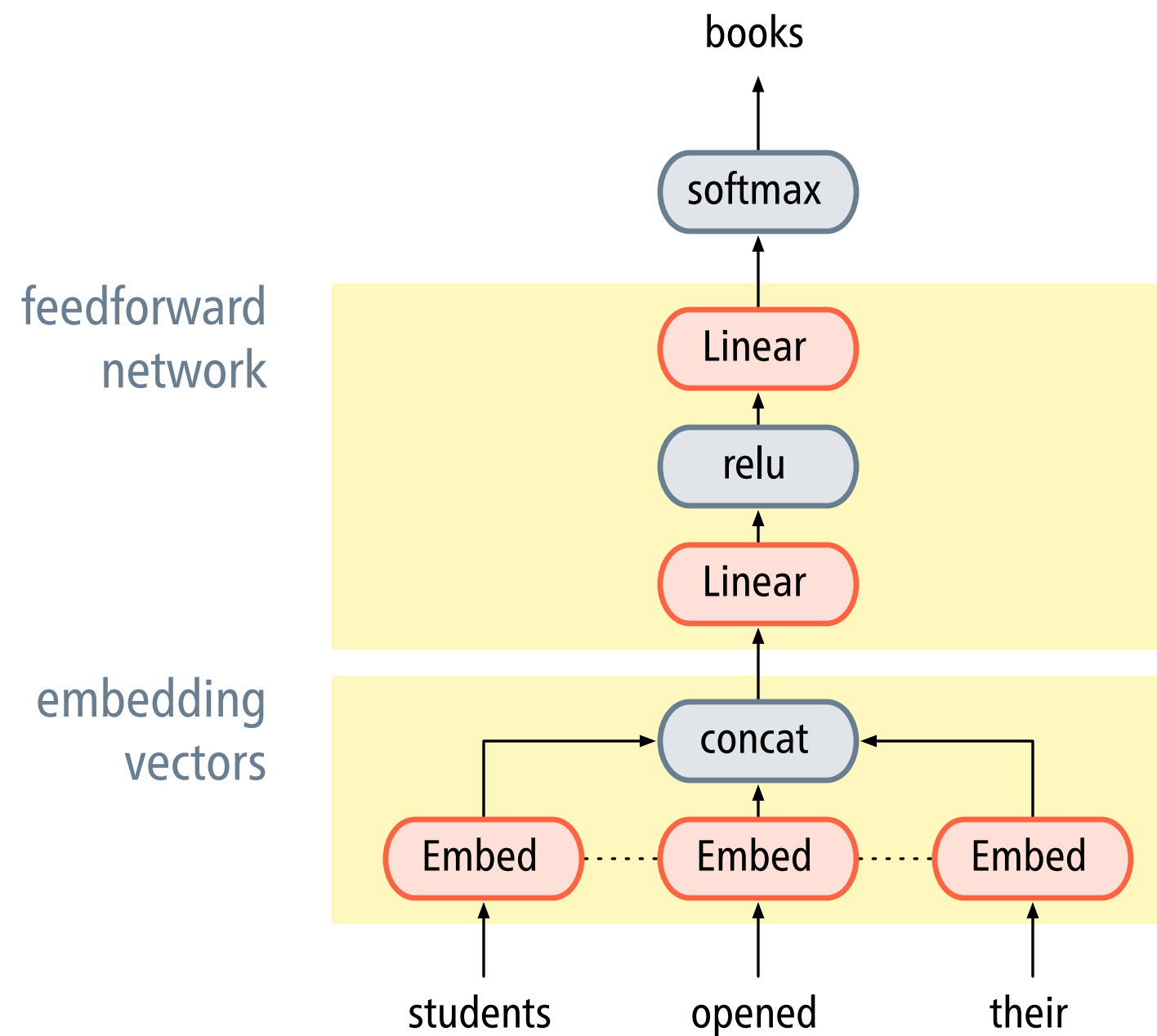


RNN language models

Marco Kuhlmann

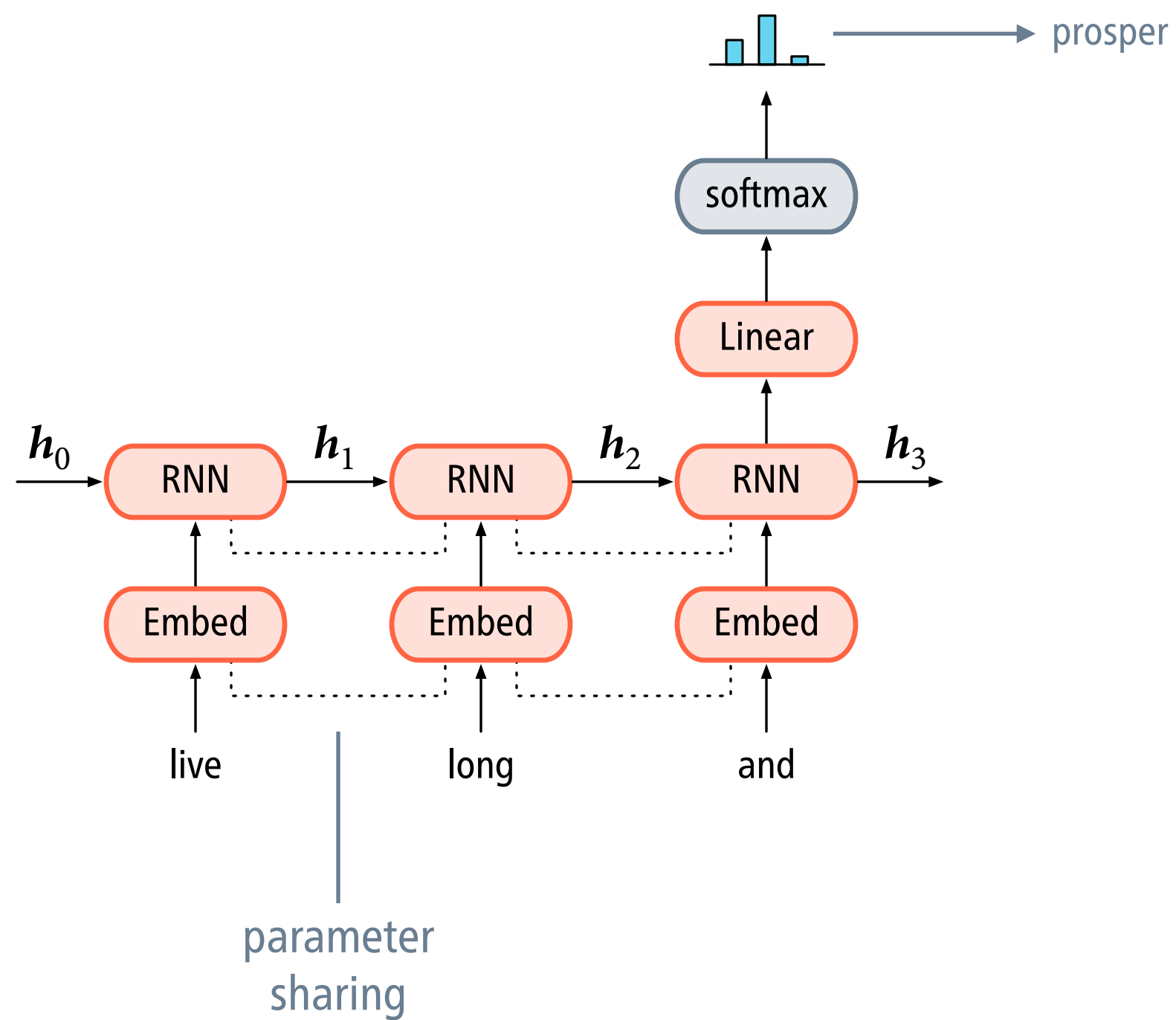
Department of Computer and Information Science

Reminder: Neural n-gram model



[Bengio et al. \(2003\)](#)

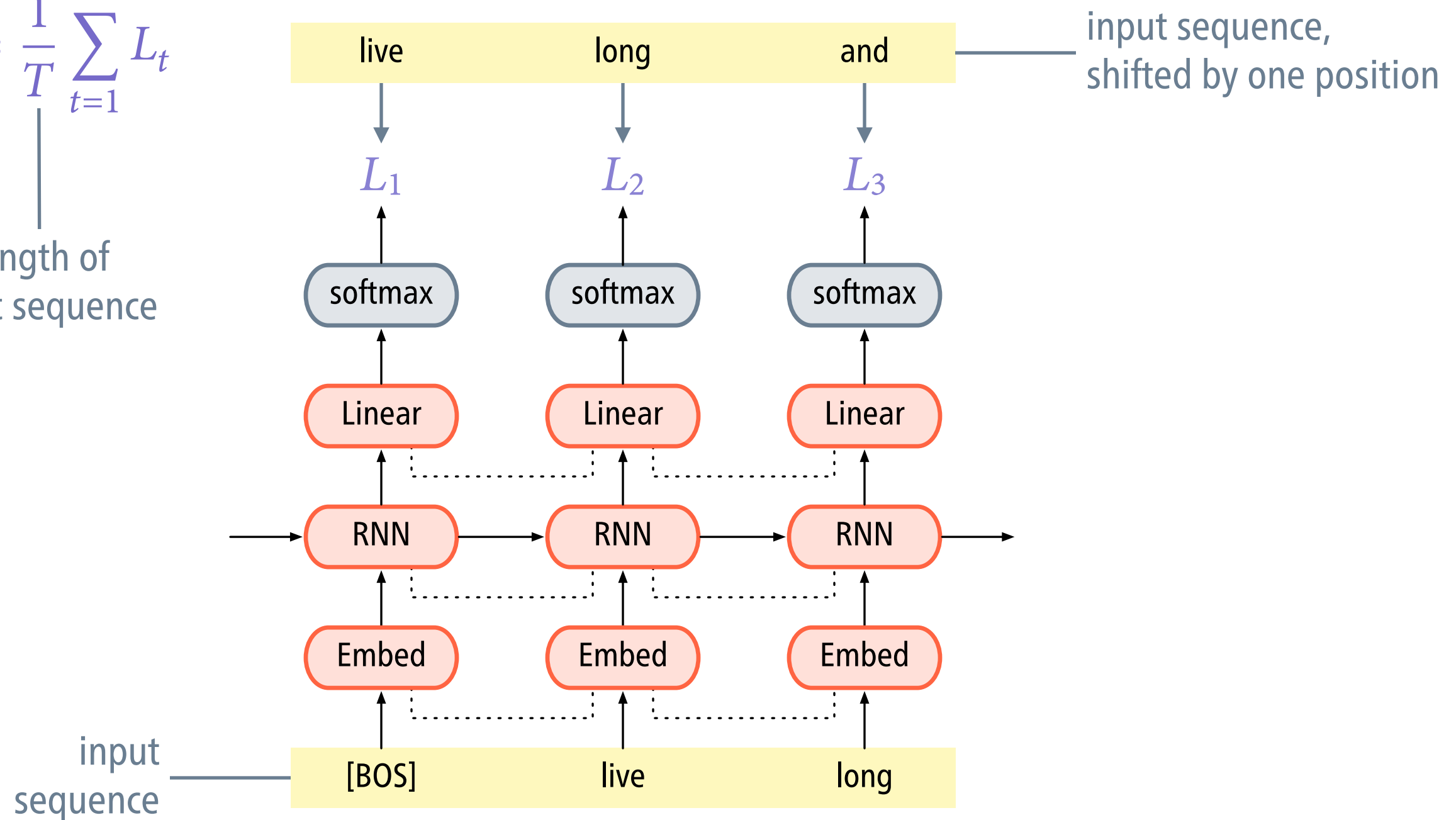
RNN language model



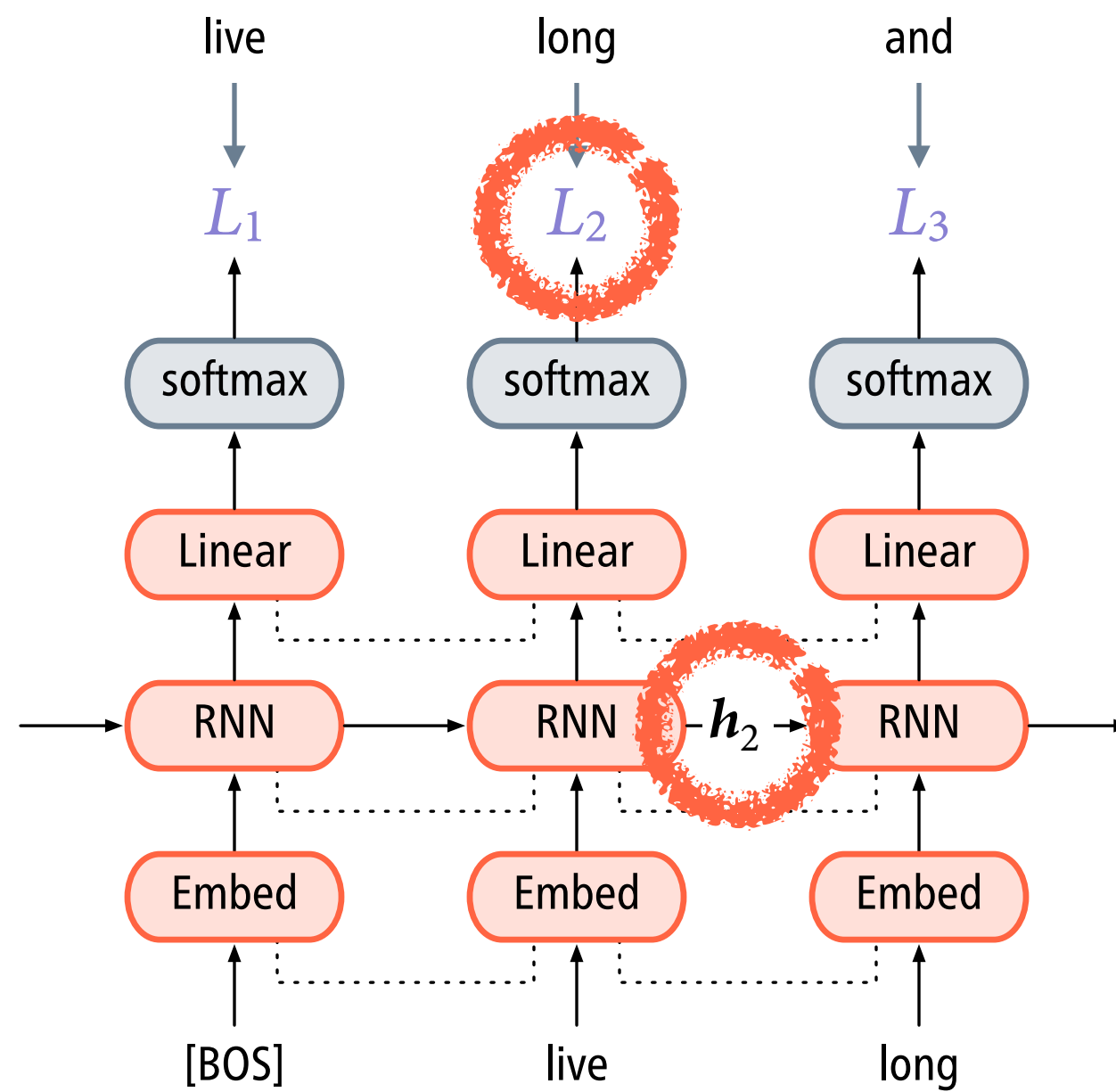
Training RNN language models

$$L = \frac{1}{T} \sum_{t=1}^T L_t$$

length of input sequence



Teacher forcing



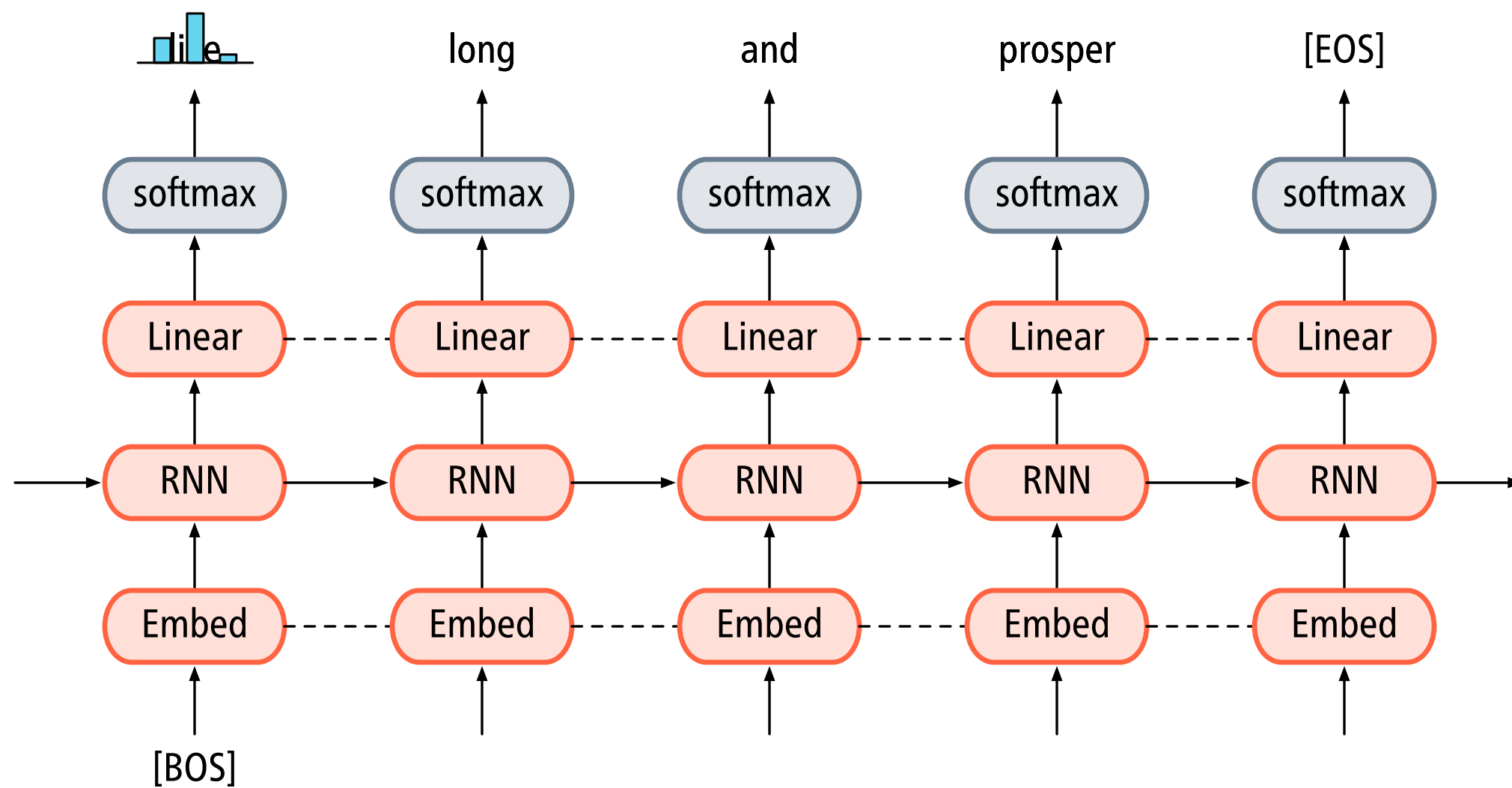
Practical issues when training RNNs

- In practice, backpropagation through time is often truncated after some fixed number of steps.

corresponds to truncating the input sequence

- When creating batches of sequences with unequal lengths, shorter sequences need to be padded.
- To minimise padding, we can sort sequences by length, and try to create minibatches in which sequences have the same length.

Generating text with an RNN language model



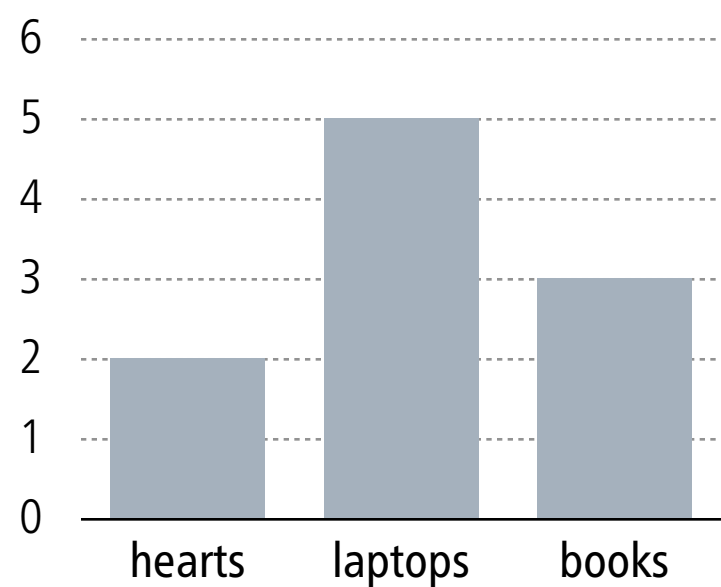
Sampling temperature

- The **temperature** parameter $T > 0$ controls how random the output of a language model is. It is applied to the softmax logits:

$$\text{softmax}(\mathbf{z})[k] = \frac{\exp(\mathbf{z}[k]/T)}{\sum_i \exp(\mathbf{z}[i]/T)}$$

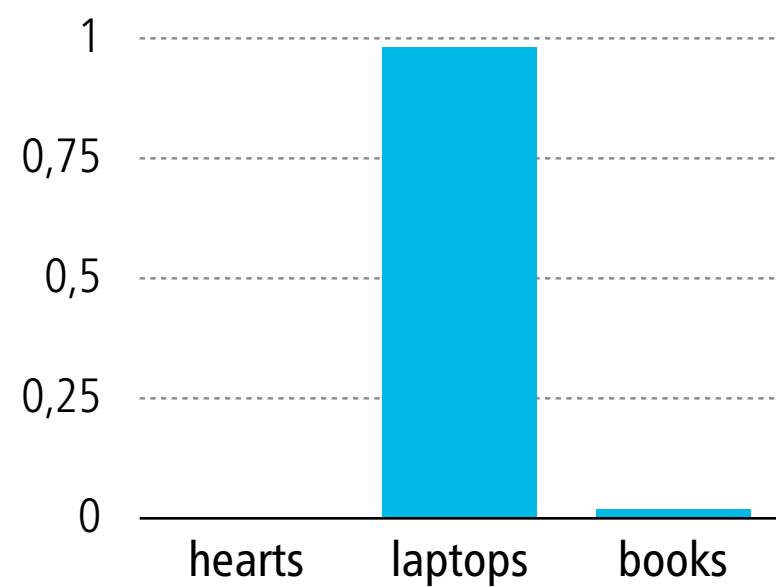
- The higher the temperature, the more uniform the distribution, and the more “creative” the output.

Impact of different sampling temperatures

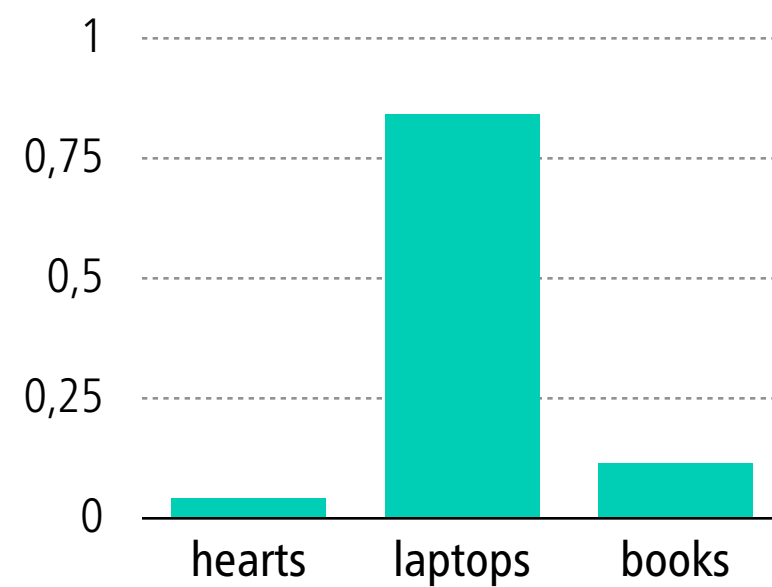


logits

T = 0.5



T = 1.0



T = 5.0

