+ Up until now we lacked context

$$\hat{y_0} \quad \hat{y_1}$$
$$\uparrow \quad \uparrow$$
$$F \longrightarrow F \longrightarrow \dots \quad \text{Recurrent Neural Network}$$
$$\uparrow \quad \uparrow$$
$$(\widehat{x_0}) \quad x_1$$
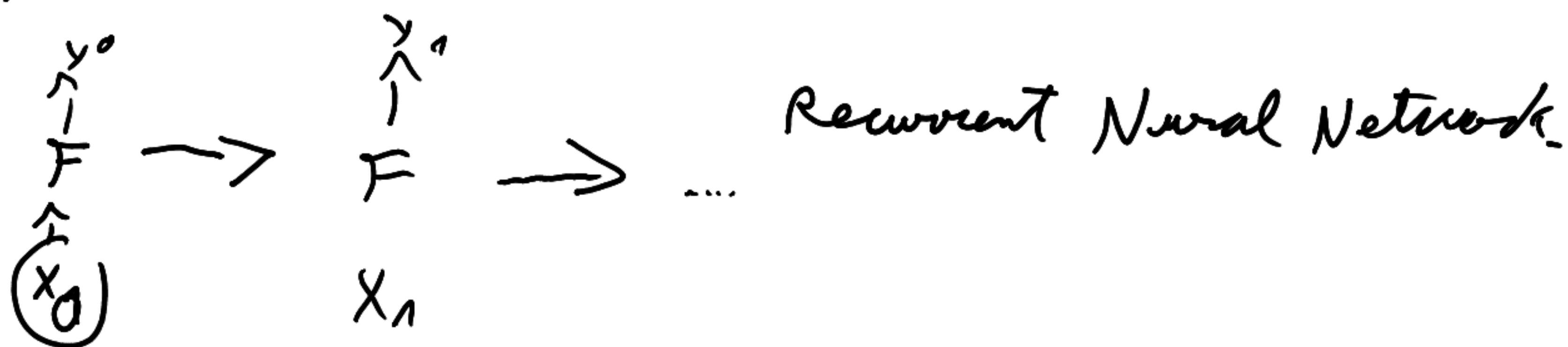
• Look into RNN.

+ <u>L S T M 's</u>

Builds on top of an RNN.
By passing a cell state on top of it
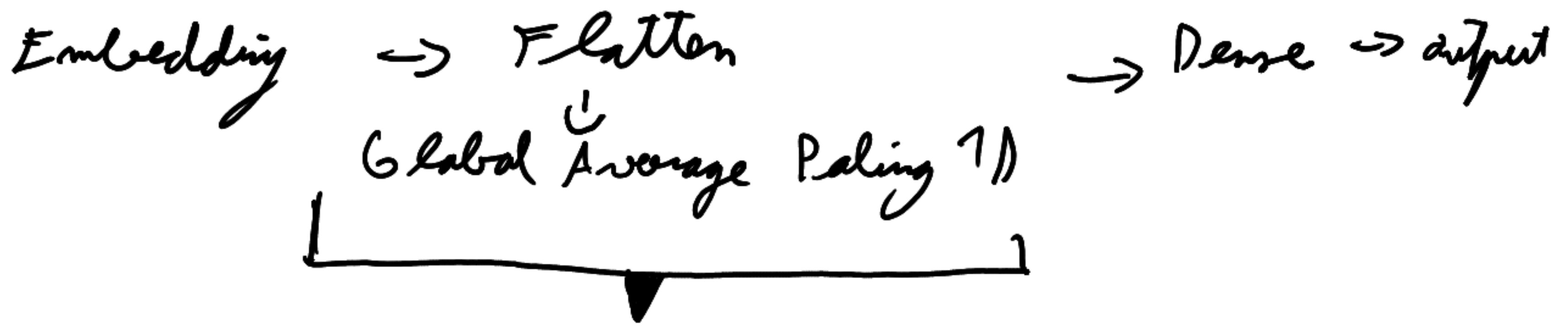in a bydirectional manner.

I.E. RNN ⟹ Value is kept but degrades,
LSTM ⟹ value is kept but the cell state
adds additional weights

1 layer vs 2 layer

2 -> very smooth, so we trust it
   more
   80% even with many out of vocabulary.
   Tommy

+ RNN with convolutions, Gated Recurrent Unit.

Embedding $\rightarrow$ Flatten $\rightarrow$ Dense $\Rightarrow$ output
                    Global Average Paling 1D)
```
        |_____|
                        ▼
```
            This can be messed around with
For example: Bidirectional (LSTM)

Overfitting happens easily in LSTM

Another layer Type: Conv 1D $\rightarrow$
                    $\rightarrow$ Global max pooling layer.

When using convolution we are applying a
filter

Will always out of Vocab words.