

Deep Learning for NLP: TEXT TRANSLATION

Jay Urbain, PhD

Professor, Electrical Engineering and Computer
Science Department

Milwaukee School of Engineering

urbain@msoe.edu

HUMAN ATTENTION FOR REASONING

The image shows a screenshot of the eyeCede tool interface. On the left, there is a code editor window titled "eyeCede (click for details)". It contains Python code with several red highlights and annotations:

```
def between(numbers, low, high):
    winners = []
    for num in numbers:
        if (low < num) and (num < high):
            winners.append(num)
    return winners

def common(list1, list2):
    winners = []
    for item1 in list1:
        if item1 in list2:
            winners.append(item1)
    return winners

x = [2, 8, 7, 9, -5, 6, 2]
x_btwn = between(x, 2, 10)
print x_btwn

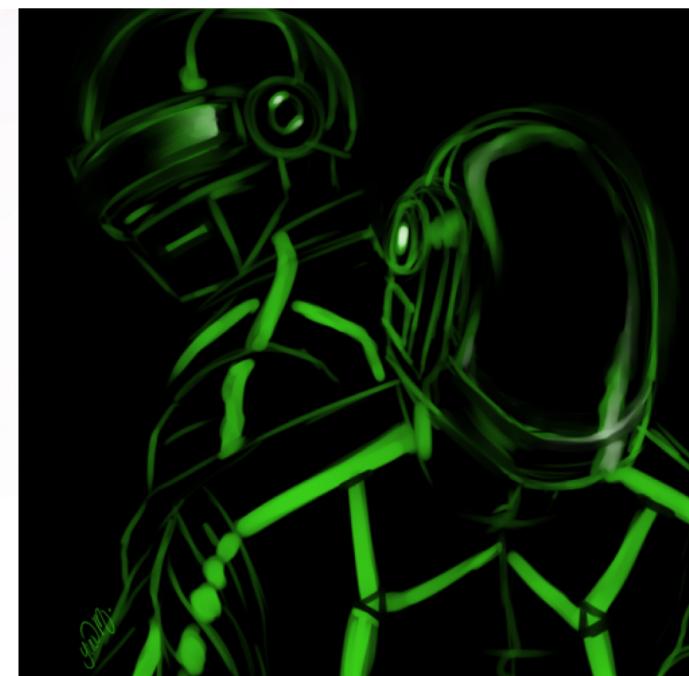
y = [1, -3, 10, 8, 3, 9, 1]
y_btwn = between(y, -2, 9)
print y_btwn

xy_common = common(x, y)
print xy_common
```

A red arrow points from the annotation "what will this program output?" to a reasoning step on the right. The reasoning step has a title "what will this program output?" and contains the text "The output of the program is [1, 8, 9, 3, 6]".

WHY GIVE ATTENTION TO DEEP LEARNING?

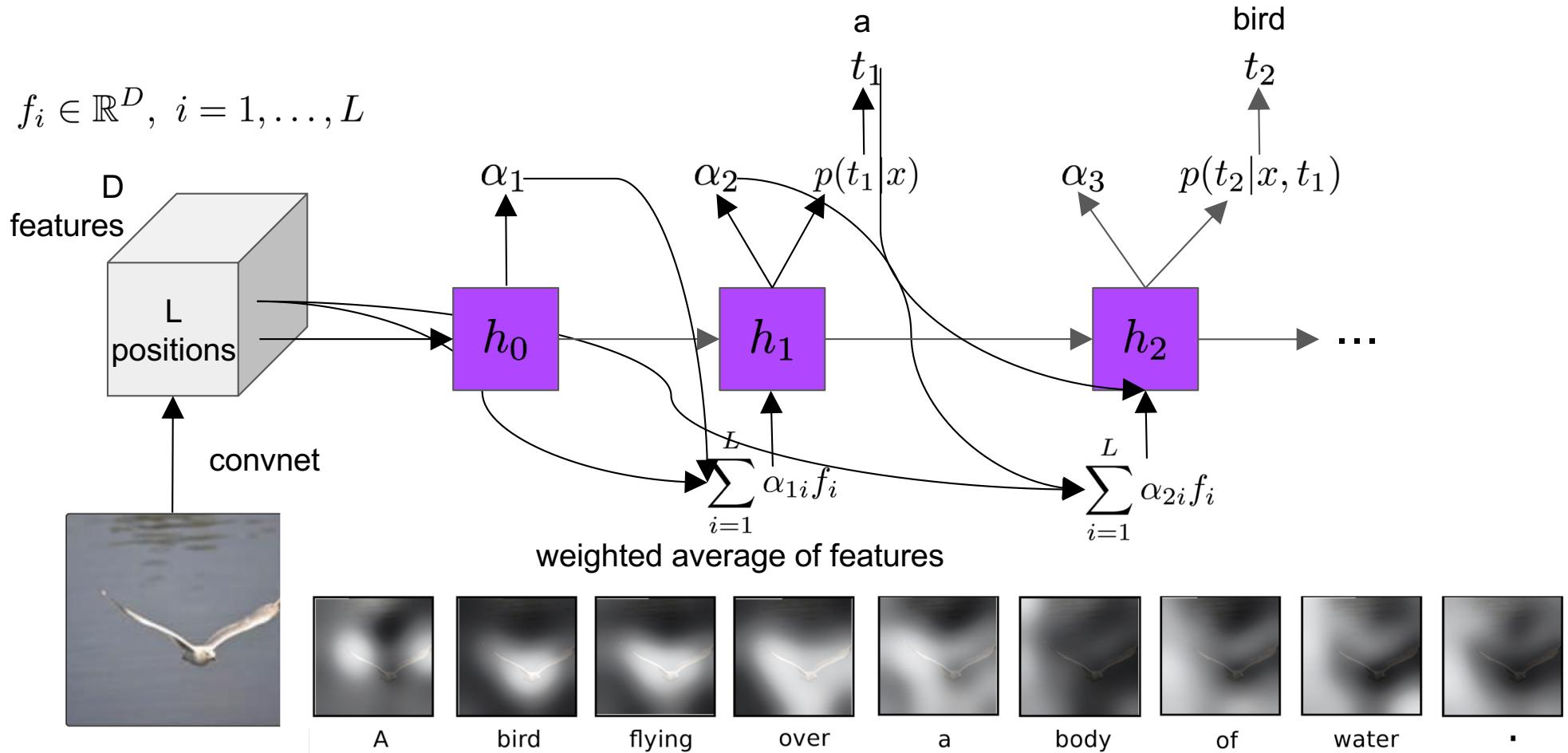
- **Better** interpretability
- **Faster** models
- **Stronger** results



Slide idea: J. Redmon, YOLO 9000 presentation
Picture from [tumblr](#)

SOFT ATTENTION FOR IMAGE CAPTIONING

$$f_i \in \mathbb{R}^D, i = 1, \dots, L$$



Xu et al. "Show, Attend and Tell: Neural Image Caption Generation with Visual Attention" ICML 2015

INTROSPECTION VIA SOFT ATTENTION



A stop sign is on a road with a mountain in the background.



A man is talking on his cell phone while another man watches.



SOFT VS. HARD ATTENTION

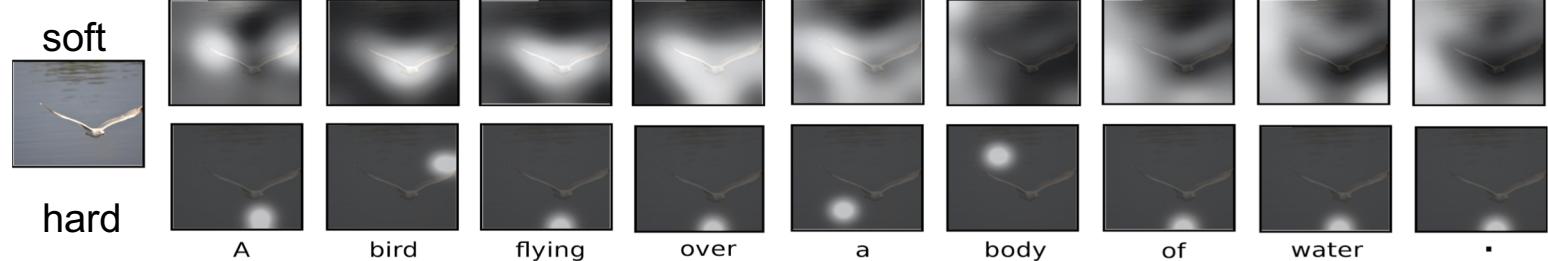
- Hard attention

- Better results (sometimes?)
- More principled
- Hard to train

Model	BLEU				METEOR
	BLEU-1	BLEU-2	BLEU-3	BLEU-4	
Log Bilinear [◦]	70.8	48.9	34.4	24.3	20.03
Soft-Attention	70.7	49.2	34.4	24.3	23.90
Hard-Attention	71.8	50.4	35.7	25.0	23.04

- Soft attention

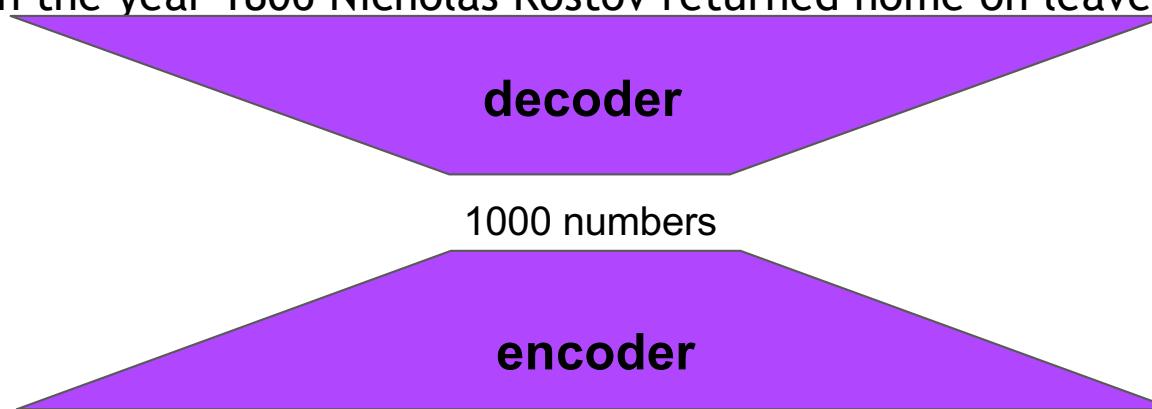
- Not always applicable
- Easy to train



Xu et al. "Show, Attend and Tell: Neural Image Caption Generation with Visual Attention" ICML 2015

NEURAL MACHINE TRANSLATION AND LONG SENTENCES

Early in the year 1806 Nicholas Rostov returned home on leave. <EOS>

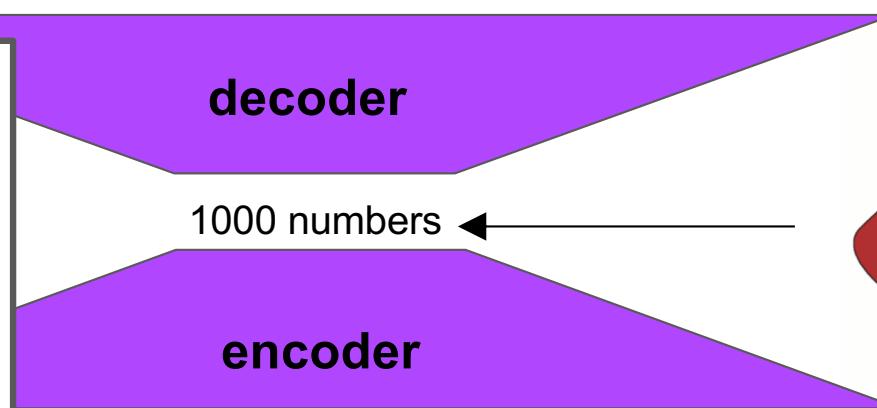
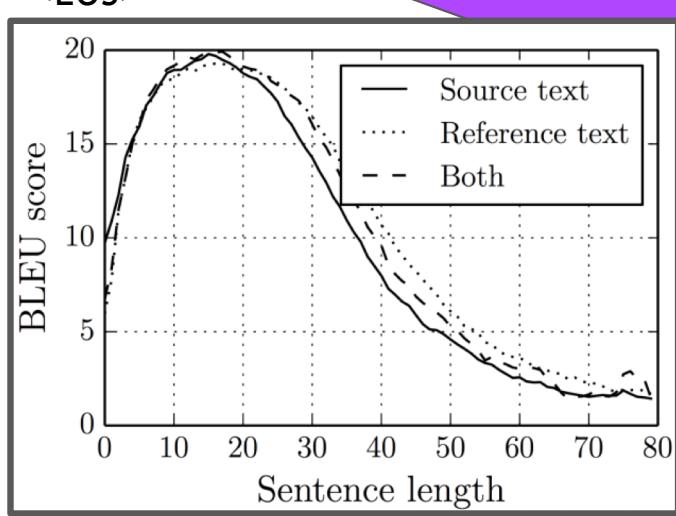


В начале 1806 года Николай Ростов вернулся в отпуск. <EOS>

NEURAL MACHINE TRANSLATION AND LONG SENTENCES

Meeting a comrade at the last post station but one before Moscow, Denisov had drunk three bottles of wine with him and, despite the jolting ruts across the snow-covered road, did not once wake up on the way to Moscow, but lay at the bottom of the sleigh beside Rostov, who grew more and more impatient the nearer they got to Moscow.

<EOS>

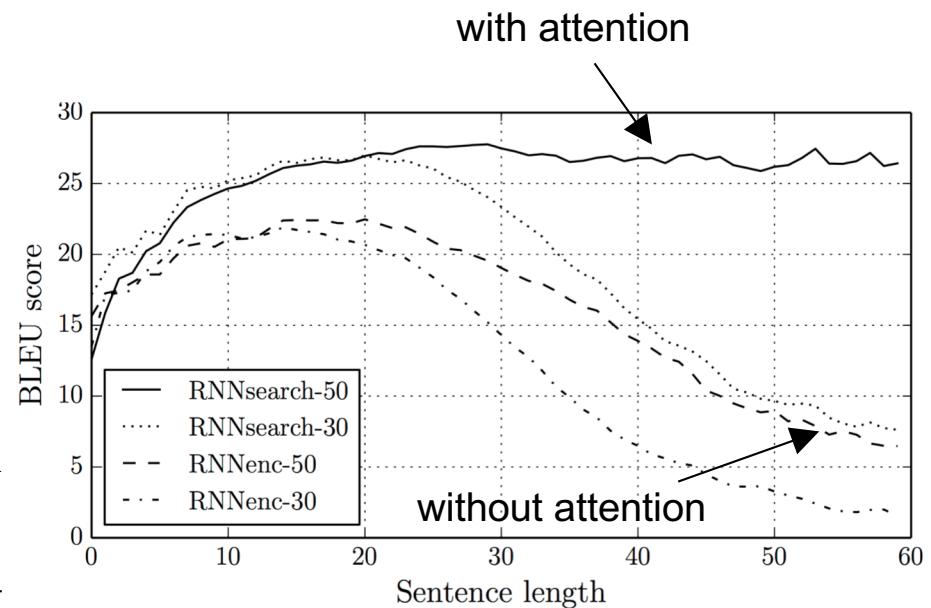
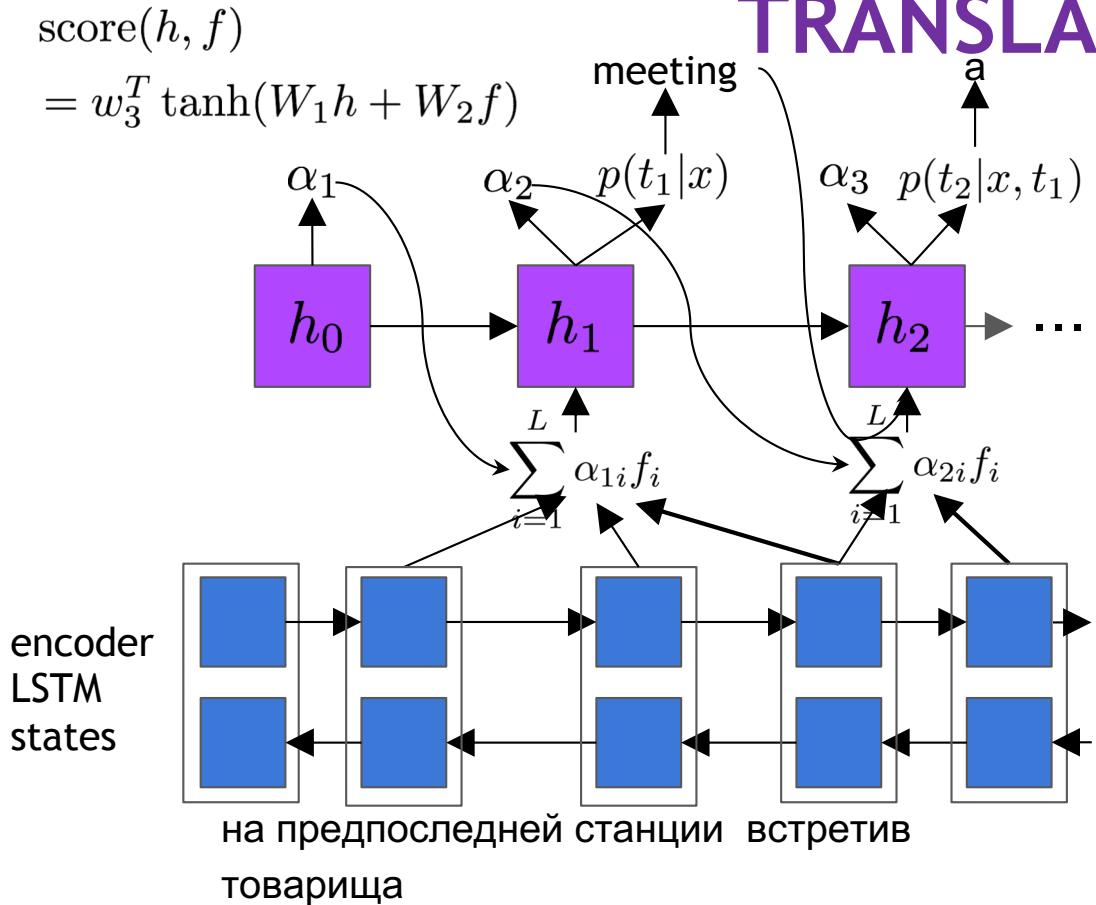


в товарища, Денисов выпил с ним три бутылки вина и подъезжая к
Москве не просыпался, лежа на дне перекладных саней, подле Ростова, который,
по мере приближения к Москве, приходил все более и более в нетерпение. <EOS>

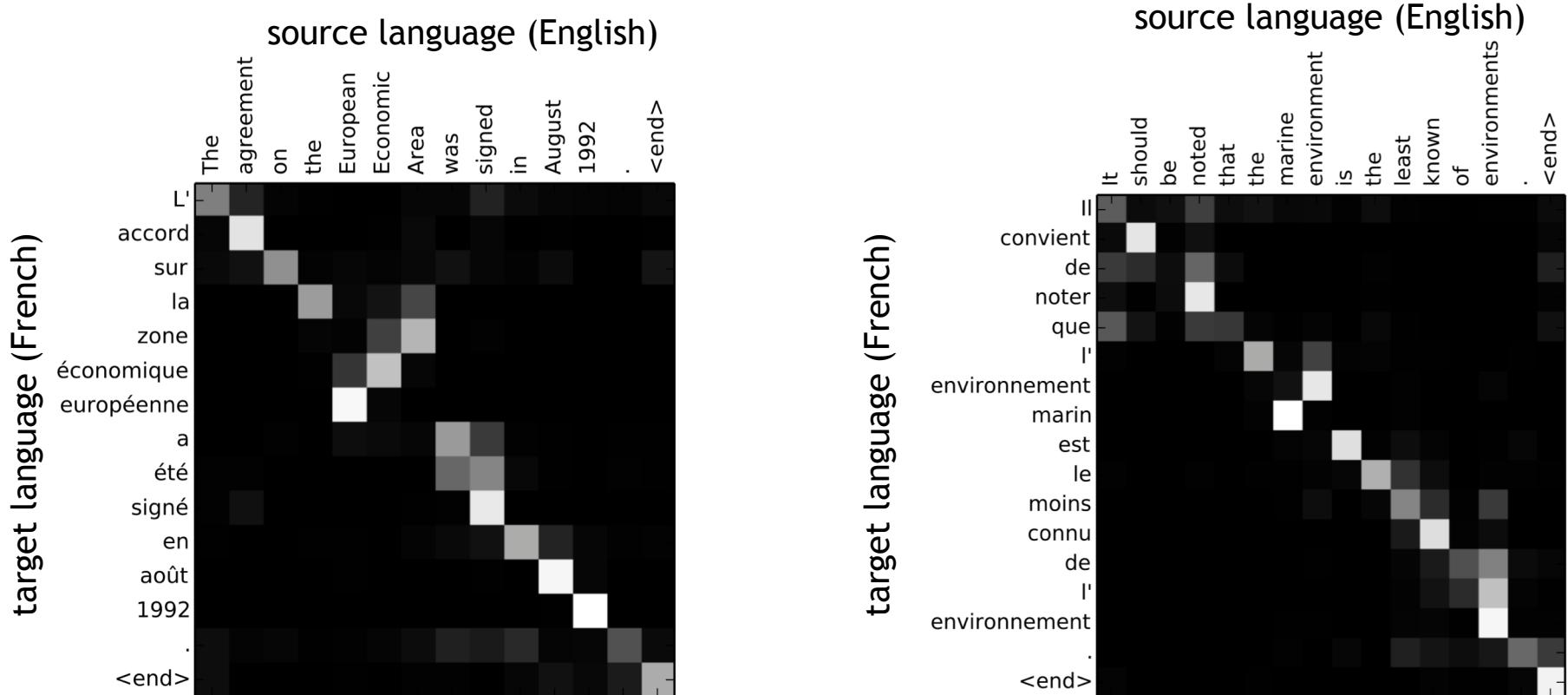
Leo Tolstoy "War and Peace", 1869

Cho et al. "On the Properties of Neural Machine Translation: Encoder-Decoder Approaches", 2014

SOFT ATTENTION FOR NEURAL MACHINE TRANSLATION



ATTENTION MAPS IN TRANSLATION



Bahdanau et al. "Neural Machine Translation by Jointly Learning to Align and Translate", 2014

