# Notes on Natural Language Processing with Transformers by Lewis Tunstall and Leandro von Wierra

compiled by D. Gueorguiev 3/30/24

## Introductory Notes

### The Encoder-Decoder Framework

State

RNN cell

Input

State

RNN cell

Input

State

RNN cell

Input

State

RNN cell

Input

State

RNN cell

Input

Figure: Unrolling an RNN in time

RNN cell

RNN cell

RNN cell

RNN cell

RNN cell

RNN cell

RNN cell

RNN cell

State

Transformers

aren’t

that

great

Transformers

sind

nicht

so toll

Figure: An encoder-decoder architecture with a pair of RNNs

## Appendix

### Unreasonable Effectiveness of Recurrent Neural Networks

#### RNNs

What makes Recurrent Networks special?

Limitation of Feed Forward Networks and also Convolutional Networks:

that their topology is too constrained: They accept a fixed size vector as input (for example an image) and produce a fixed-sized vector as an output (for example probabilities of different classes). This kind of networks also perform their mapping using a fixed amount of computational steps because there are fixed number of layers in their topology.

RNNs do not have some of these constraints. The topology of the RNNs allow to operate over a *sequence of vectors* rather than just individual vectors.

## References

## [Natural Language Processing with Transformers, Lewis Tunstall, Leandro von Werra, Thomas Wolf, 2022](https://github.com/dimitarpg13/transformers_intro/blob/main/articles_and_books/natural-language-processing-with-transformers-revised-edition-book.pdf)

[The Unreasonable Effectiveness of Recurrent Neural Networks, Andrej Karpathy blog, May 21, 2015](https://karpathy.github.io/2015/05/21/rnn-effectiveness/)