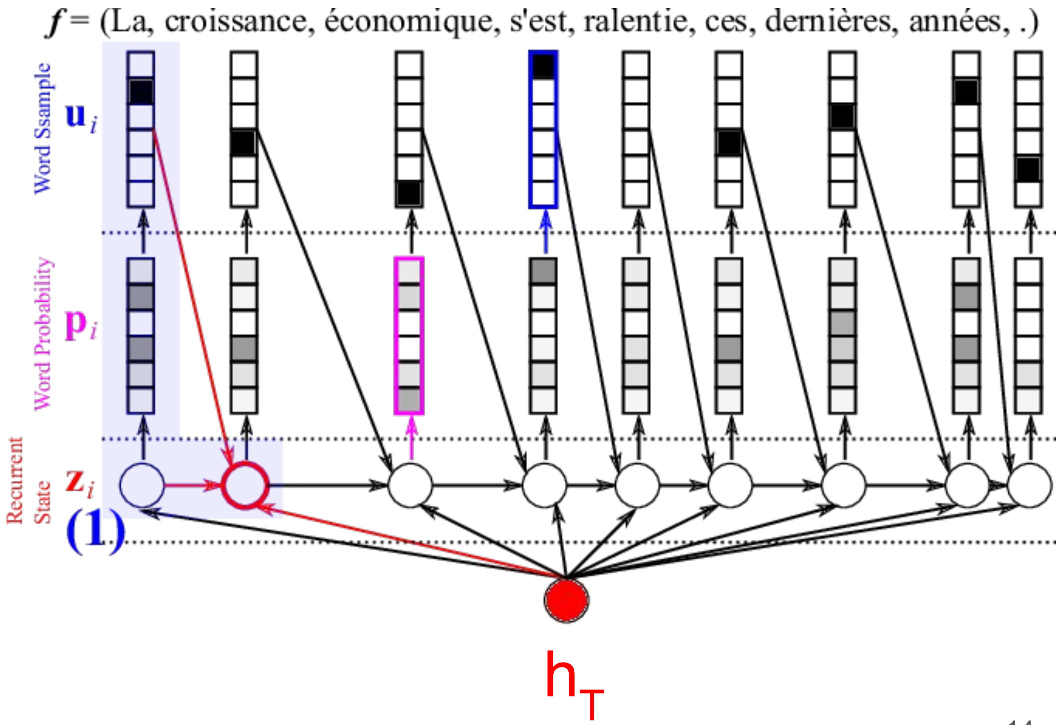
Decoder

Decoder

The Recurrent State (z_i) of the decoder is determined by:

1) summary vector h_T
2) previous output word u_{i-1}
3) previous state z_{i-1}

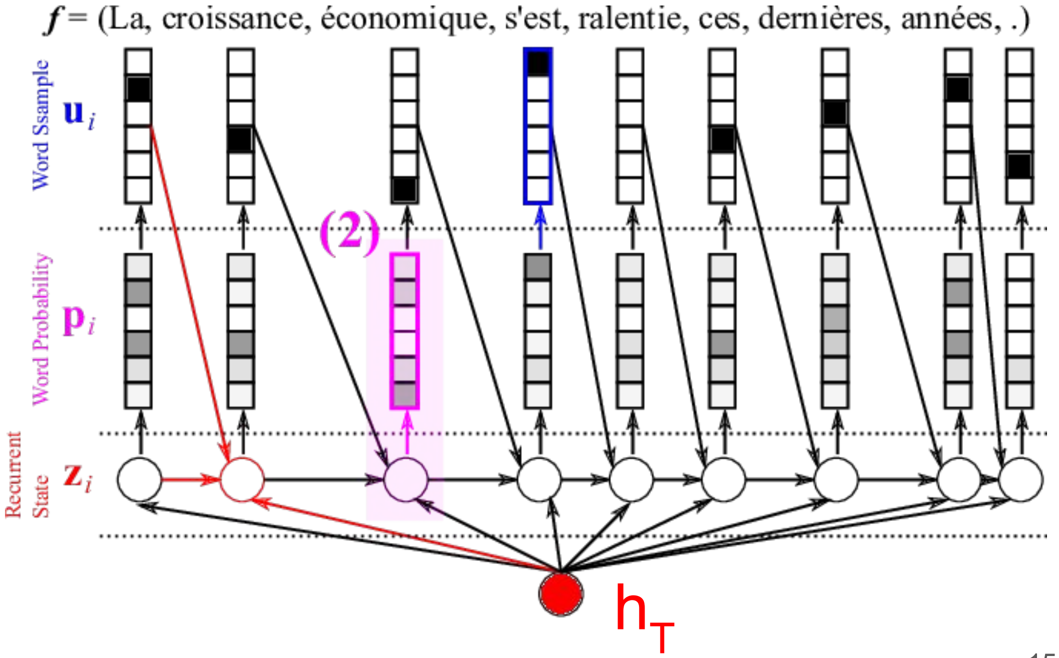
$$z_i = \phi_{\theta'}(h_T, u_{i-1}, z_{i-1}).$$



Kyunghyun Cho, "Introduction to Neural Machine Translation with GPUs" (2015)

Decoder

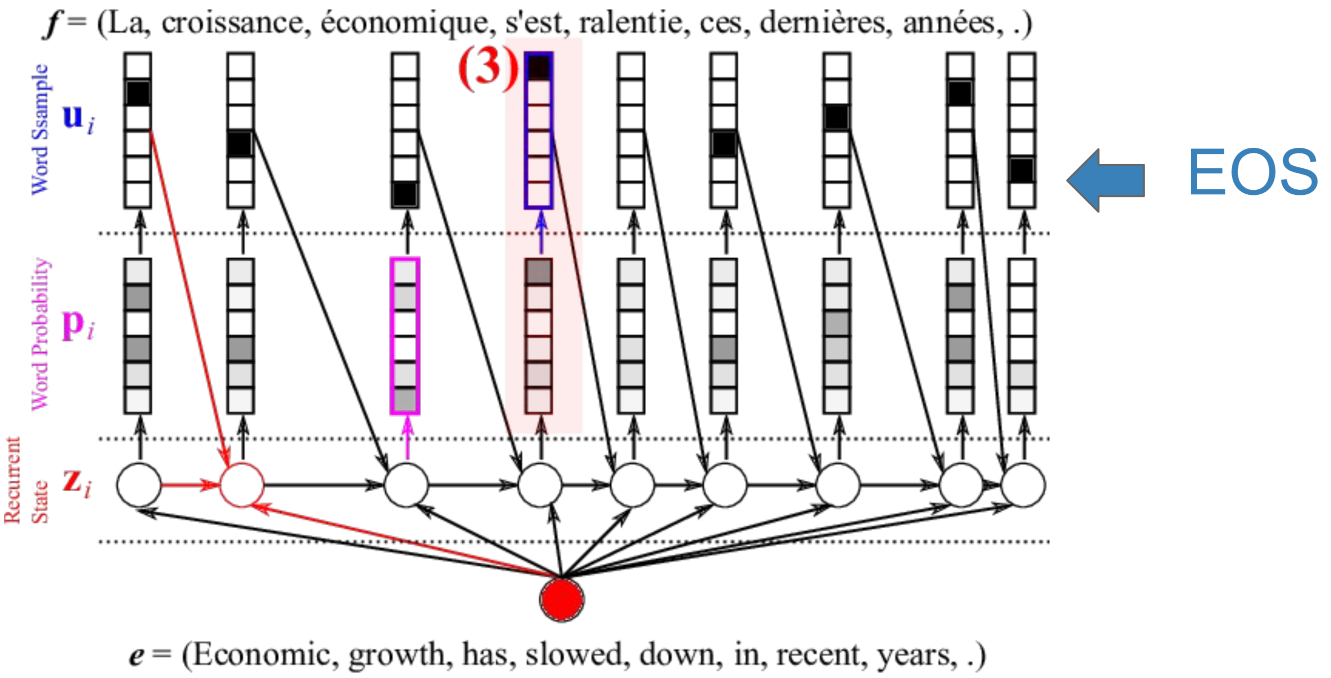
With z_i updated, we can compute a **probability p_i** for each word *i* as an output of the RNN:



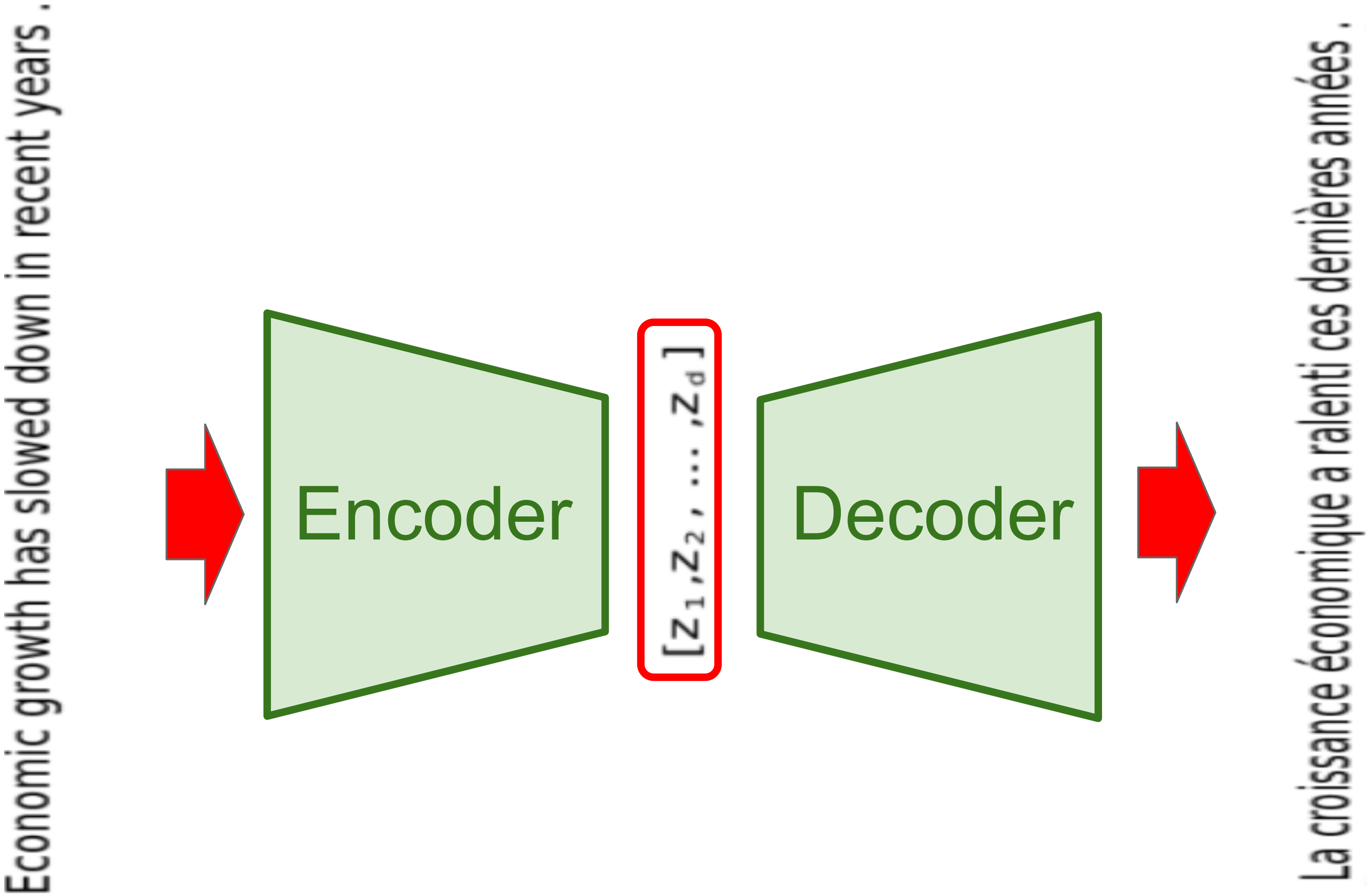
Kyunghyun Cho, “Introduction to Neural Machine Translation with GPUs” (2015)

Decoder

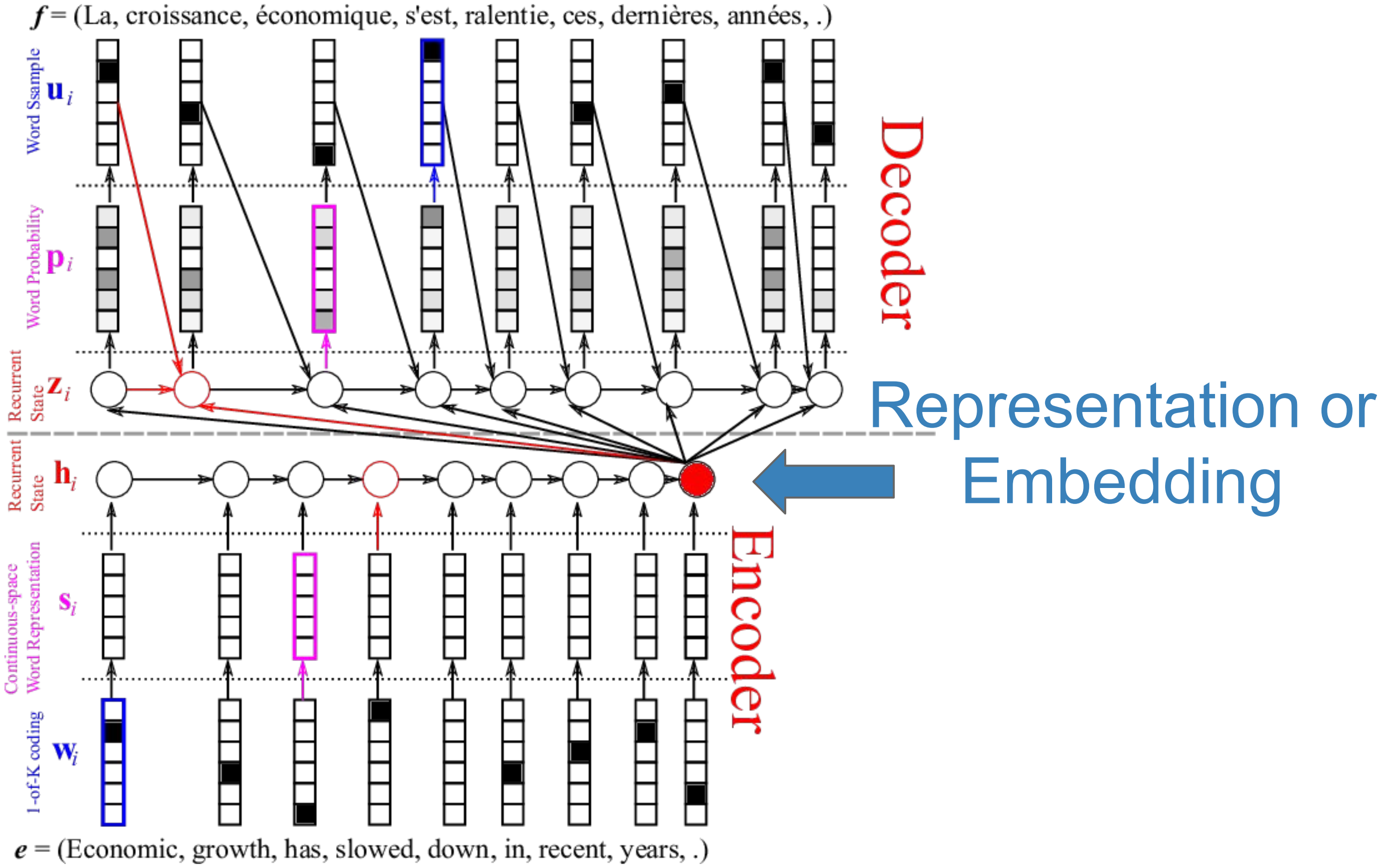
More words for the decoded sentence are generated until a <EOS> (End Of Sentence) “word” is predicted.



Kyunghyun Cho, “Introduction to Neural Machine Translation with GPUs” (2015)


Economic growth has slowed down in recent years .
Encoder
[z1,z2, ... ,zd]
Decoder
La croissance économique a ralenti ces dernières années .

Encoder-Decoder



Kyunghyun Cho, [“Introduction to Neural Machine Translation with GPUs”]() (2015)

Parallel corpus

Training requires a large dataset of pairs of sentences in the two languages to translate.

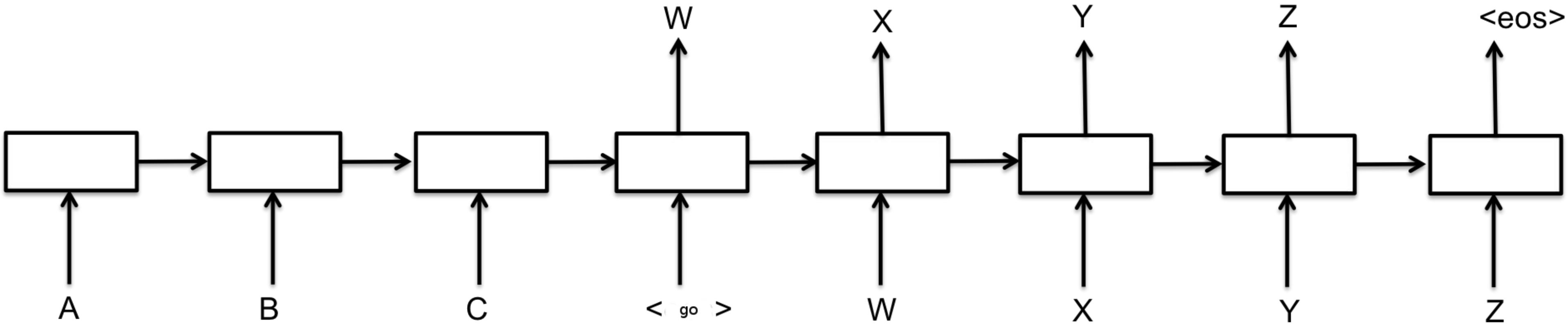
Source	Translation Model
at the end of the	[a la fin de la] [ŕ la fin des années] [être supprimés à la fin de la]
for the first time	[r © pour la premirëre fois] [été donnés pour la première fois] [été commémorée pour la première fois]
in the United States and	[? aux ?tats-Unis et] [été ouvertes aux États-Unis et] [été constatées aux États-Unis et]
, as well as	[?s , qu'] [?s , ainsi que] [?re aussi bien que]
one of the most	[?t ?l' un des plus] [?l' un des plus] [être retenue comme un de ses plus]

Cho, Kyunghyun, Bart Van Merriënboer, Caglar Gulcehre, Dzmitry Bahdanau, Fethi Bougares, Holger Schwenk, and Yoshua Bengio. ["Learning phrase representations using RNN encoder-decoder for statistical machine translation."]() AMNLP 2014.

Seq2Seq

The **Seq2Seq** variation:

- trigger the output generation with an input <go> symbol.
- the predicted word at timestep *t*, becomes the input at *t+1*.



Sutskever, Ilya, Oriol Vinyals, and Quoc V. Le. "Sequence to sequence learning with neural networks." NIPS 2014.

Seq2Seq

Sutskever, Ilya, Oriol Vinyals, and Quoc V. Le. "Sequence to sequence learning with neural networks." NIPS 2014.

Questions?