

Assignment 2

NLP Project

Tim Rocktäschel & Sebastian Riedel
COMP0087 Natural Language Processing



Submission Deadline

- 16:00 (UK time) on Wednesday 12 May 2021

Project Checklist

- Do you have a **research question**?
 - Is it possible to do X?
 - How can we do better on X?
 - Why does X work?
 - ...
- Do you have a **research hypothesis**?
 - X is possible if we do Y
 - X works because of Y
 - ...
- Do you know **why** this question matters?

Project Checklist

- Do you have an **empirical** question (that requires NLP to answer)
- Do you have the right **data**?
- Do you have a way to **measure** success?
- Do you have identified the **risks**?
- Do you have identified means to **mitigate** the risks?
- Do you have evidence for **feasibility**?
- Is your idea truly **novel**? What previous work exists?
- Do you have the right **tools**?
- Do you have a **timeline** and assignment of work (packages)?

Other Things

- Things/Ideas don't have to work (and they often don't)
- But you should analyse and explain why (we want to learn something)
- Also ... Look at the data

Managing risks

- "We spent three weeks trying to get the existing framework **X** to work."
- "We asked researcher **X** to send us code for their paper and they never replied."
- "We used DL library **X** and nobody helped us when we ran into problems."
- "We tried a bunch of stuff and nothing worked."
- "We didn't have enough training data."
- "We couldn't get access to enough GPUs to make it work."
- "Our team member didn't do any work so we had a disadvantage."
- "Our team member left UCL to do a startup and now we have a disadvantage."

Fame awaits...

MIT Technology Review



8 October 2019

Artificial Intelligence / Machine Learning

Instead of practicing, this AI mastered chess by reading about it

Machines that appreciate “brilliant” and “dumb” chess moves could learn to play the game—and do other things—more efficiently.

by Will Knight

Jul 31, 2019



MSc Machine Learning students achieved double success with published research paper; “SentiMATE: Learning to play Chess through Natural Language Processing”, presented at AIIDE-19. It also received a write-up in MIT Technology Review.

Some previous projects

- Review of Dropout Schemes in Recurrent Neural Networks
- Abstractive Text Summarisation using Pointer-Generator Network
- BERT meets Goldilocks: Pay Cloze Attention to Transfer Learning on the Children's Book Test
- Handle With Poincaré: Question Answering in Hyperbolic Space
- Using Graph Convolutional Networks for Abstractive Text Summarization
- Topic-Conditioned CNN for wikiHow Text Summarization
- Comparing the performance of RNNs and CNNs for Fake News Detection
- An Exploratory Study of How Political Orientation Affects Language Use and Communication

Ethics

- Before you commit to a project, please send a 100 word abstract of your proposed work
- This will help us to understand and approve the ethical implications of your projects
- We also aim to give light feedback if we see any obvious practical pitfalls
- Please include a description of the data you plan to use
- **Use the private Forum for this**

Rules for publishing

- We encourage submission of outstanding work to conferences/workshops
- However, do not submit without consulting and involving us
- We are happy to help you identify whether your project qualifies, and what needs to be done to improve your chances of acceptance
- **Do not submit your coursework to arXiv!**
- **Seriously!**
- **Don't do it!**



Thomas G. Dietterich @tdietterich · Mar 20, 2019
Dear academic colleagues. It is not appropriate for class projects to be submitted to [@arxiv_org](#) unless they are also being submitted for publication. I and my fellow moderators are seeing more and more of this behavior. It wastes our time and the time of your colleagues

20 210 977

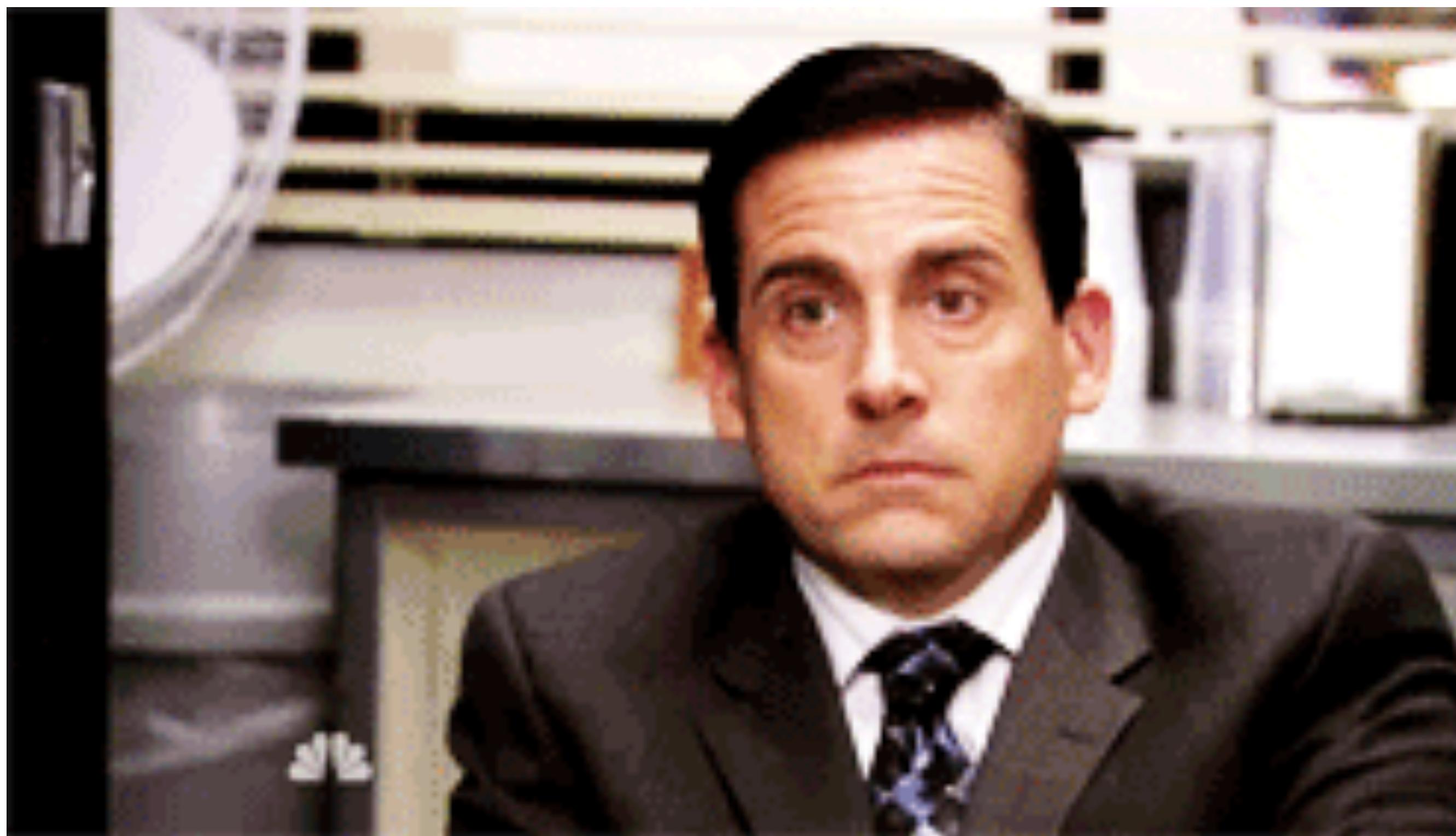
Rebecca Mason @adhdphd · Mar 20, 2019
Is there a policy against this? I've seen undergraduate class projects on arxiv, but it seems as though students at certain universities are more likely to get them in than others.

2 1

Thomas G. Dietterich
@tdietterich

missions in the form of an article by a conventional publication that contain original or substantive undergraduate research, course proposals may be removed."

We're extremely excited to see
what you'll come up with!

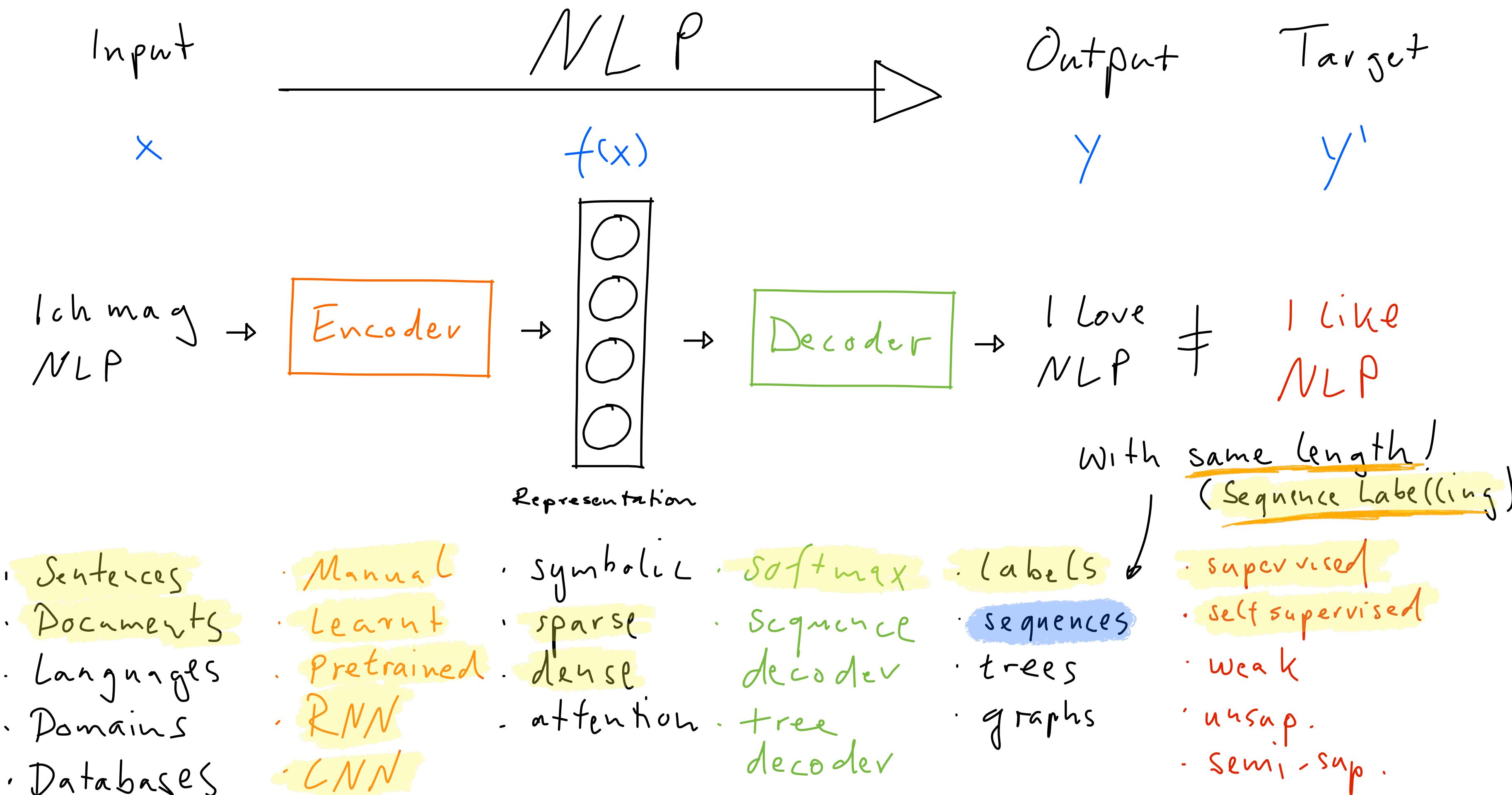


Conditional Language Models

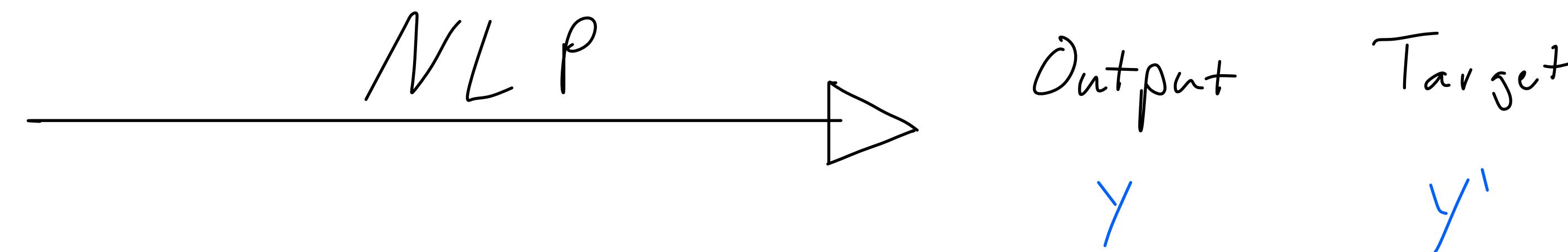
Tim Rocktäschel & **Sebastian Riedel**
COMP0087 Natural Language Processing



NLP in a Nutshell



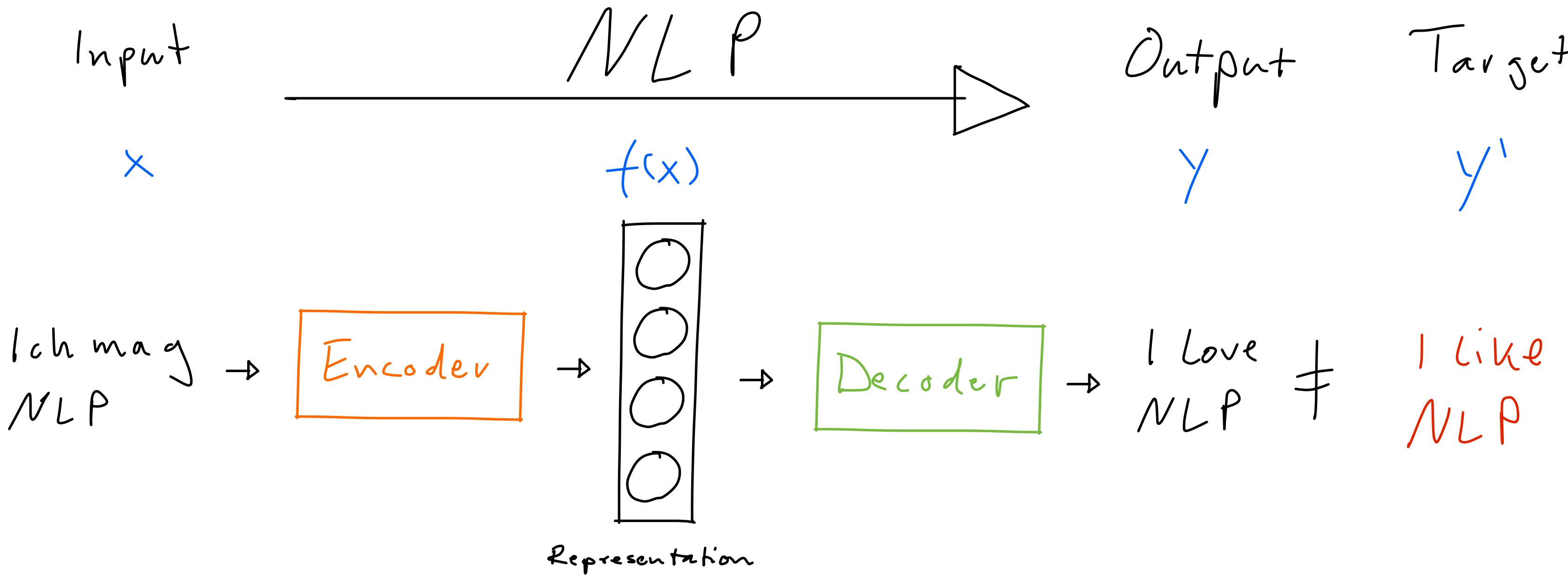
Language Models



→ Decoder → I Love
NLP ≠ I like
NLP

- labels
- sequences
- trees
- graphs

Conditional Language Models



- Sentences
- Documents
- Languages
- Domains
- Databases

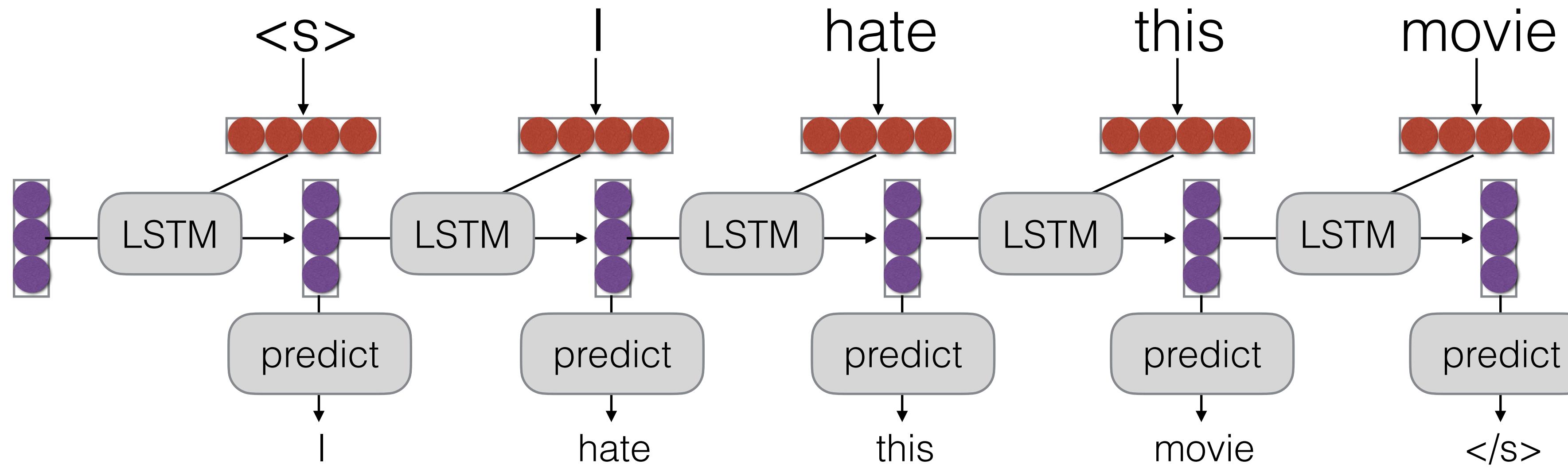
- Labels
- Sequences
- trees
- graphs

Conditioned Language Models

- Not just generate text, generate text according to some specification

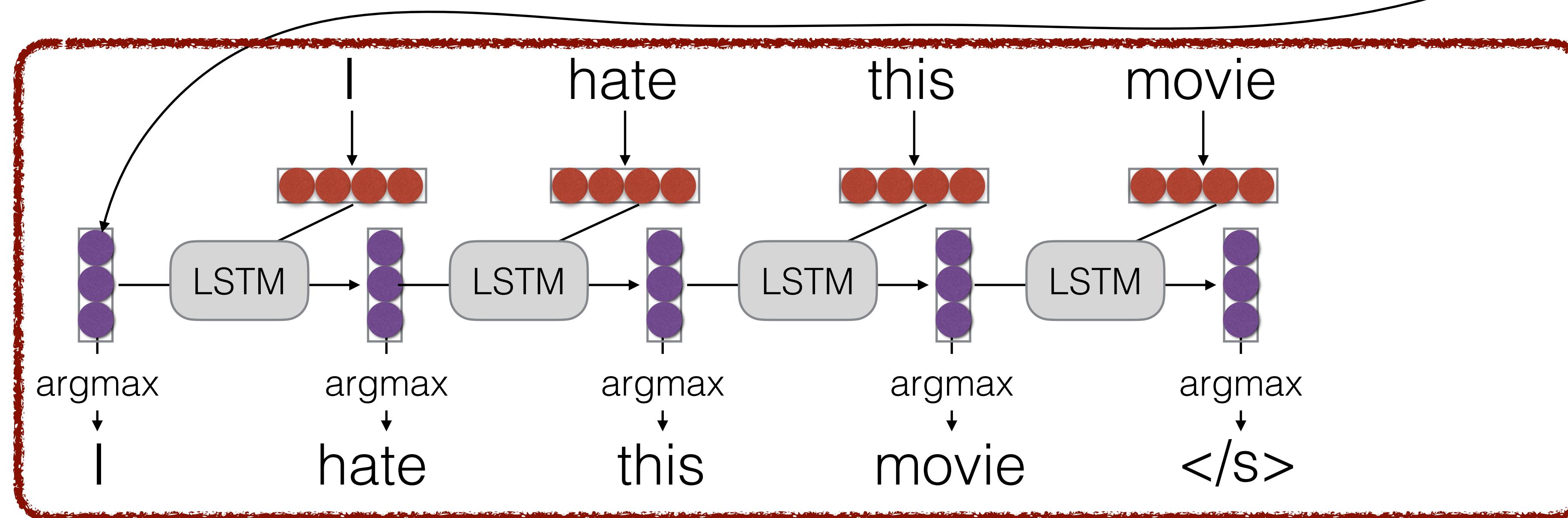
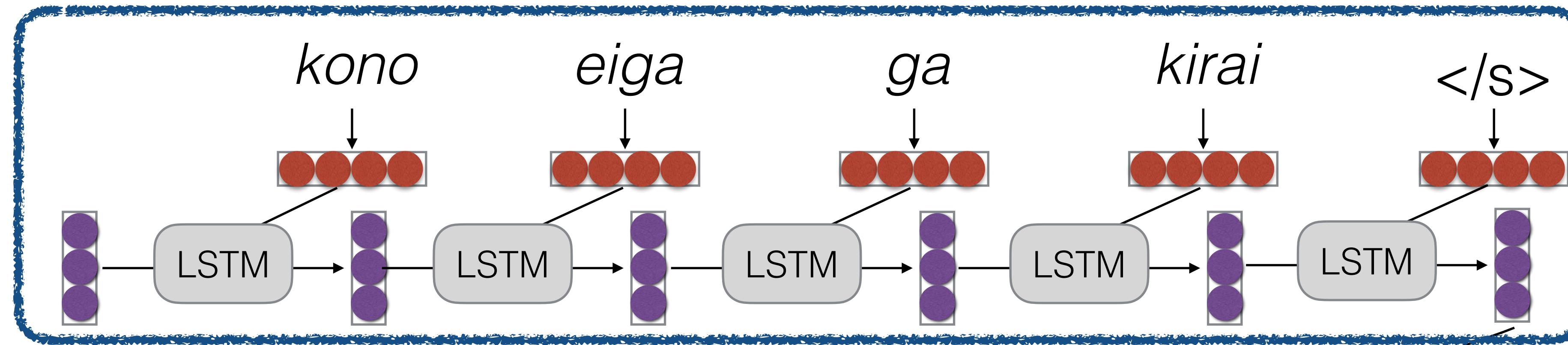
| <u>Input X</u> | <u>Output Y (Text)</u> | <u>Task</u> |
|-----------------|------------------------|---------------------|
| Structured Data | NL Description | NL Generation |
| English | Japanese | Translation |
| Document | Short Description | Summarization |
| Utterance | Response | Response Generation |
| Image | Text | Image Captioning |
| Speech | Transcript | Speech Recognition |

(One Type of) Language Model (Mikolov et al. 2011)



(One Type of) Conditional Language Model (Sutskever et al. 2014)

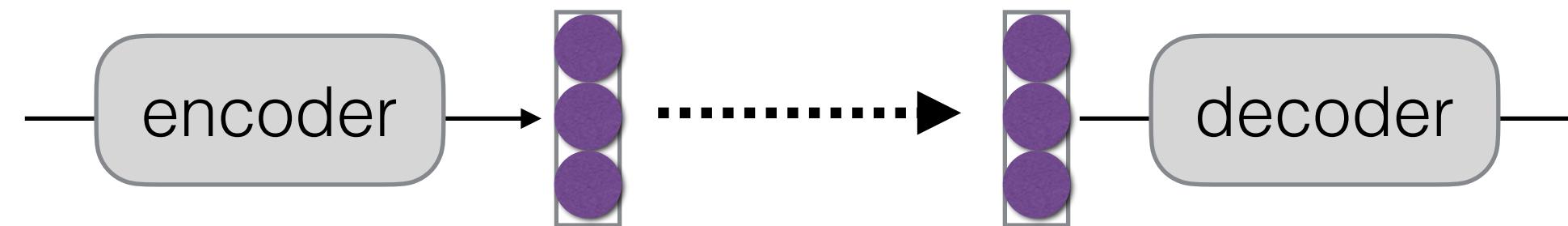
Encoder



Decoder

How to Pass Hidden State?

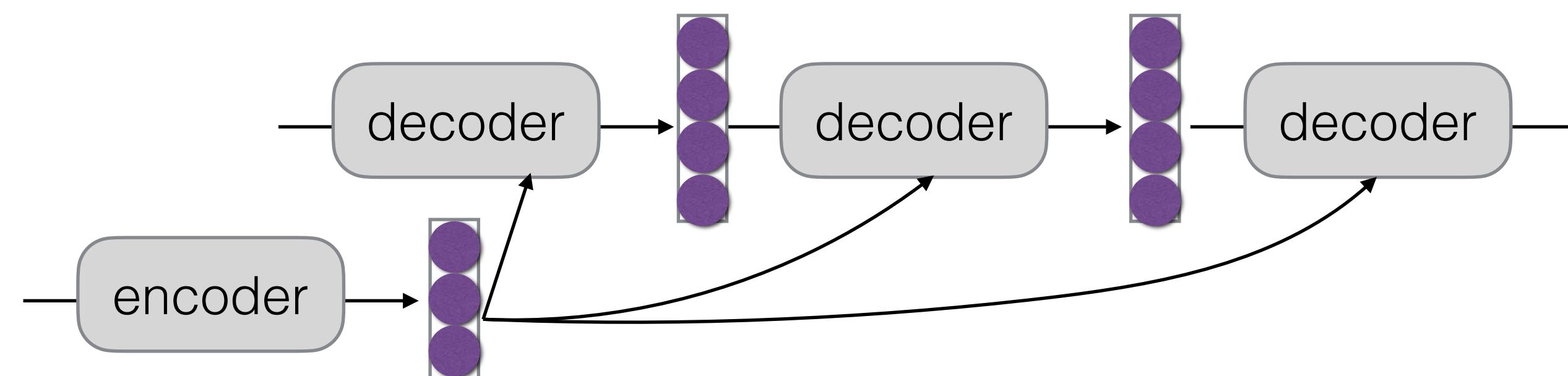
- Initialize decoder w/ encoder (Sutskever et al. 2014)



- Transform (can be different dimensions)



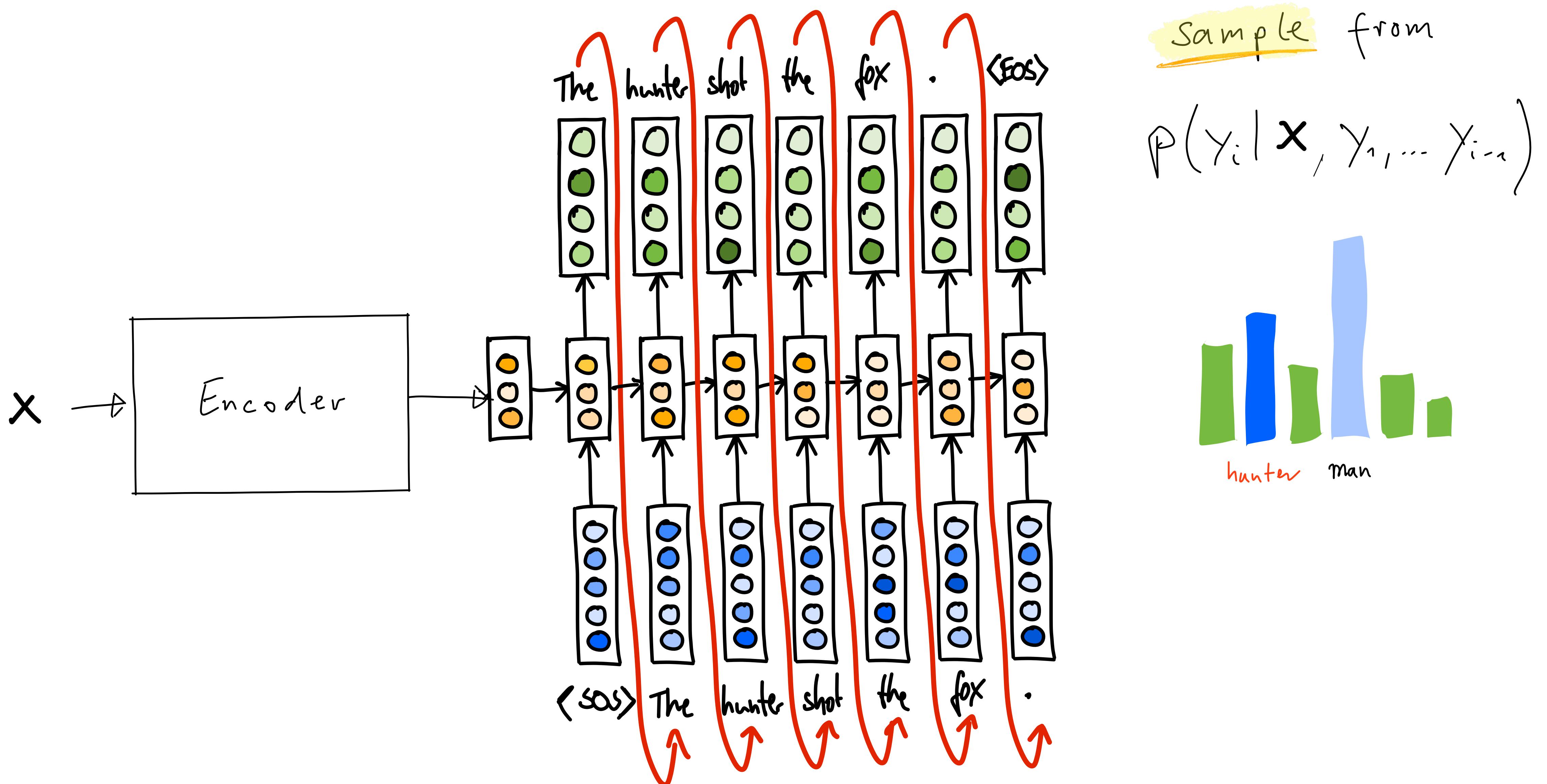
- Input at every time step (Kalchbrenner & Blunsom 2013)



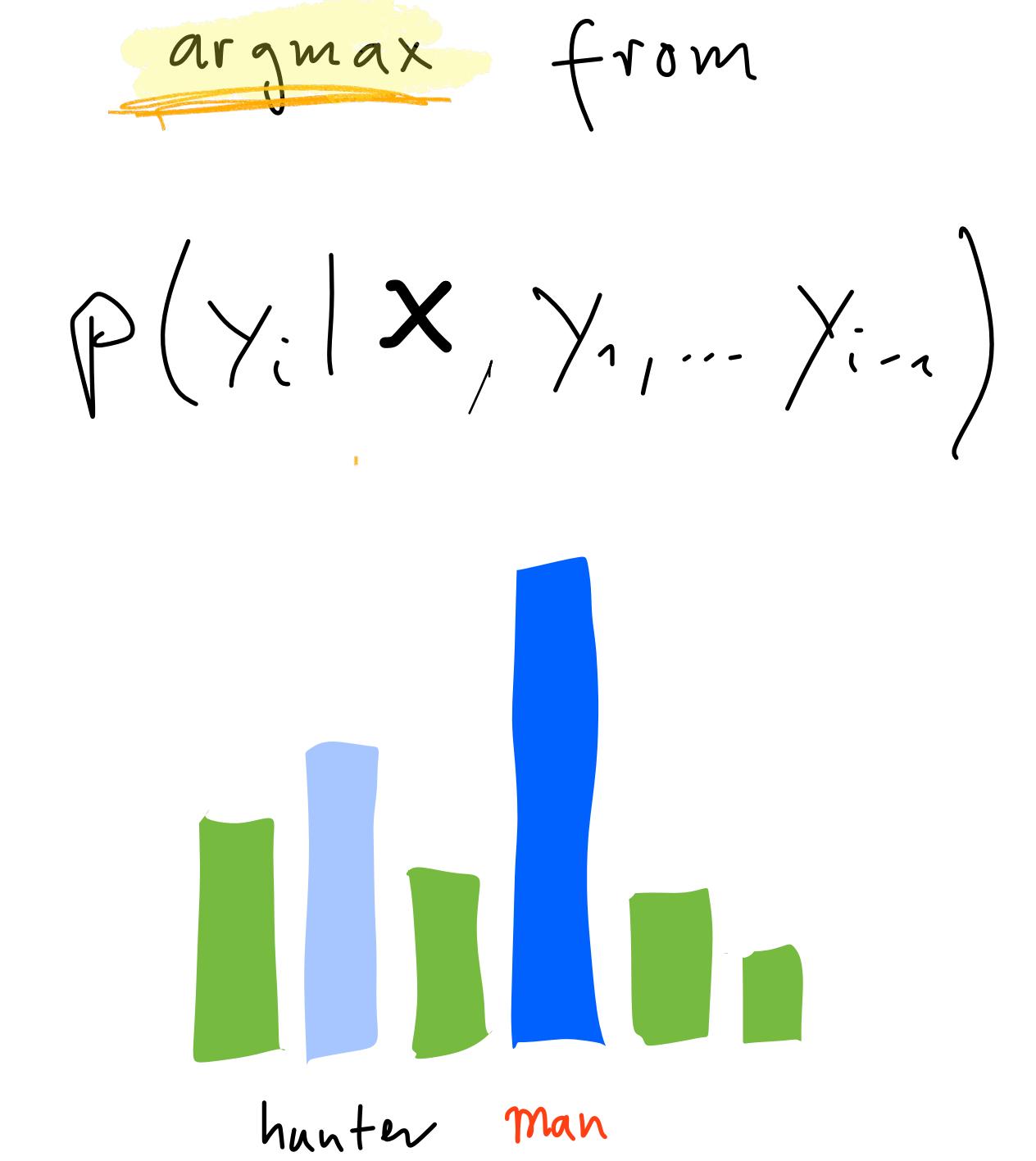
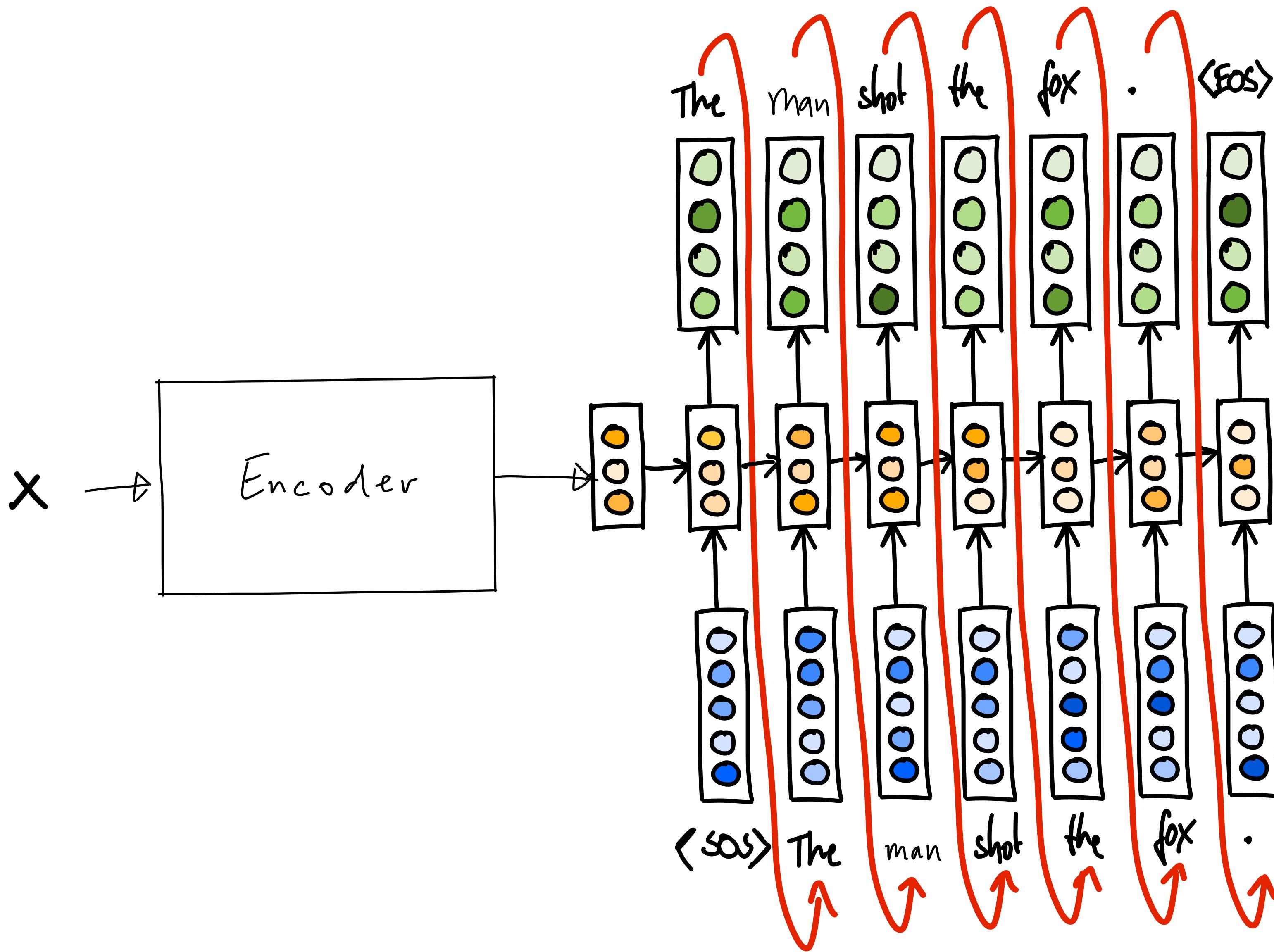
The Generation Problem

- We have a model of $P(Y|X)$, how do we use it to generate a sentence?
- Two methods:
 - **Sampling:** Try to generate a *random* sentence according to the probability distribution.
 - **Argmax:** Try to generate the sentence with the *highest* probability.

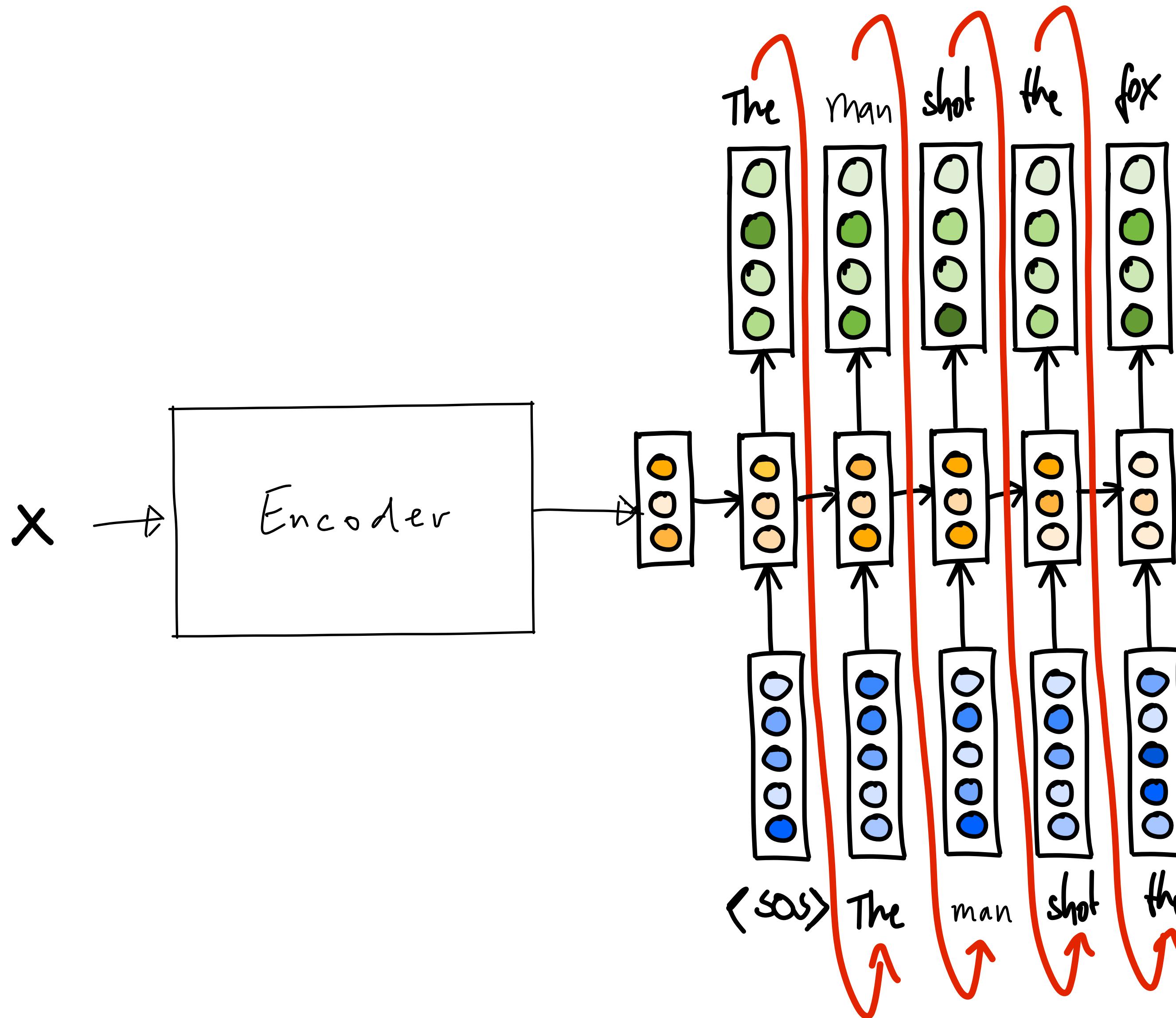
Generating from a Language Model



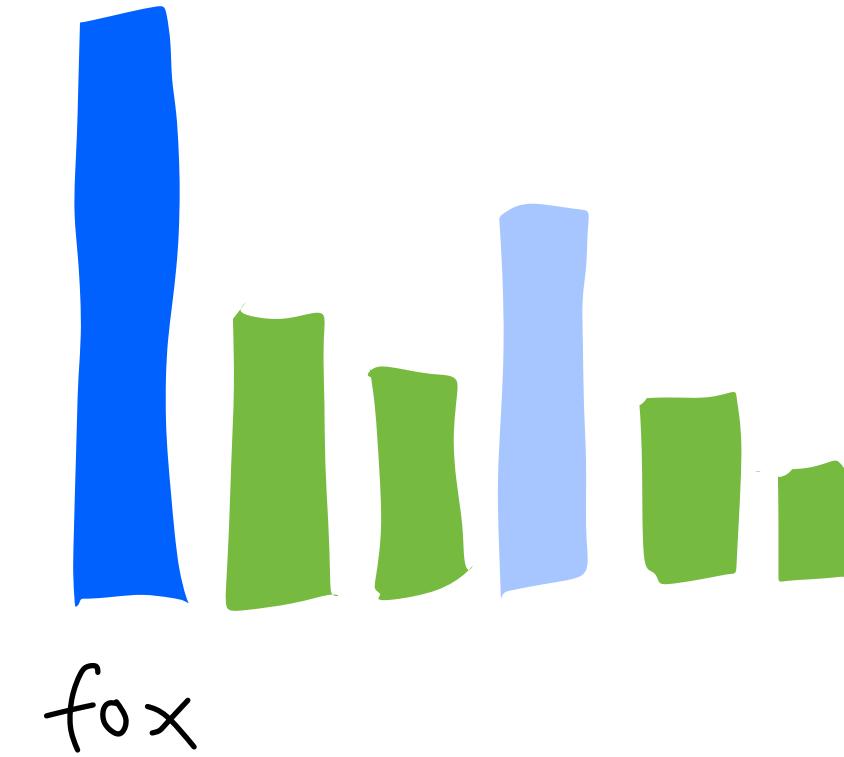
Greedy Search



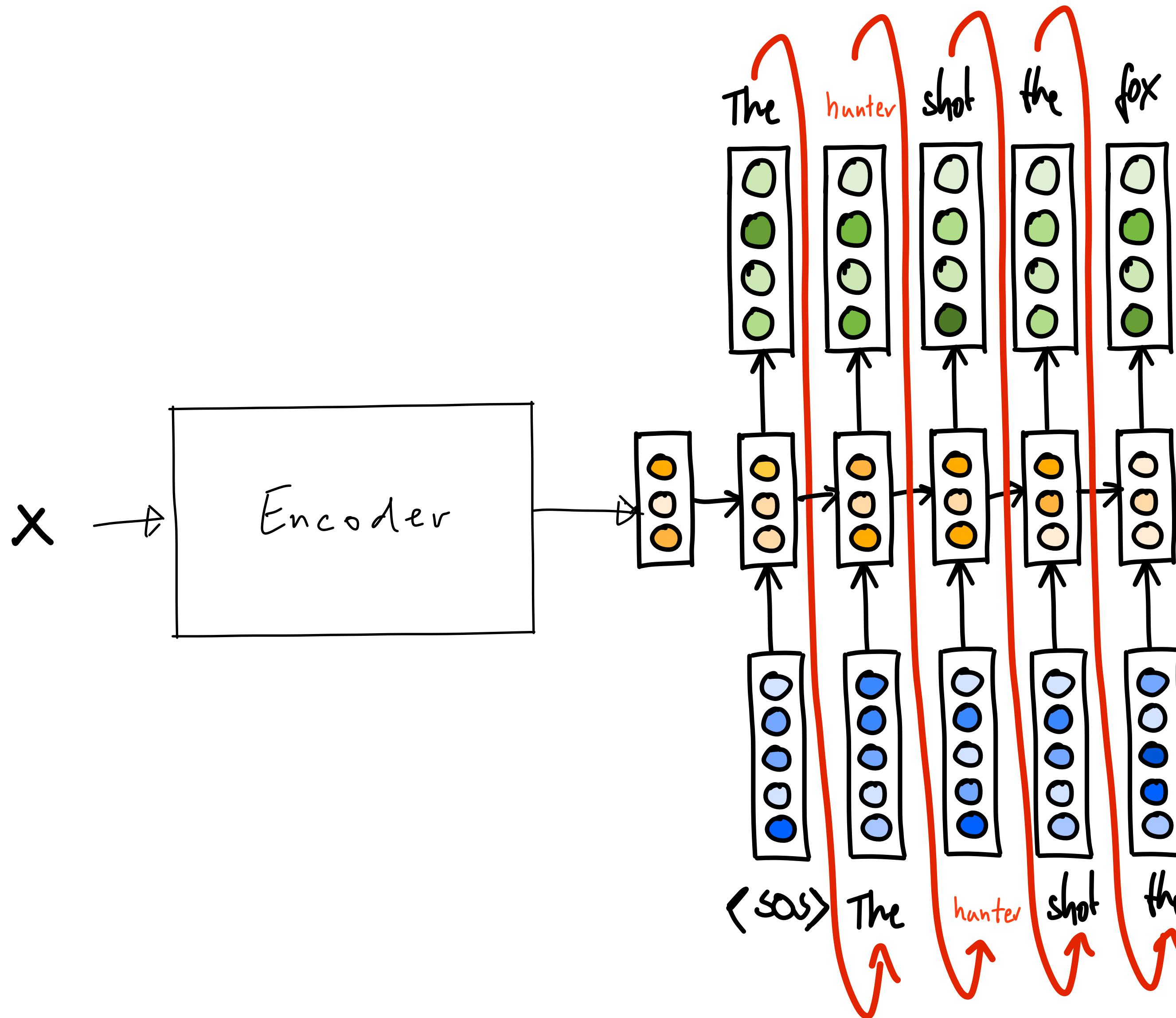
Greedy Search Errors



argmax from
 $p(y_i | x, \langle \text{SOS} \rangle, \text{the}, \text{man}, \text{shot}, \text{the})$



Greedy Search Errors

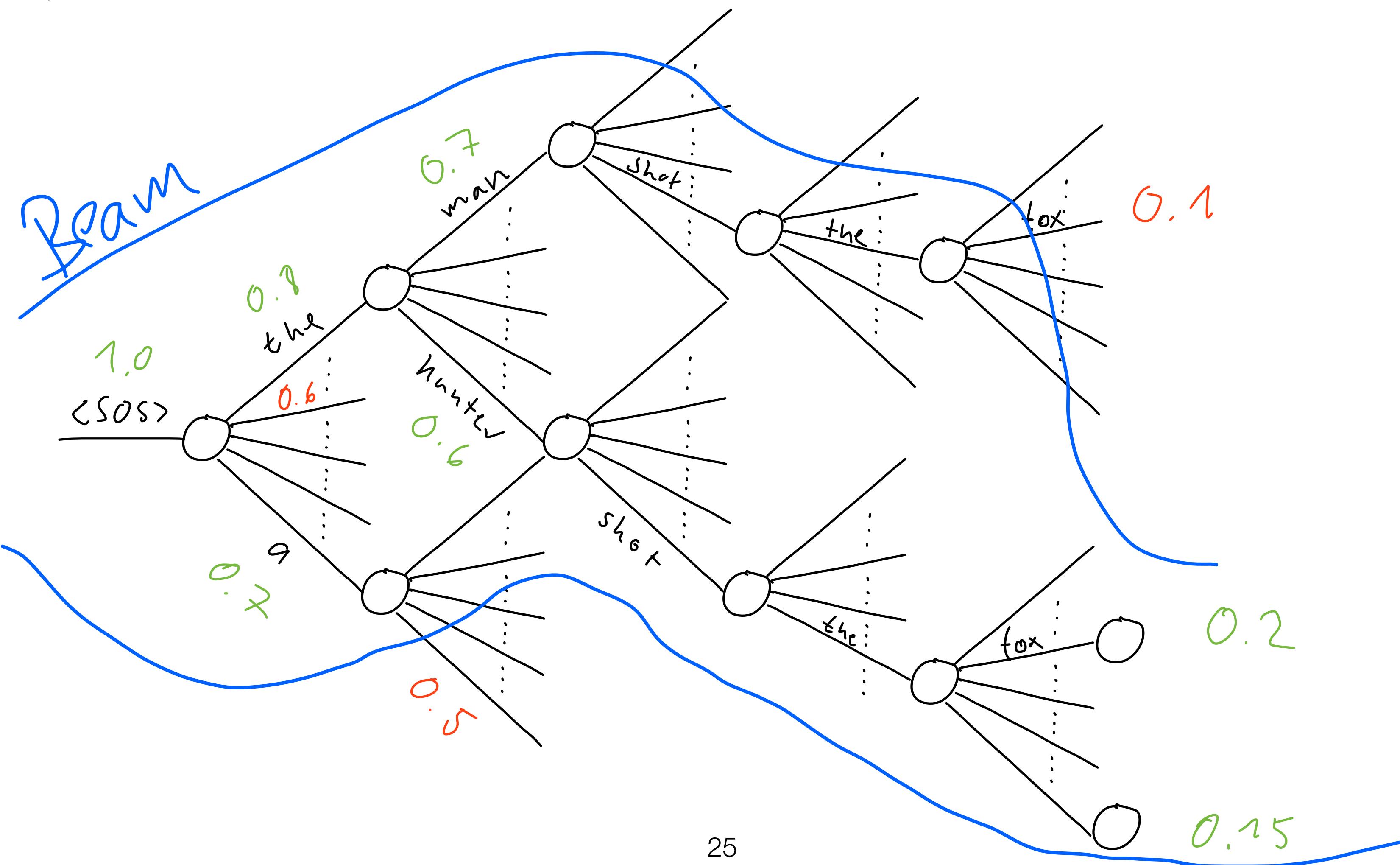


argmax from
 $P(y_i | x, \langle \text{SOS} \rangle, \text{the}, \text{hunter}, \text{shot}, \text{the})$

fox

Beam Search

$$p(y_1) \cdot p(y_2 | \dots) \cdot p(y_3 | \dots) \cdot p(y_4 | \dots) \cdot p(y_5 | \dots) \cdot p(y_6 | \dots)$$



How do we Evaluate?

Basic Evaluation Paradigm

- Use parallel test set
- Use system to generate translations
- Compare target translations w/ reference

Human Evaluation

- Ask a human to do evaluation

| | | | |
|---------------------|-----------------------------|---------------------|-----|
| 太郎が花子を訪れた | | | |
| Taro visited Hanako | the Taro visited the Hanako | Hanako visited Taro | |
| Adequate? | Yes | Yes | No |
| Fluent? | Yes | No | Yes |
| Better? | 1 | 2 | 3 |

- Final goal, but slow, expensive, and sometimes inconsistent

BLEU

- Works by comparing n-gram overlap w/ reference

Reference: Taro visited Hanako

System: the Taro visited the Hanako

1-gram: 3/5

2-gram: 1/4

Brevity: $\min(1, |\text{System}|/|\text{Reference}|) = \min(1, 5/3)$

brevity penalty = 1.0

$$\begin{aligned}\text{BLEU-2} &= (3/5 * 1/4)^{1/2} * 1.0 \\ &= 0.387\end{aligned}$$

- **Pros:** Easy to use, good for measuring system improvement
- **Cons:** Often doesn't match human eval, bad for comparing very different systems

Perplexity

- Calculate the perplexity of the words in the held-out set *without* doing generation
- **Pros:** Naturally solves multiple-reference problem!
- **Cons:** Doesn't consider decoding or actually generating output.
- May be reasonable for problems with lots of ambiguity.

Case Studies in Conditional Language Modeling

From Structured Data

(e.g. Wen et al 2015)

- When you say “Natural Language Generation” to an old-school NLPer, it means this

| | SF Restaurant | SF Hotel |
|----------|---|---|
| act type | inform, inform_only, reject, confirm, select, request, reqmore, goodbye | |
| shared | name, type, *pricerange, price, phone, address, postcode, *area, *near | |
| specific | *food *goodformeal *kids-allowed | *hasinternet *acceptscards *dogs-allowed |

bold=binary slots, *=slots can take “don’t care” value

Still a Difficult Problem!

- e.g. "Challenges in data-to-document generation" (Wiseman et al. 2017)

| TEAM | WIN | LOSS | PTS | FG_PCT | RB | AS ... |
|-----------------|-----|------|-----|--------|-----|----------|
| Heat | 11 | 12 | 103 | 49 | 47 | 27 |
| Hawks | 7 | 15 | 95 | 43 | 33 | 20 |
| PLAYER | AS | RB | PT | FG | FGA | CITY ... |
| Tyler Johnson | 5 | 2 | 27 | 8 | 16 | Miami |
| Dwight Howard | 4 | 17 | 23 | 9 | 11 | Atlanta |
| Paul Millsap | 2 | 9 | 21 | 8 | 12 | Atlanta |
| Goran Dragic | 4 | 2 | 21 | 8 | 17 | Miami |
| Wayne Ellington | 2 | 3 | 19 | 7 | 15 | Miami |
| Dennis Schroder | 7 | 4 | 17 | 8 | 15 | Atlanta |
| Rodney McGruder | 5 | 5 | 11 | 3 | 8 | Miami |
| Thabo Sefolosha | 5 | 5 | 10 | 5 | 11 | Atlanta |
| Kyle Korver | 5 | 3 | 9 | 3 | 9 | Atlanta |
| ... | | | | | | |

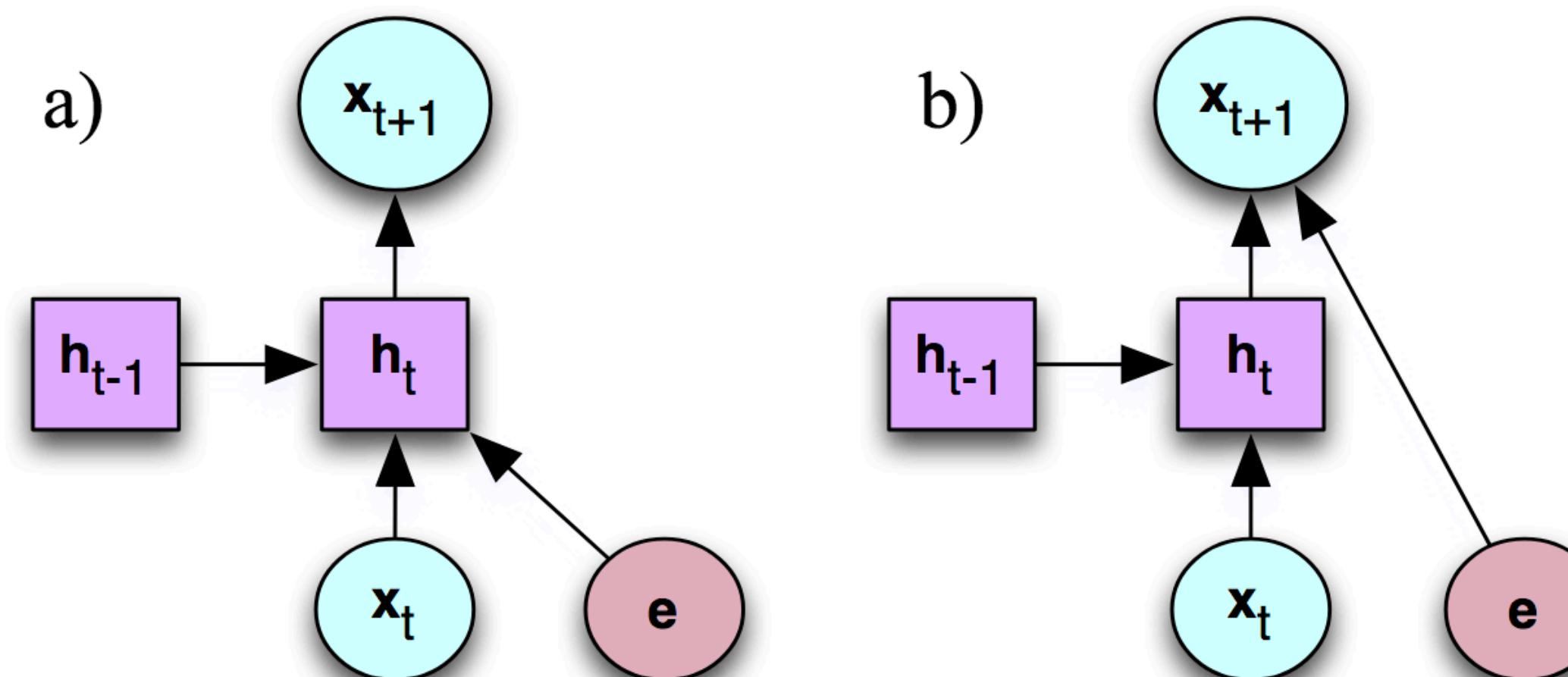
The Utah Jazz (38 - 26) defeated the Houston Rockets (38 - 26) 117 - 91 on Wednesday at Energy Solutions Arena in Salt Lake City . The Jazz got out to a quick start in this one , out - scoring the Rockets 31 - 15 in the first quarter alone . Along with the quick start , the Rockets were the superior shooters in this game , going 54 percent from the field and 43 percent from the three - point line , while the Jazz went 38 percent from the floor and a meager 19 percent from deep . The Rockets were able to out - rebound the Rockets 49 - 49 , giving them just enough of an advantage to secure the victory in front of their home crowd . The Jazz were led by the duo of Derrick Favors and James Harden . Favors went 2 - for - 6 from the field and 0 - for - 1 from the three - point line to score a game - high of 15 points , while also adding four rebounds and four assists

Figure 2: Example document generated by the Conditional Copy system with a beam of size 5. Text that accurately reflects a record in the associated box- or line-score is highlighted in blue, and erroneous text is highlighted in red.

From Speaker/Document Traits

(Hoang et al. 2016)

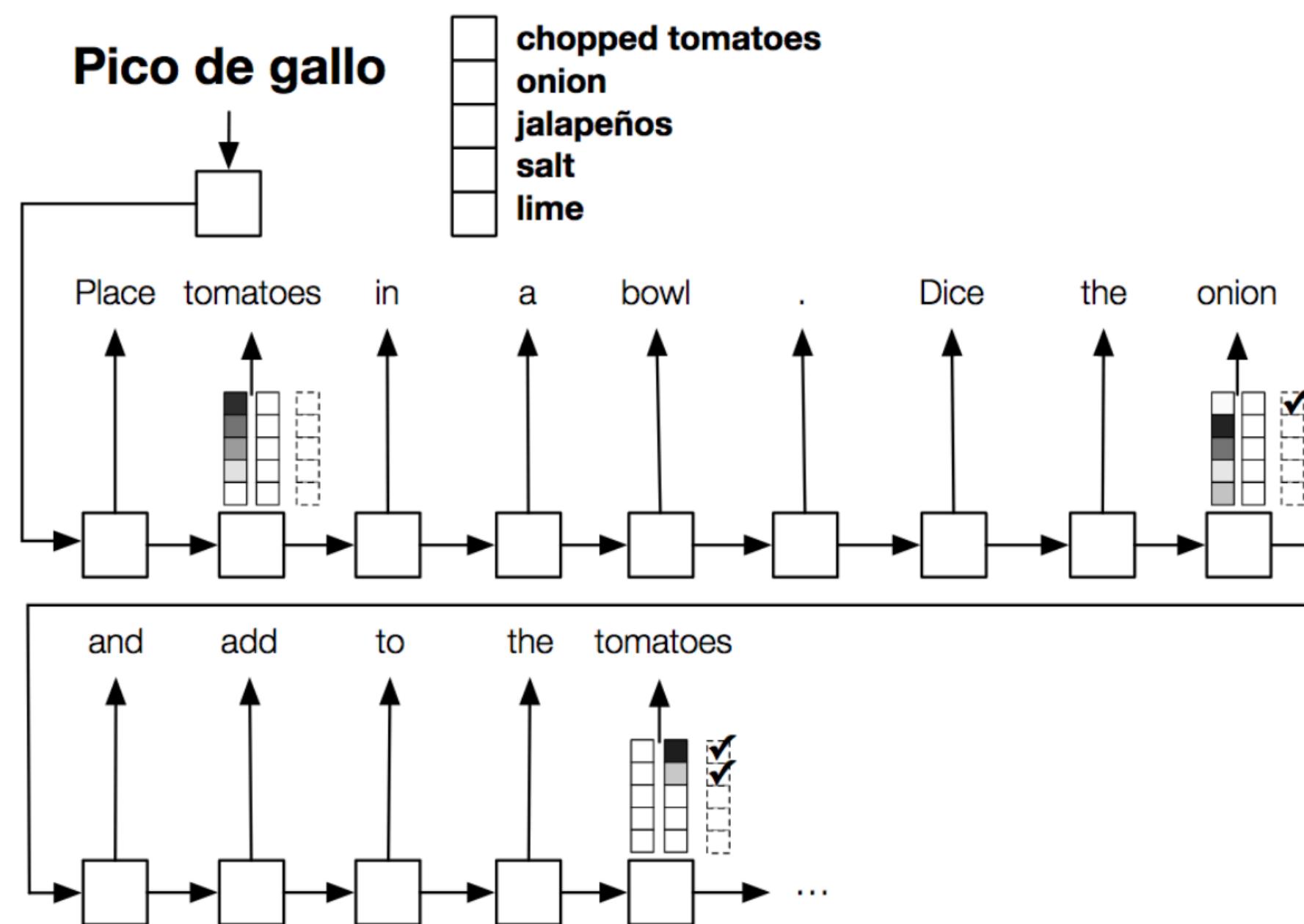
- e.g. TED talk description -> TED talk
- Encode title, description, keywords, author embedding
- Various encoding methods: BOW, CNN, RNN
- Various integration methods: in recurrent layer or softmax layer



From Lists of Traits

(Kiddon et al. 2016)

- Name of a recipe + ingredients -> recipe
- "Neural Checklist Model" that tells when a particular item in the list has been generated



From Images

(e.g. Karpathy et al. 2015)

- Input is image features, output is text



*"A Tabby cat is leaning
on a wooden table, with
one paw on a laser
mouse and the other on
a black laptop"*

- Standard to use CNN-based image encoders
- Often pre-trained on large databases such as ImageNet

Image Captioning a Year Ago



AI: A plate with a sandwich and salad on it.



Image Captioning with a Personality

Sweet AI: This is a lovely sandwich.

Dramatic AI: This sandwich looks so delicious! My goodness!

Anxious AI: I'm afraid this might make me sick if I eat it.

Sympathetic AI: I feel so bad for that carrot!

Arrogant AI: I make better sandwiches than this.

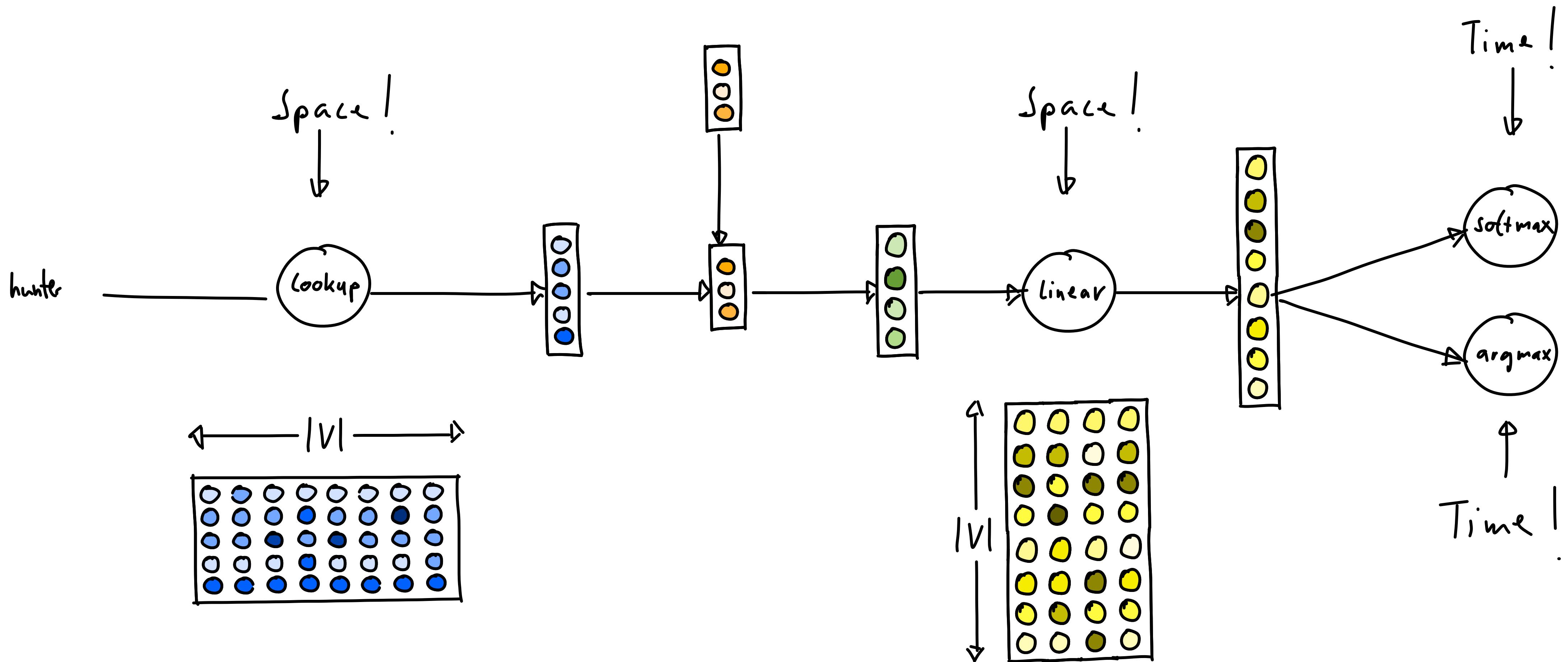
From Word Embeddings (Noraset et al. 2017)

- Baseline: standard sequence-to-sequence model
- Additional information about the affixes and hypernyms

| Word | Generated definition |
|--------------|--|
| brawler | a person who fights |
| butterfish | a marine fish of the atlantic coast |
| continually | in a constant manner |
| creek | a narrow stream of water |
| feminine | having the character of a woman |
| juvenility | the quality of being childish |
| mathematical | of or pertaining to the science of mathematics |
| negotiate | to make a contract or agreement |
| prance | to walk in a lofty manner |
| resent | to have a feeling of anger or dislike |
| similar | having the same qualities |
| valueless | not useful |

Open Vocabulary Generation

Vocabulary Size



Out of Vocabulary Words

V = ?

compounds

solar system → Sonnensystem

named entities

Obama

→ オバマ

loanwords

claustraphobia → Klaustraphobia

Byte Pair Encoding

- Iteratively replace most frequent pair with new word
- apply on dictionary, not full text
- Vocabulary = characters + new words

| word | freq |
|-------------|------|
| low </w> | 5 |
| lower </w> | 2 |
| newest </w> | 6 |
| widest </w> | 3 |

| freq | symbol pair | new symbol |
|------|-------------|------------|
| | | |