



CMPT 413/713: Natural Language Processing

Sequence to Sequence Models (Seq2Seq)

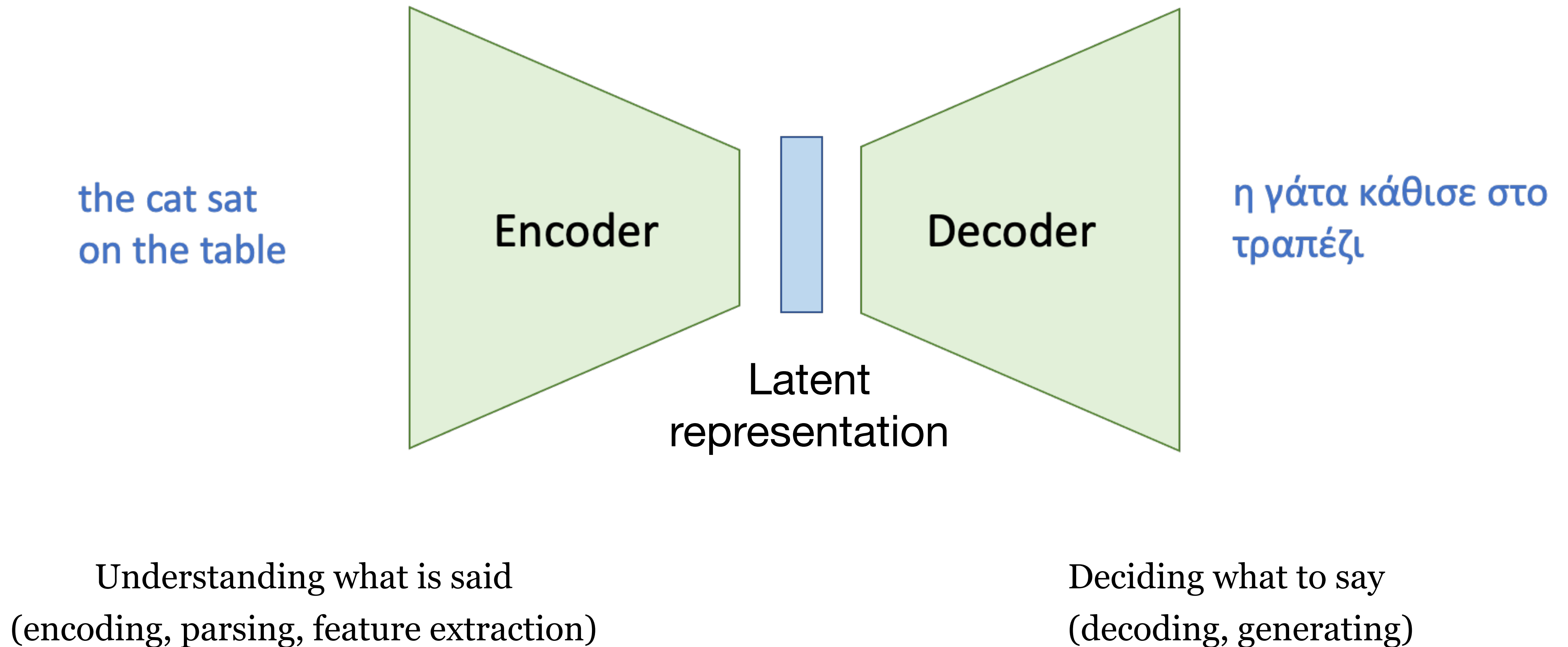
Spring 2026
2026-02-02

Adapted from slides from Danqi Chen and Karthik Narasimhan
(with some content from slides from Abigail See, Graham Neubig)

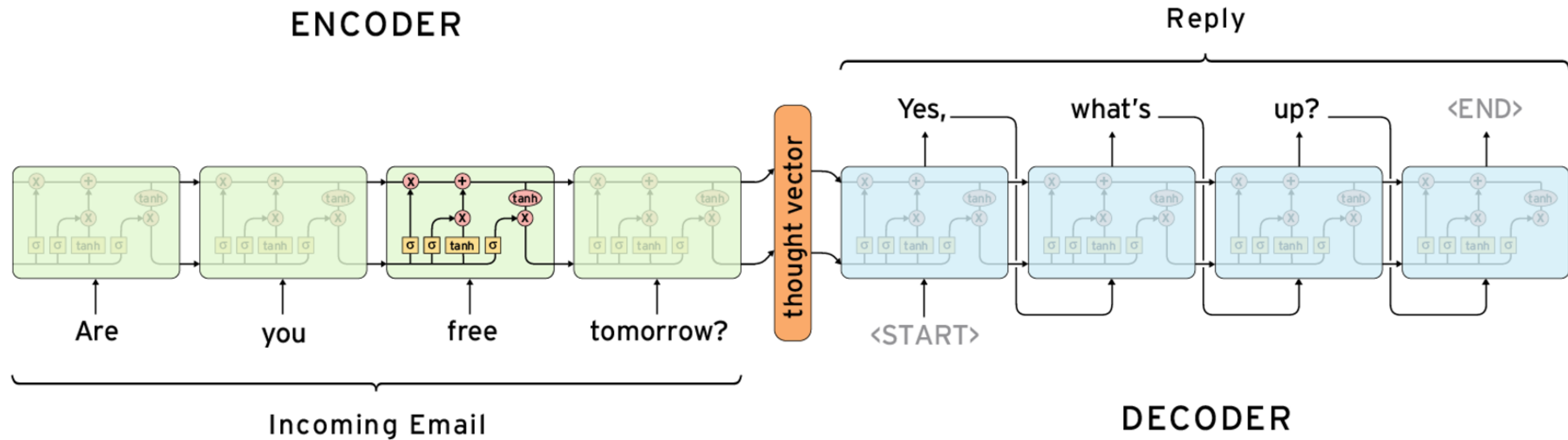
Overview

- Sequence generation tasks
- Seq2Seq models - Encoder/Decoder
- Decoding strategies
- Evaluating text generation
- Attention

Want computer friendly representation for applications



Encoder-Decoder Model



Seq2Seq Tasks and Applications

Task/Application	Input	Output
Machine Translation	French	English
Summarization	Document	Short Summary
Dialogue	Utterance	Response
Parsing	Sentence	Parse tree (as sequence)
Question Answering	Context + Question	Answer

Cross-Modal Seq2Seq

Task/Application	Input	Output
Speech Recognition	Speech Signal	Transcript
Image Captioning	Image	Text
Video Captioning	Video	Text
Vision-Language Navigation	Text	Actions

Seq2Seq Tasks and Applications

Task/Application	Input	Output
Machine Translation	French	English
Summarization	Document	Short Summary
Dialogue	Utterance	Response
Parsing	Sentence	Parse tree (as sequence)
Question Answering	Context + Question	Answer

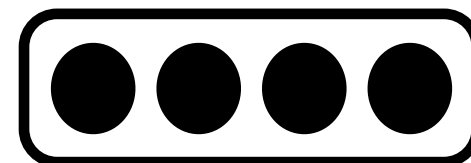
Neural Machine Translation

- ▶ A **single neural network** is used to translate from source to target
- ▶ Architecture: Encoder-Decoder
- ▶ Two main components:
 - ▶ **Encoder:** Convert source sentence (input) into a vector/matrix
 - ▶ **Decoder:** Convert encoding into a sentence in target language (output)

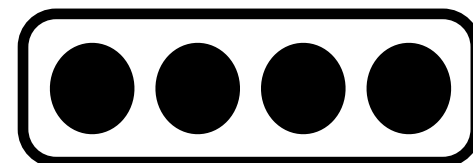
Encoder

Sentence: This cat is cute

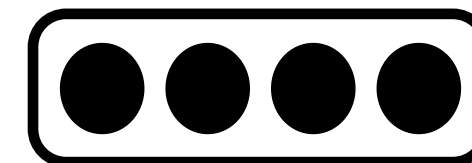
word
embedding



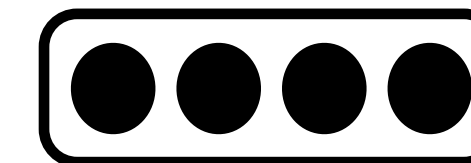
This



cat



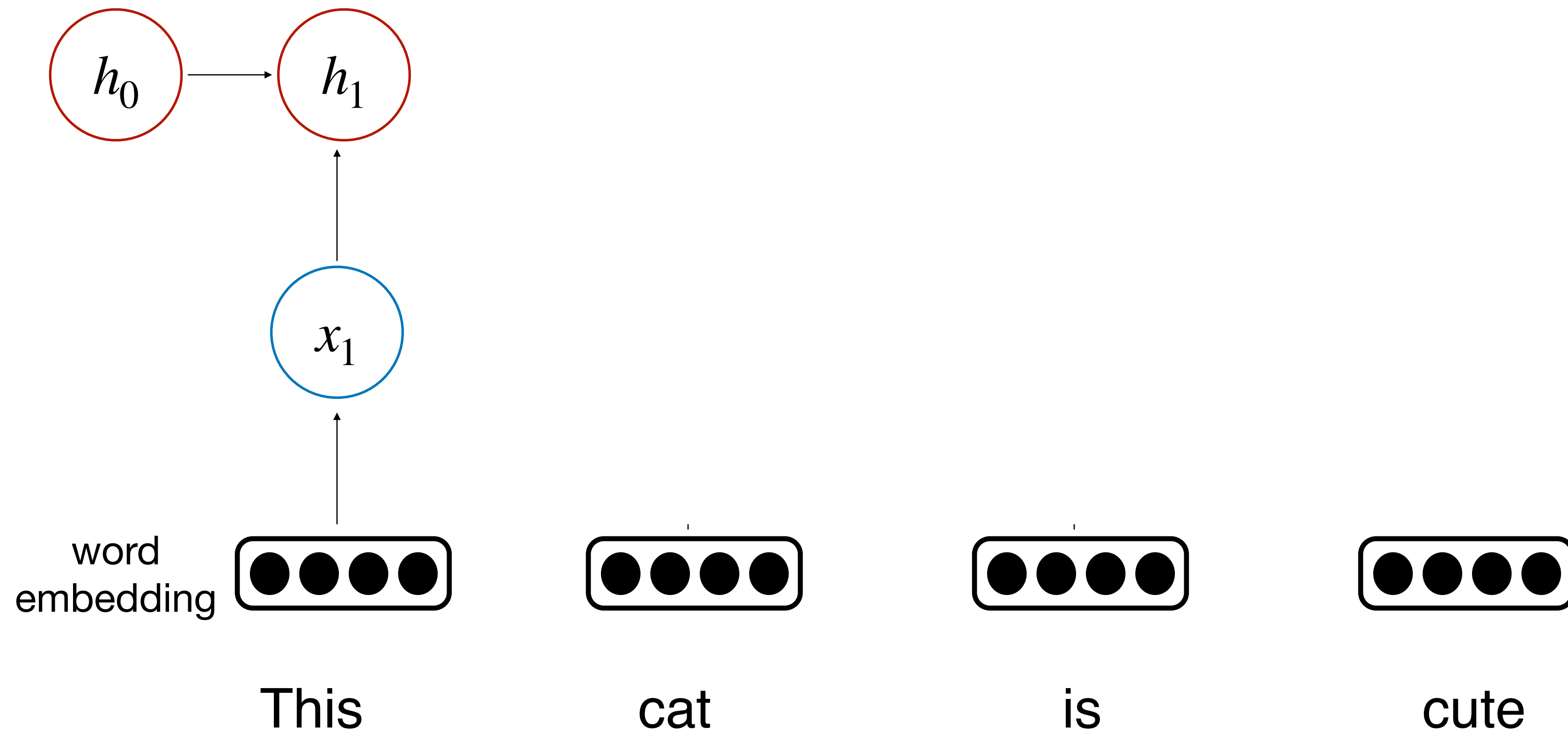
is



cute

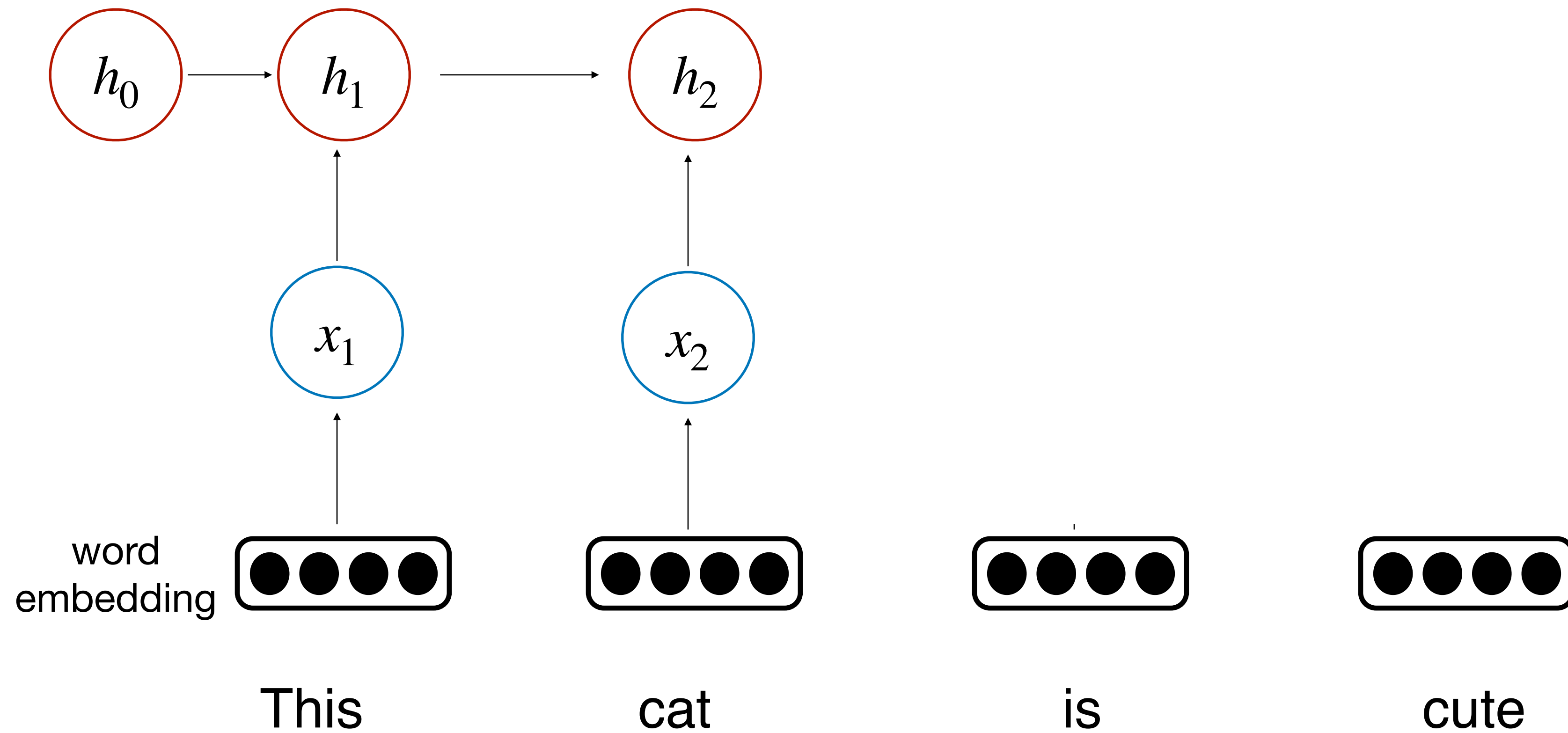
Encoder

Sentence: This cat is cute



Encoder

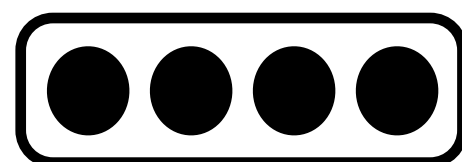
Sentence: This cat is cute



Decoder

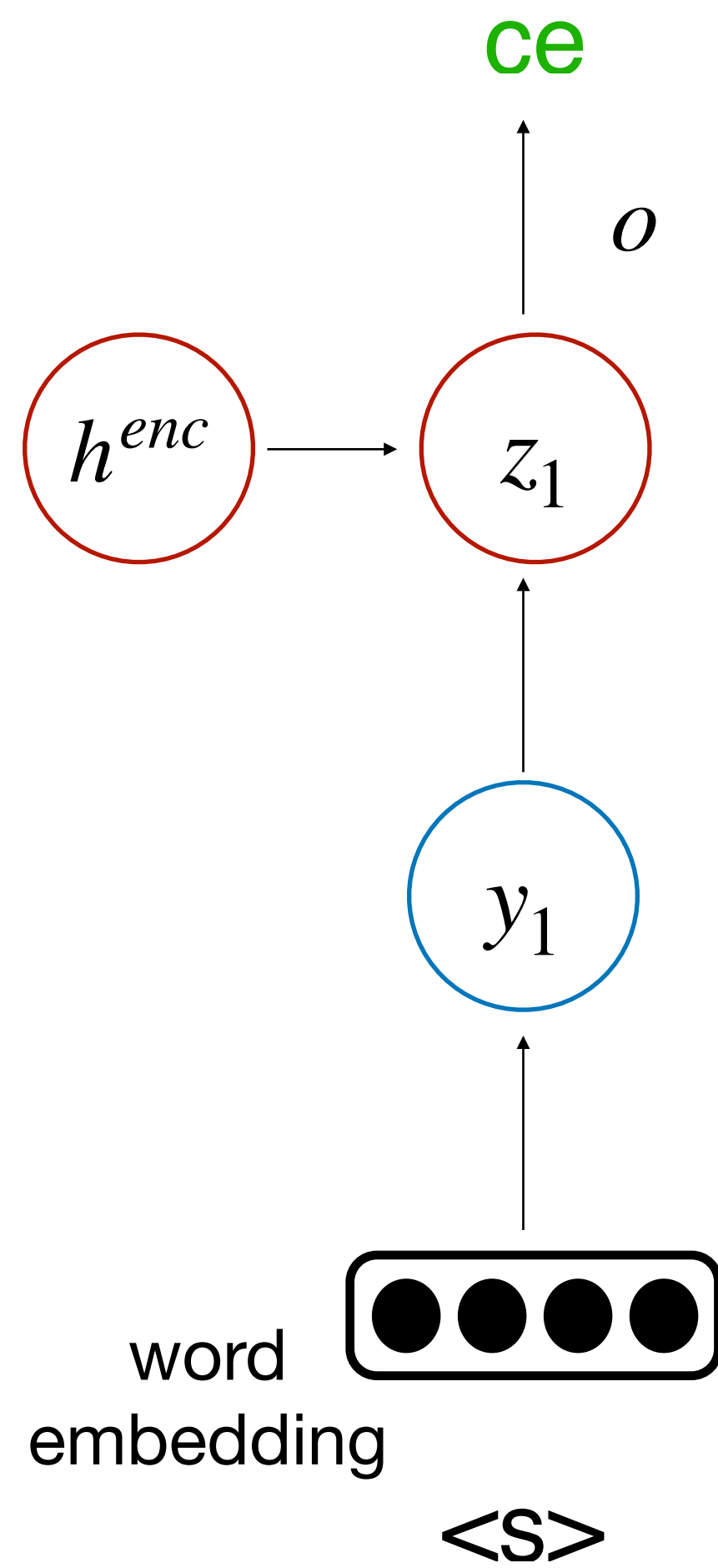
h^{enc}

word
embedding

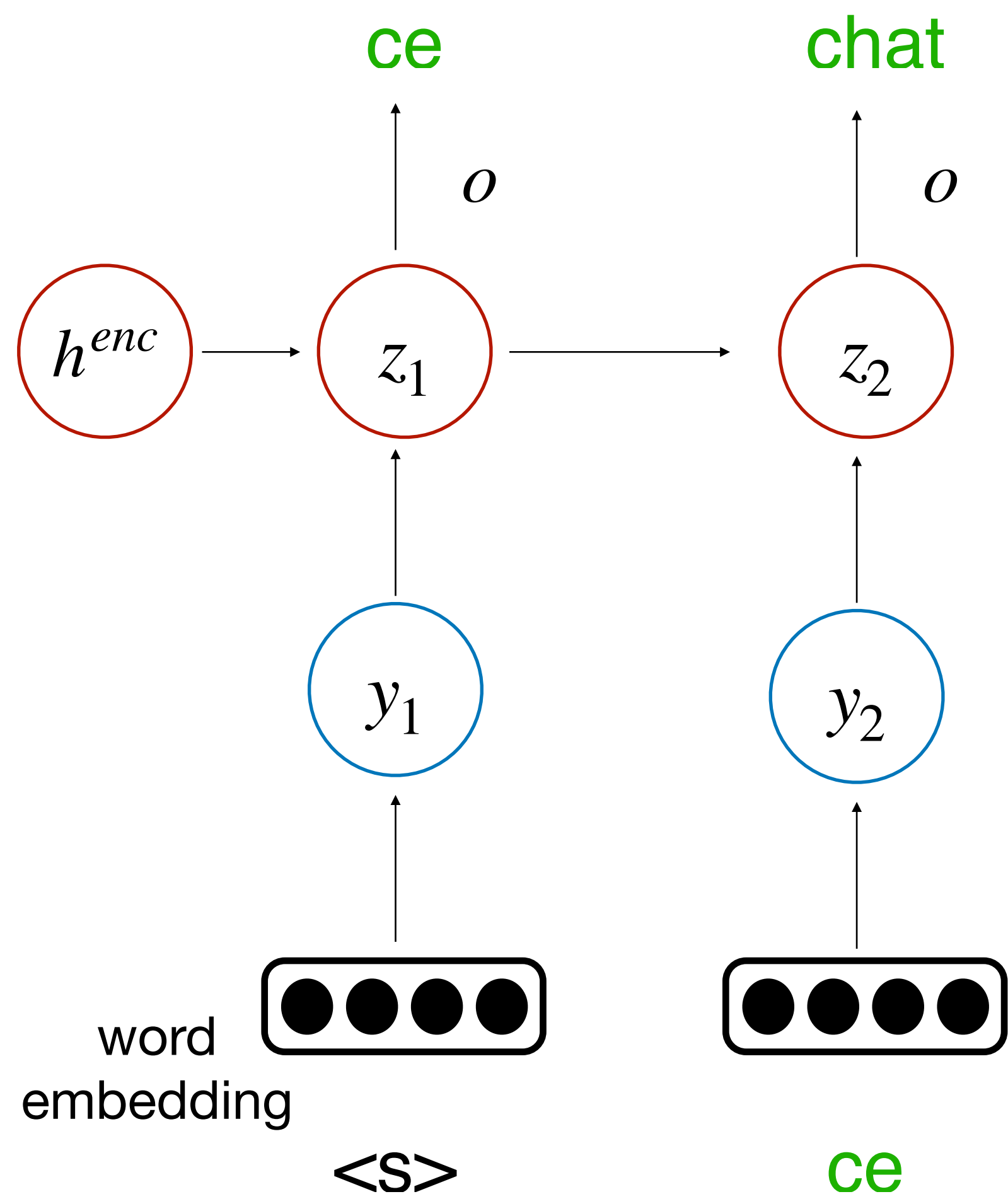


<S>

Decoder



Decoder



Efficient Training: Batching

- Apply RNNs to batches of sequences
- Present data as 3D tensor of $(T \times B \times F)$
- Use mask matrix to aid with computations that ignore padded zeros

Padded sequences

1	1	1	1	0	0
1	0	0	0	0	0
1	1	1	1	1	1
1	1	1	0	0	0

Lengths

4
1
6
3

