



University of
Zurich^{UZH}

Institute of Computational Linguistics

Neural Machine Translation

Brainhack Zurich 2018

Mathias Müller



1



Translate

[Turn off instant translation](#)

English French German Detect language ▾



French German English ▾

Translate



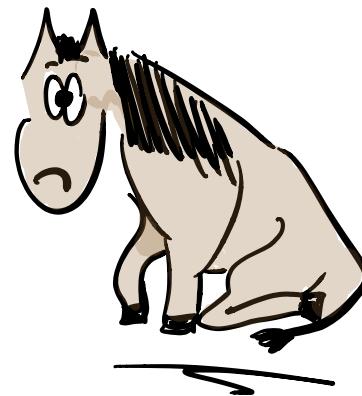
0/5000

Type text or a website address or [translate a document](#).



WorldViews

In Ukraine, Google translates Russia as ‘Mordor’ and top diplomat’s name as ‘sad little horse’





howstuffworks



machine translation



Health

Science

Home & Garden

Auto

Tech

Culture

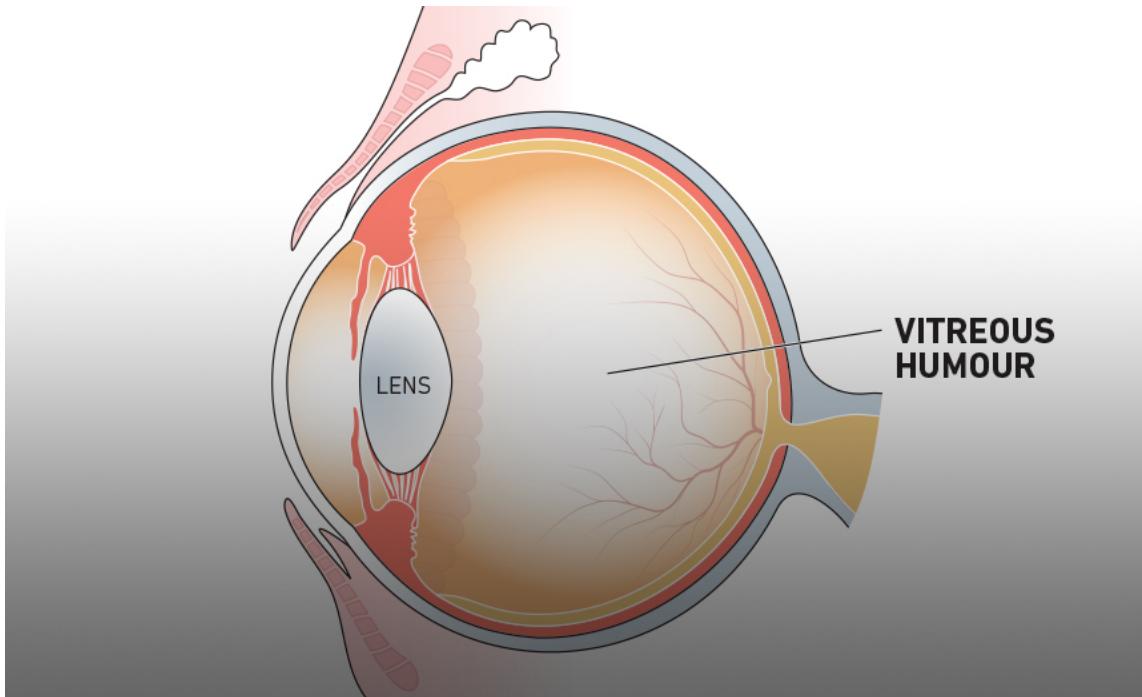
Money

Lifestyle

More ▾



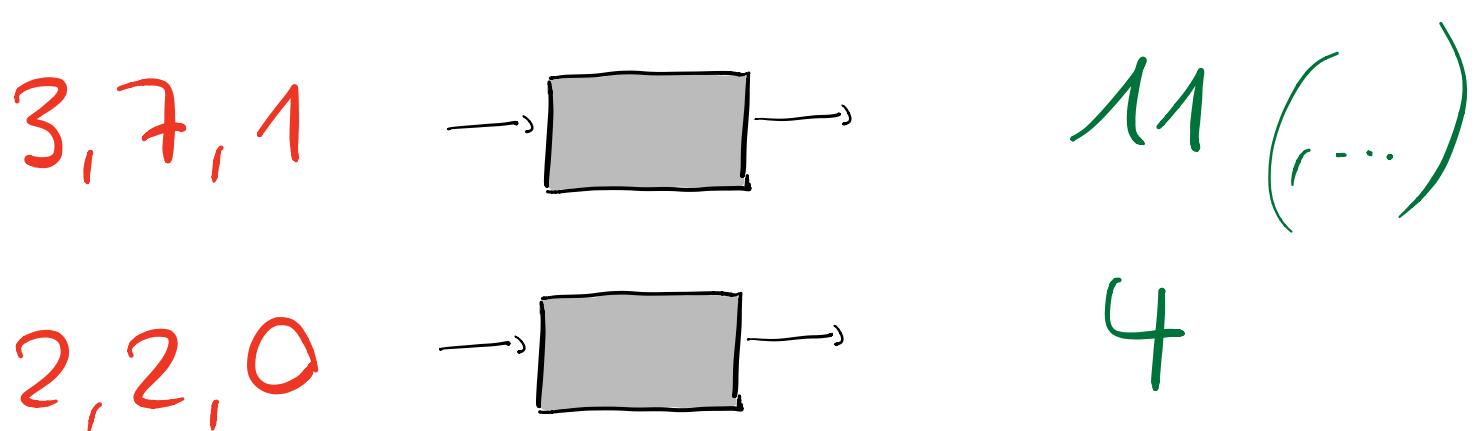
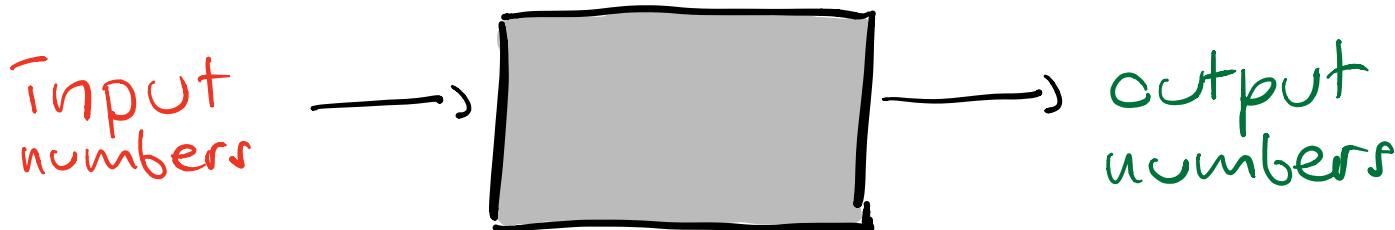
Funny coincidence: modern machine translation is “neural”



Neural Machine Translation

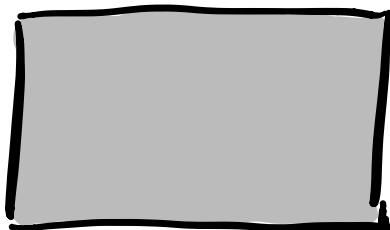
- is “neural” because it uses artificial neural networks
- ~~because it is similar to human brains~~
- neural networks are a particular type of machine learning model

Neural networks: black-box description



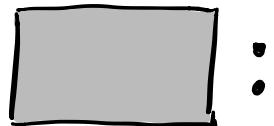
Black-box description of neural machine translation

He is out of
his mind.



Er ist
ausser sich.

prior knowledge we put into :



Putting language into neural networks

~~"Haus"~~

3,7,1

Vocabulary:

"Haus", "er", "im",
"ist"

$$\text{Haus} = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

$$\text{er} = \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix}$$

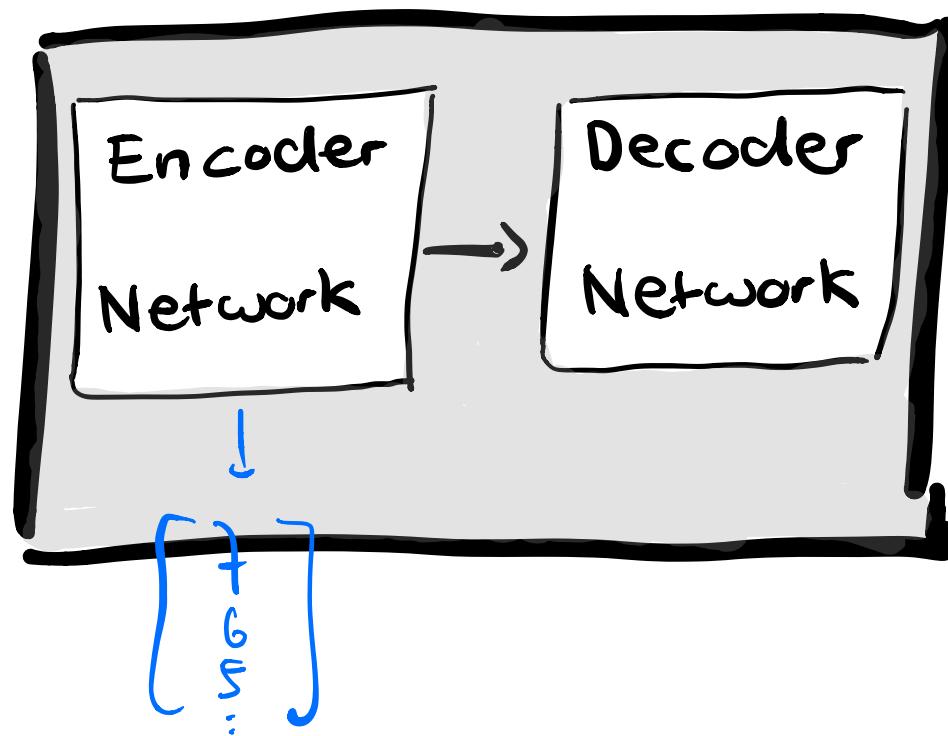
$$\text{im} = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \end{bmatrix}$$

$$\text{ist} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}$$

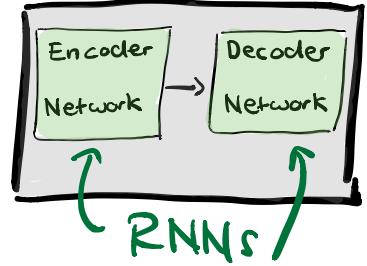
Glass-box neural machine translation

He is out of
his mind.

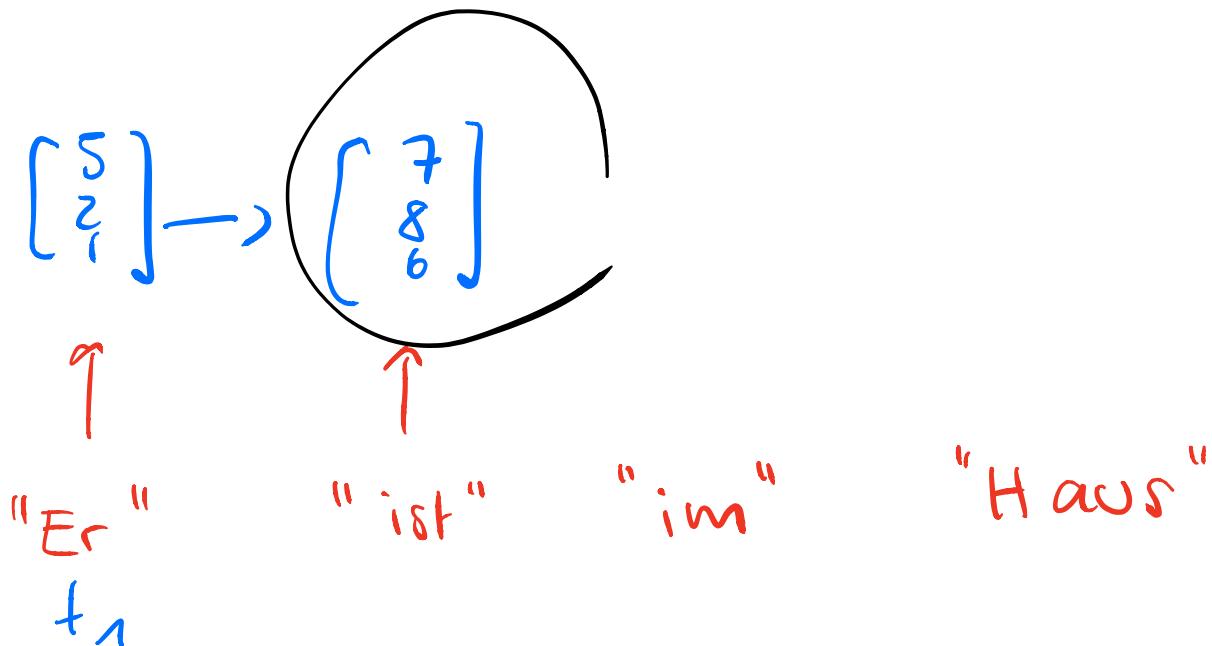
Er ist
ausser sich.



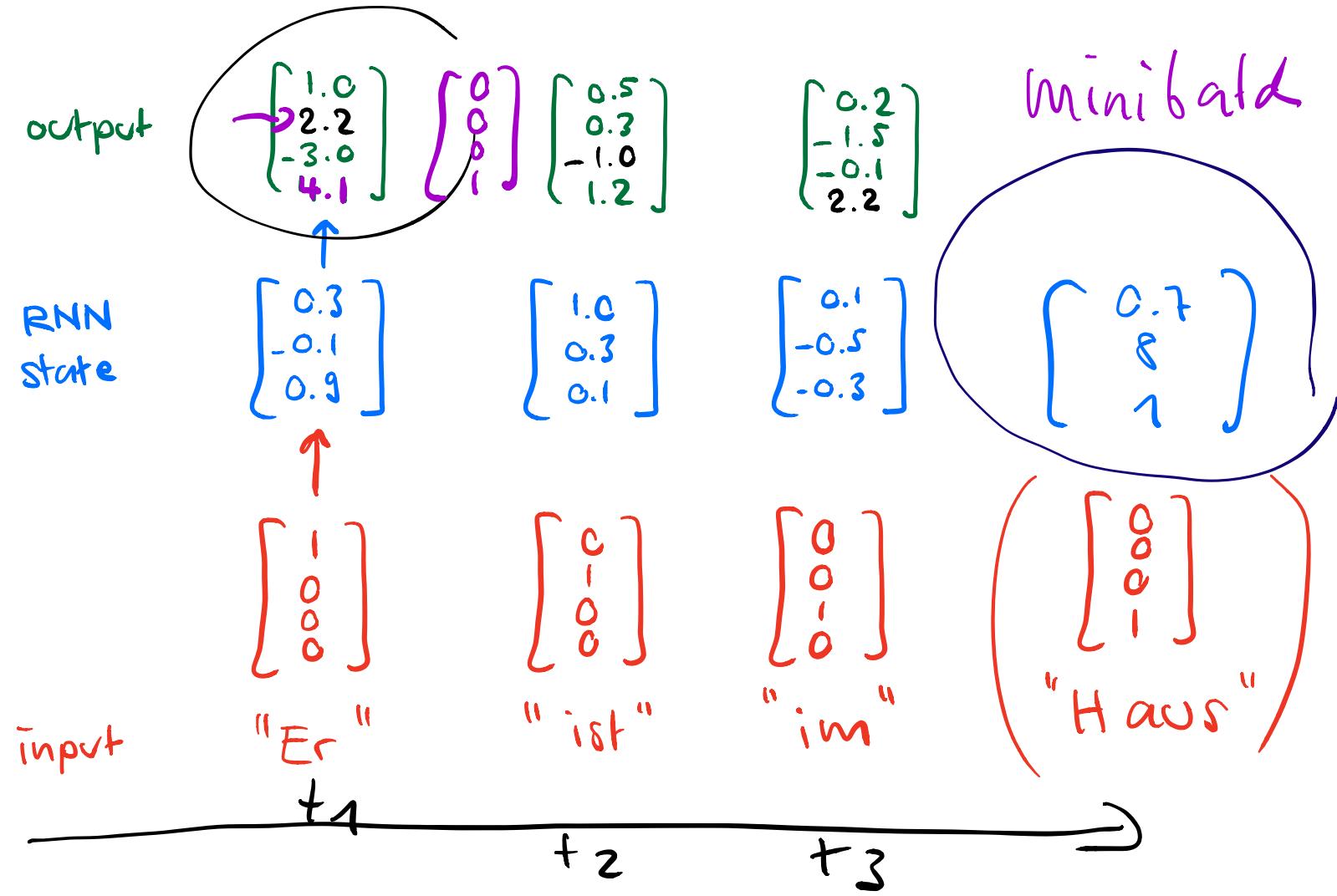
Recurrent neural networks (RNNs)



- specialized for **sequence input**
- have a **state**
- reads words one at a time



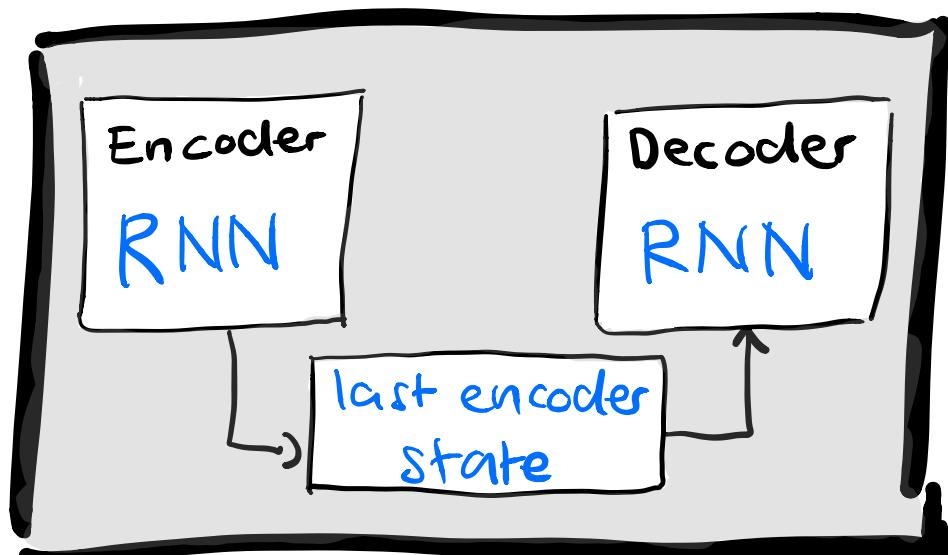
Simple RNN example: a language model



Review glass-box neural machine translation

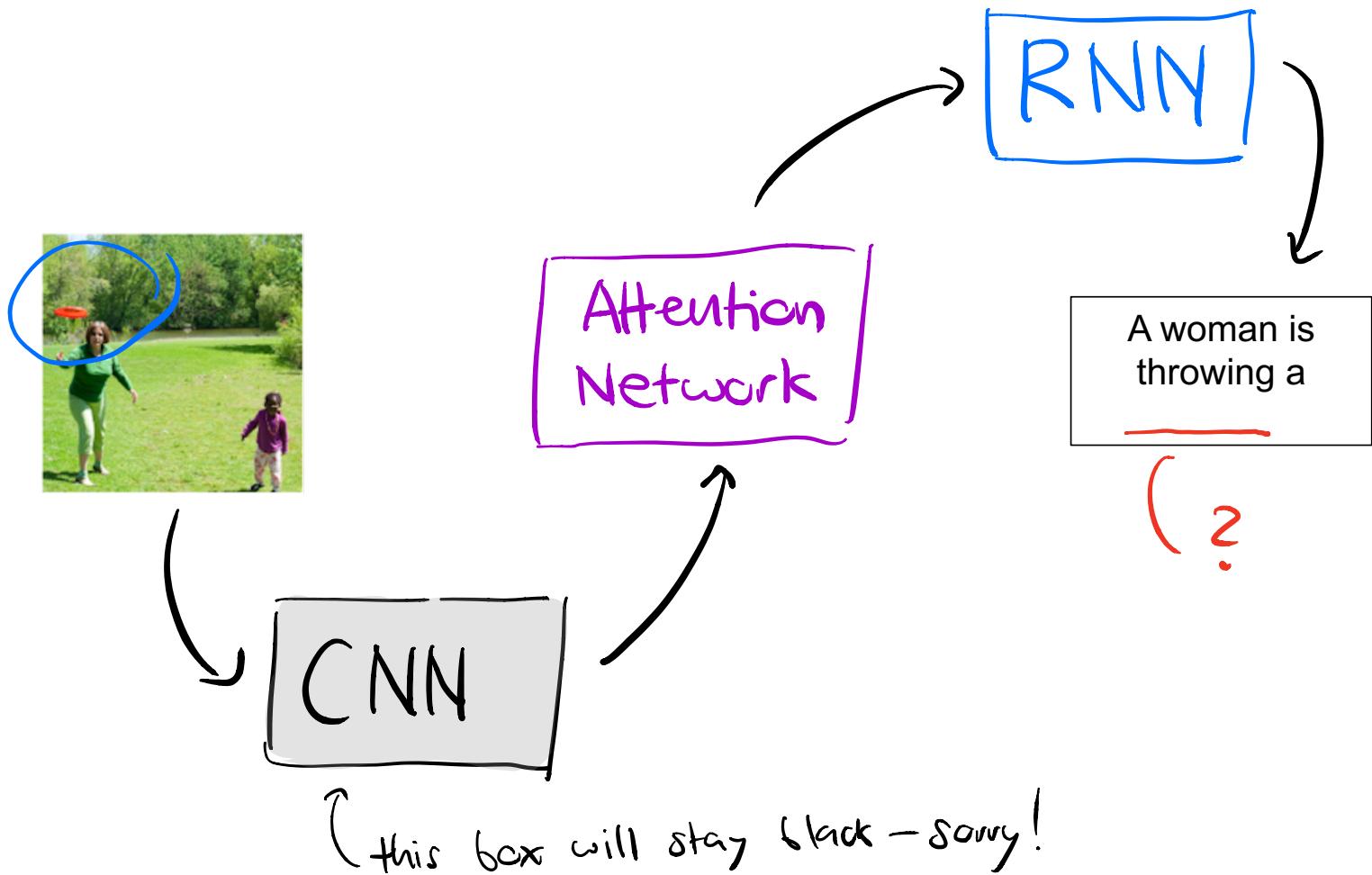
He is out of
his mind.

Er ist
ausser sich.

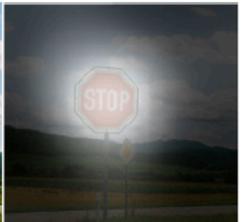


[:]

Attention networks: visual attention



Attention networks: visual attention



A woman is throwing a frisbee in a park.

A dog is standing on a hardwood floor.

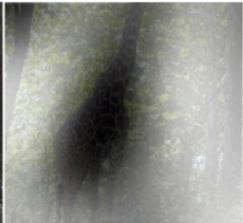
A stop sign is on a road with a mountain in the background.



A little girl sitting on a bed with a teddy bear.



A group of people sitting on a boat in the water.



A giraffe standing in a forest with trees in the background.

Attention networks: textual attention

input

Eine Frau wirft einen Frisbee in einem Park.



output

A woman is throwing a frisbee in a park.

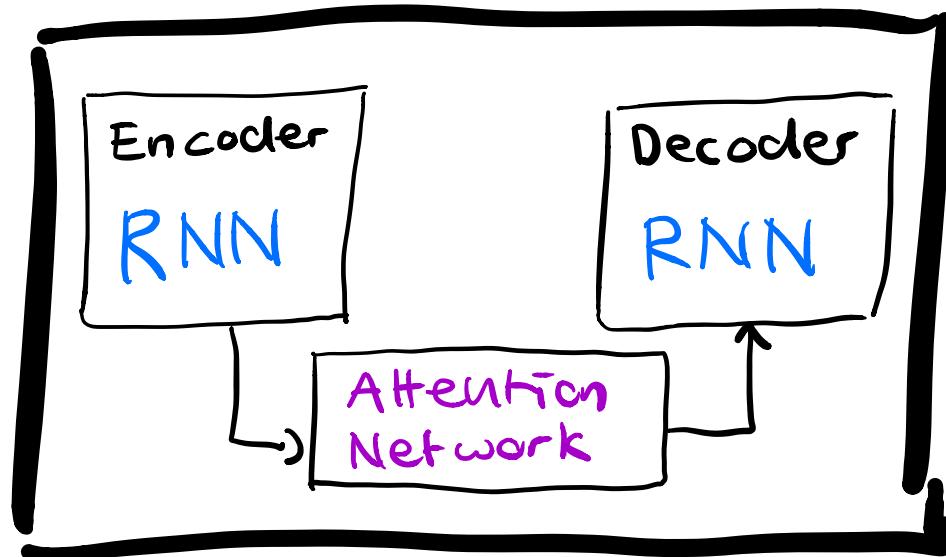
Review glass-box neural machine translation

He is out of
his mind.



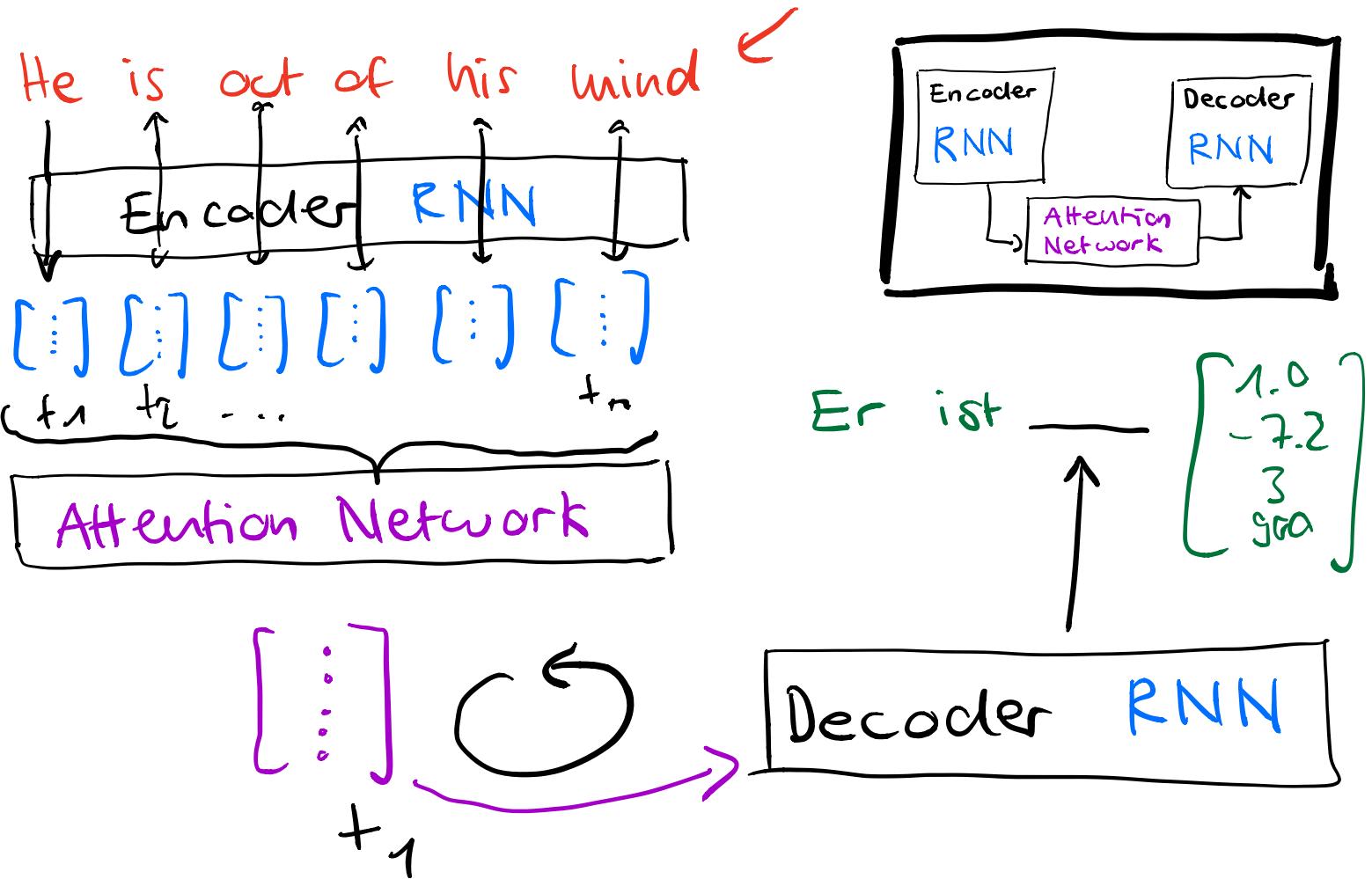
Er ist

ausser sich.



Encoder - Decoder with Att.

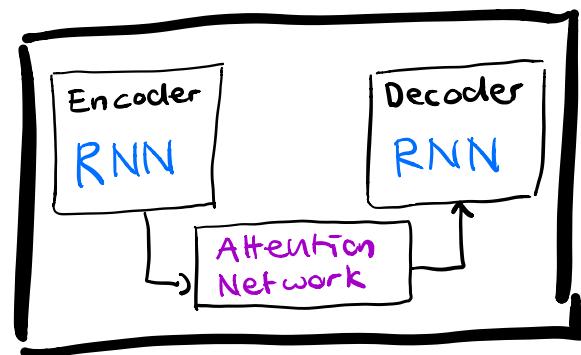
Review glass-box neural machine translation



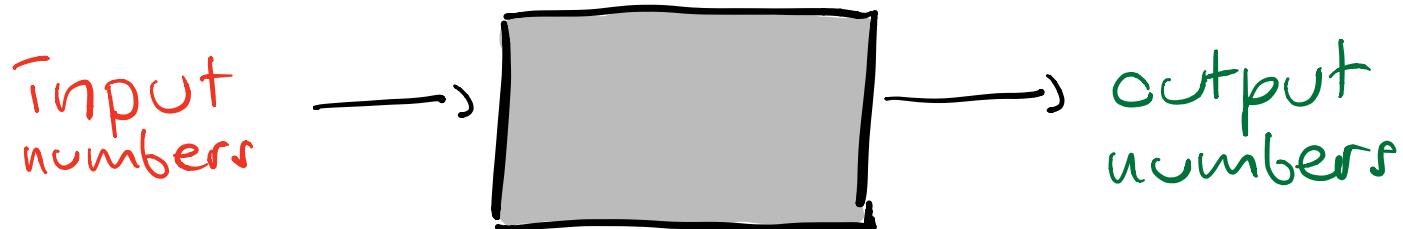
Why is this approach appealing?

- end-to-end optimization
- very general, broad class of problems

End-to-end optimization



Generalizability



speech
heart rate yesterday

sequence 2

speech
heart rate today

sequence

Summary

- Modern machine translation uses recurrent neural networks
- A popular setup is: RNN encoder, RNN decoder and attention network
- Approach very general: any sequence -to- sequence problem can be tackled
- I would love to hear your thoughts!

Thanks!



github.com/ZurichNLP/daikon

mmueller@cl.uzh.ch

Picture Credits

Sad little horse:
my own neural network

Google translates Russia as Mordor:

<https://www.washingtonpost.com/news/worldviews/wp/2016/01/05/where-is-russia-actually-mordor-in-the-world-of-google-translate/>

Vitreous Humour:

<https://visioneyeinstitute.com.au/wp-content/uploads/vitreous-humour.jpg>

Eating Carpet strictly prohibited:

https://www.upi.com/Odd_News/2015/08/26/Indian-airports-salacious-mistranslation-Eating-carpet-strictly-prohibited/4801440614077/

End-to-end slide:

<https://www.slideshare.net/silverroad/hris-optimizationmethods-to-improve-your-hr-system-capabilities>

Visual attention:

“Show, attend and tell”: <https://arxiv.org/pdf/1502.03044.pdf>

word embedding

"Haus" $\begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \end{bmatrix}$ $\begin{bmatrix} 7 \\ -1.2 \\ 0.9 \\ 1 \end{bmatrix}$

1-hot embedding

$$\begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \end{bmatrix} \longrightarrow \begin{bmatrix} 7 \\ -1.2 \\ 0.9 \\ 1 \end{bmatrix}$$

shack
house
shell

Trump
Obama