

COMP 4446 / 5046

Lecture 13: Review

Jonathan K. Kummerfeld

Semester 1, 2025

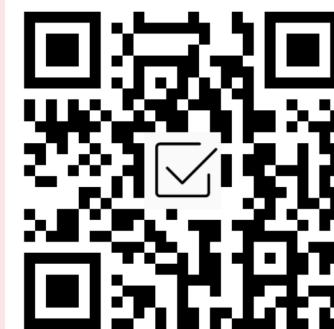
Do the Unit of Study Survey!

[https://student-
surveys.sydney.edu.au/
students/](https://student-surveys.sydney.edu.au/students/)



THE UNIVERSITY OF
SYDNEY

menti.com 6574 1888



Turing Test

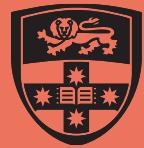
TURING TEST EXTRA CREDIT:
CONVINCE THE EXAMINER
THAT HE'S A COMPUTER.

YOU KNOW, YOU MAKE
SOME REALLY GOOD POINTS.
I
I'M ... NOT EVEN SURE
WHO I AM ANYMORE.



[Hit Turing right in
the test-ees.]

Source: <https://xkcd.com/329/>



Can you give us an overall visualisation (diagram) of how each lecture components link to each other?

Inference

Exhaustive
Greedy and Search
Dynamic
programming

Models

Linear
Non-linear
Encoder-decoder
Transformer
Large Language Models
Agents

Training

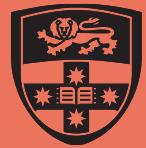
Supervised
Unsupervised
Reinforcement Learning

Data

Annotation and crowdsourcing



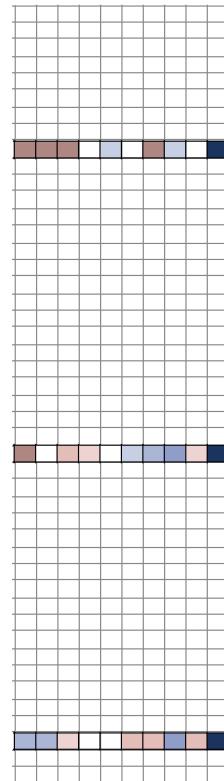
[menti.com 6574 1888](https://menti.com/65741888)



Differences between static and contextual embeddings

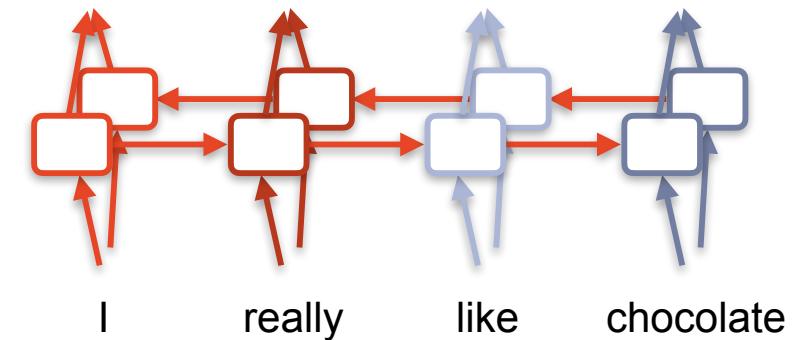
Static

chocolate



Contextual

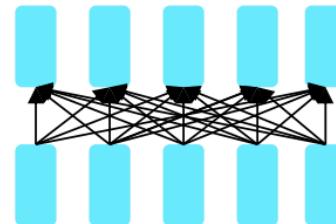
Contextual vectors for each word



Key difference: Static vectors are the same everywhere, contextual vectors vary across sentences



Why it is hard for encoder-only models to do generation tasks



Encoders

$$P(x_i | x_{1\dots i-1}, x_{i+1\dots N})$$

Recipe for hot chocolate:



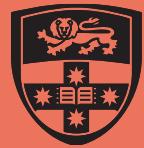
Recipe for hot chocolate:
40g



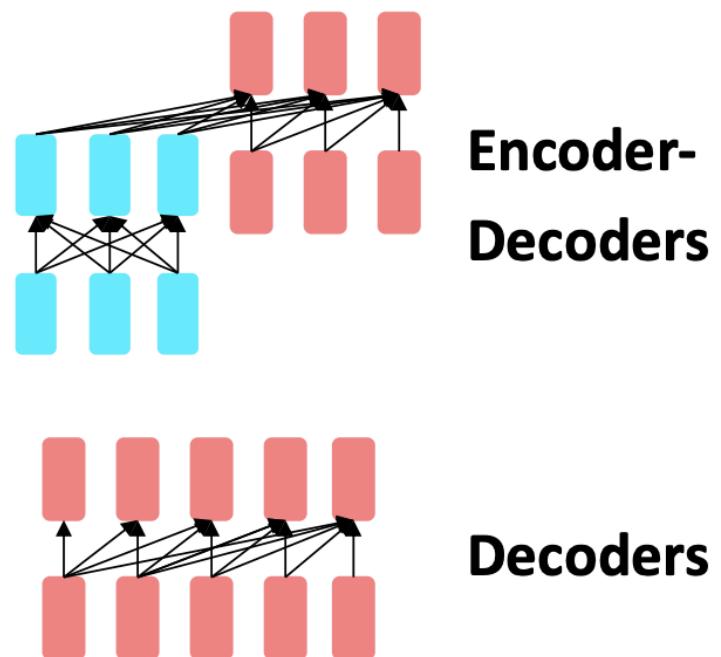
Recipe for hot chocolate:
40g chocolate



Key issue: The encoder generates a fixed number of outputs that matches its number of inputs. Self-attention looks everywhere, so you can't add new inputs in.



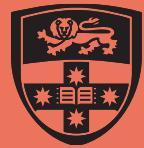
Why it is hard for encoder-only models to do generation tasks



With a decoder, we process left-to-right, so we can easily do another step of output



[menti.com 6574 1888](https://menti.com/65741888)



LLM Prompting and different ways to train LLM models

Review: The cinematography was stellar; great movie!
Sentiment (positive or negative): positive

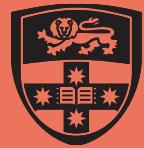
Review: The plot was boring and the visuals were subpar.
Sentiment (positive or negative): negative

Review: The movie's acting could've been better, but the
visuals and directing were top-notch.
Sentiment (positive or negative):

Context: Christopher agrees with Kevin. [...] Q: Who hangs out with a student?
Because Mary hangs out with Danielle and Danielle is a student, the answer is Mary.

Explain-predict: answer is conditioned on output explanation (Chain of Thought)





LLM Prompting and different ways to train LLM models

<s> I like Sydney University's NLP course .

<s> I like Stanford University's AI class .



Model

<s> I like Sydney University's [MASK] course .

Sydney
talk



[menti.com 6574 1888](https://menti.com/65741888)



General principles of training a large language model for tasks

Options for use:

- (1) Put a task-specific model on top
- (2) Use prompting
- (3) A combination of the two!

Options for training:

- (1) Fine-tuning parameters
- (2) Prompt tuning
- (3) A combination of the two!

General suggestions:

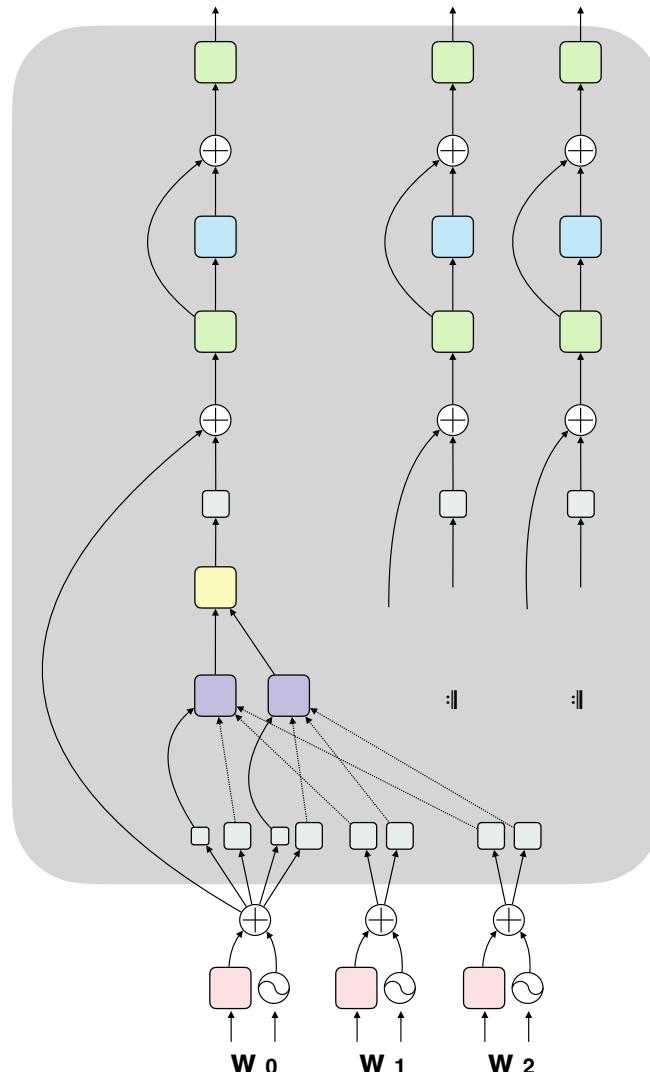
- Have a carefully designed held-out evaluation dataset
- Apply all the methods from ML classes
- Vary your data and model



[menti.com 6574 1888](https://menti.com/65741888)



Review of the transformer and the importance of each component for its performance /The inner workings on Transformer



Feed Forward

$$\max(\mathbf{0}, \mathbf{xW}_1 + \mathbf{b}_1)\mathbf{W}_2 + \mathbf{b}_2$$

Layer Norm

⊕ Element-wise sum

Concatenate

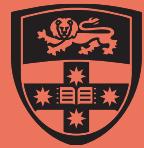
Attention:

$$\text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V$$

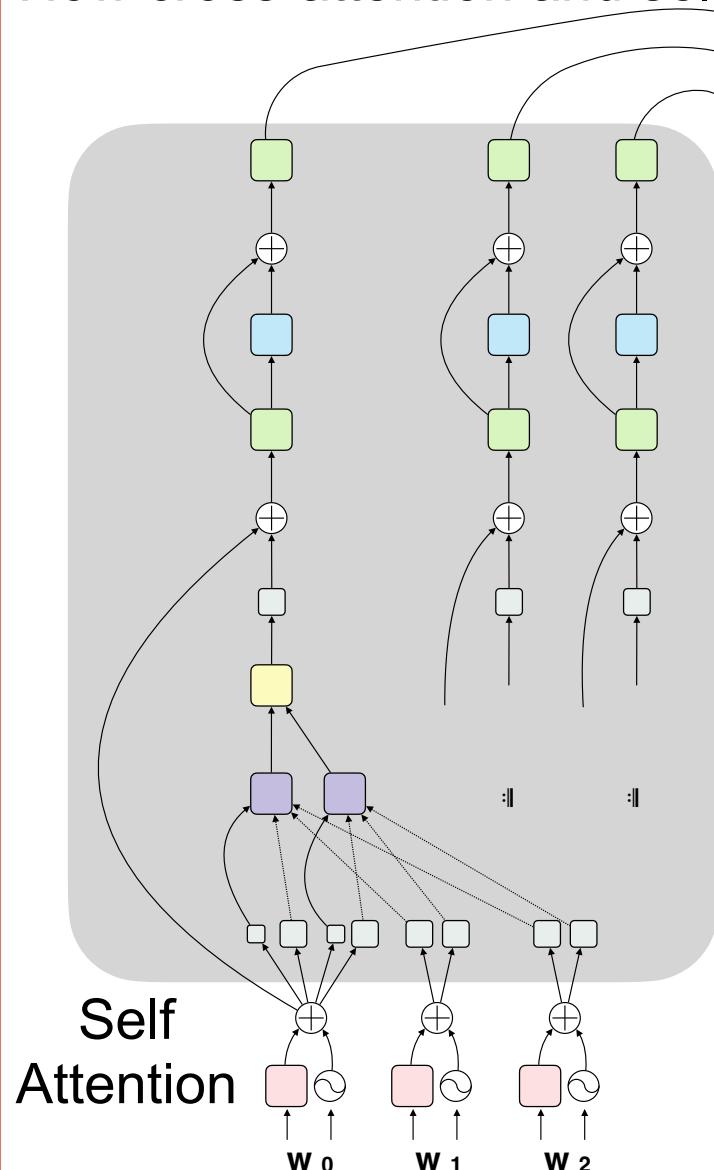
Linear transform

Look up vector

⊕ Position as vector



How cross-attention and self-attention differ

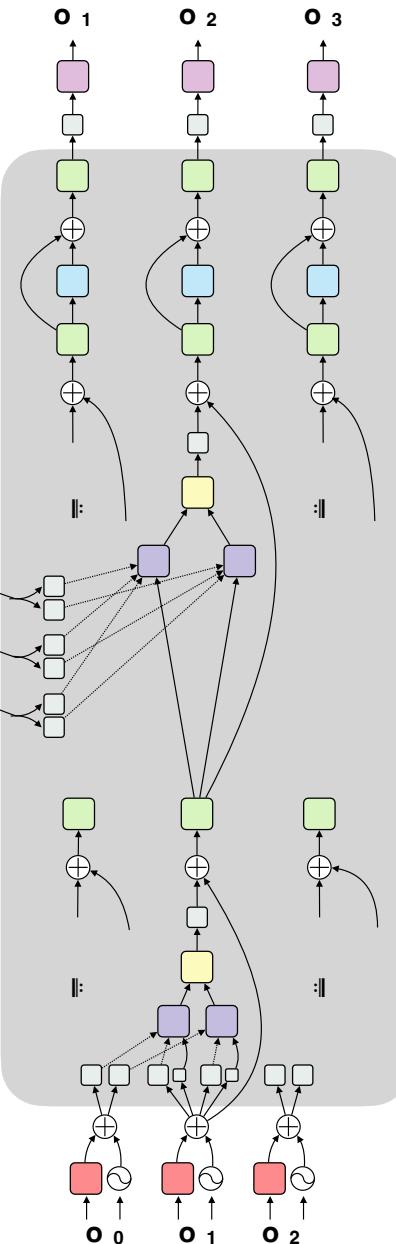


Cross
Attention.

*input
from
encoder*

Self
Attention

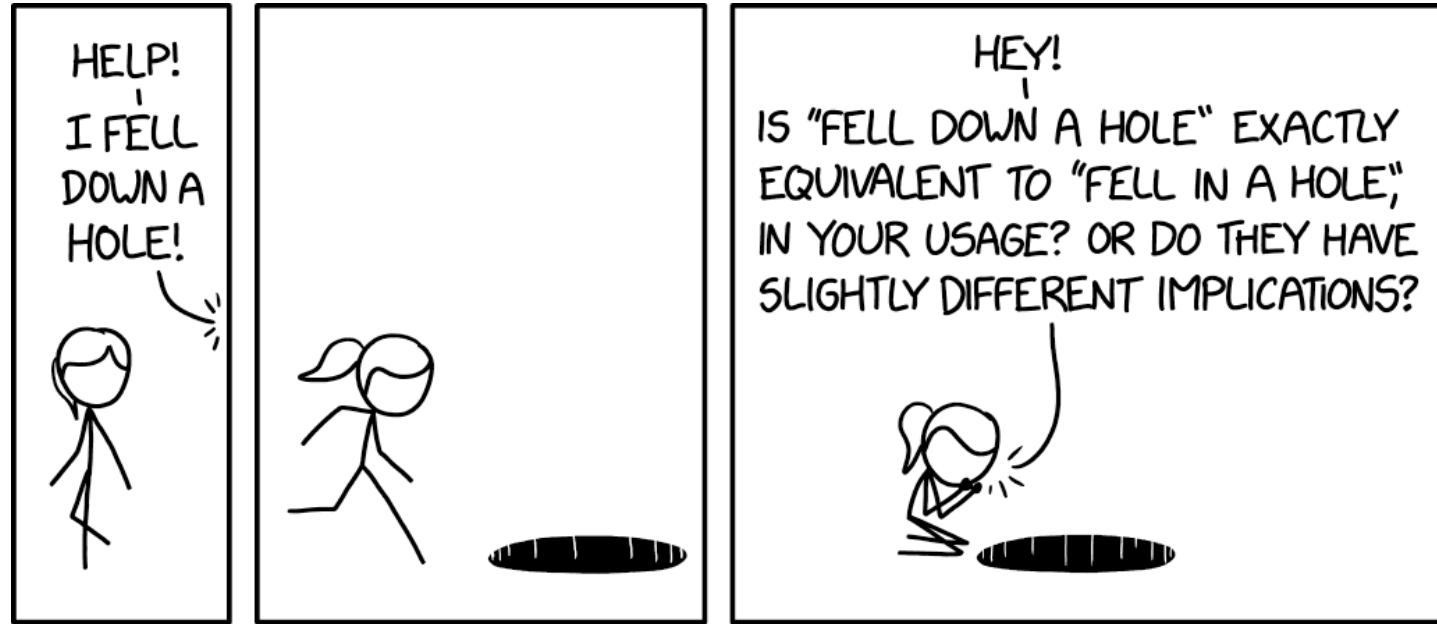
Query from decoder





My
Survey

University
Survey



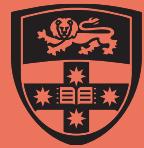
THERE'S A MYTH THAT LINGUISTS ARE PEDANTS WHO LOVE CORRECTING PEOPLE, BUT THEY'RE ACTUALLY JUST ENTHUSIASTIC ABOUT UNDERSTANDING LANGUAGE IN ALL ITS INFINITE VARIETIES, WHICH IS MUCH WORSE.

[“Do you feel like the answer depends on whether you’re currently in the hole, versus when you refer to the events later after you get out? Assuming you get out.”]

Source: <https://xkcd.com/2390/>

menti.com
6574 1888





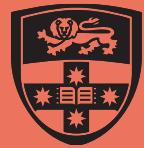
Can you briefly explain the marking rubric ie if its a 2 marker should we aim to give 2 points, if its a 5 marker how descriptive are we expected to be(do we need to give 5 points). For coding also...

Factors going into value:

- How many aspects in answer
- How much time I expect it to take
- How much weight I want this concept to have
- How difficult the question is



[menti.com 6574 1888](https://menti.com/65741888)



Exam notes

Content

Lectures

Labs

Assignments

1-4 marks
each

Other notes

Equation sheet provided

Spare pages for scratch paper

No calculators or other electronic devices permitted

1 A4 page is permitted (double sided, printed or written)

Translating dictionary permitted if it has been checked

Structure

Multiple Choice

- 10 questions, 10 marks

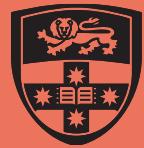
Short Answer

- 20 questions, 40 marks

Code (like quizzes)

- 2 questions, 10 marks





Greedy vs. exhaustive inference methods

Exhaustive

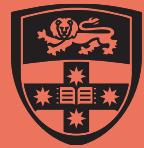
- Always finds the highest score option
- Can consider the entire output when scoring

Greedy

- Faster
- When using top-K / top-P / etc, it can be variable

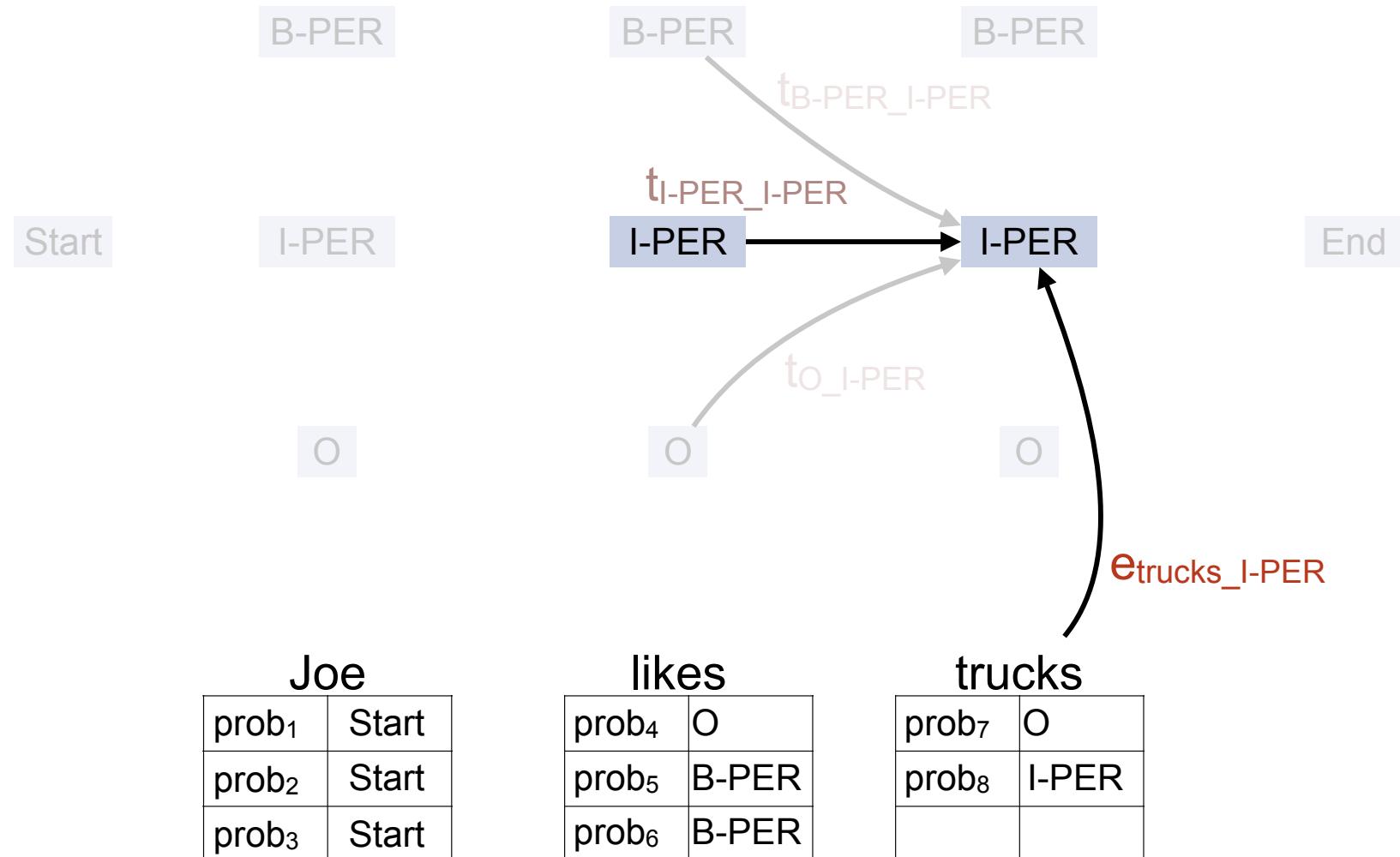


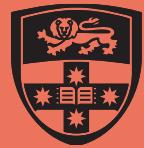
[menti.com 6574 1888](https://menti.com/65741888)



How the Viterbi algorithm works

$$\text{prob}_8 = e_{\text{trucks_I-PER}} * \max(t_{\text{prev_I-PER}} * \text{prob}_{\text{prev}})$$

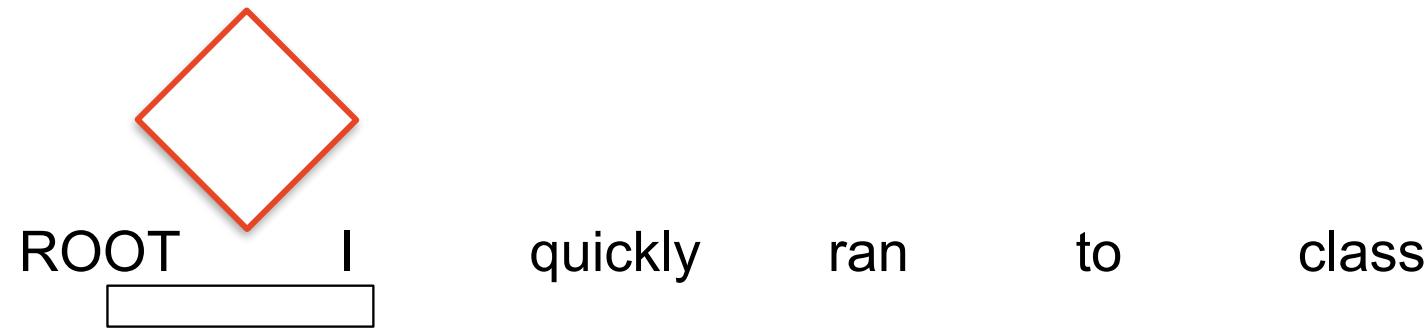


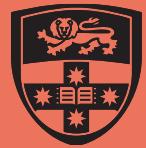


Can we go through the CYK algorithm for dependency parsing? /
Can you “Explain Like I’m 5” the CYK/CKY algorithm in dynamic
programming parsing in another way?



[menti.com 6574 1888](https://menti.com/65741888)

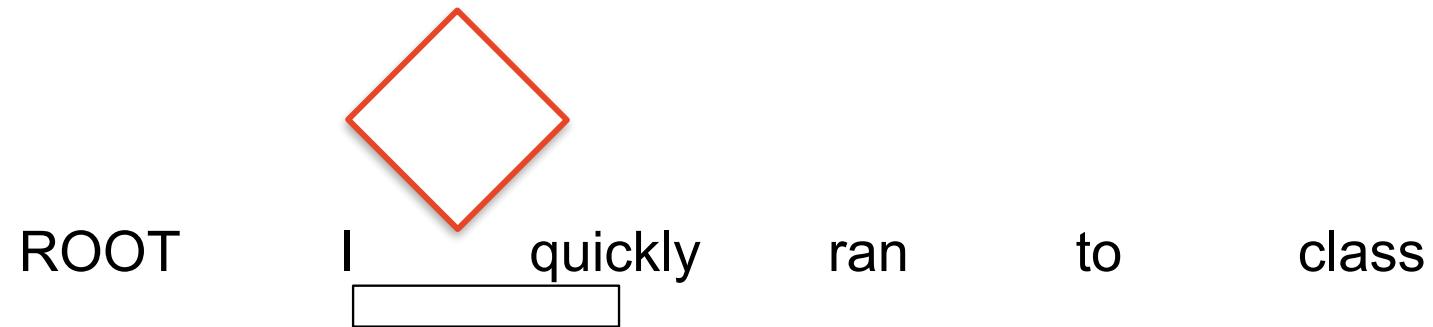


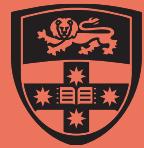


Can we go through the CYK algorithm for dependency parsing? /
Can you “Explain Like I’m 5” the CYK/CKY algorithm in dynamic
programming parsing in another way?



[menti.com 6574 1888](https://menti.com/65741888)

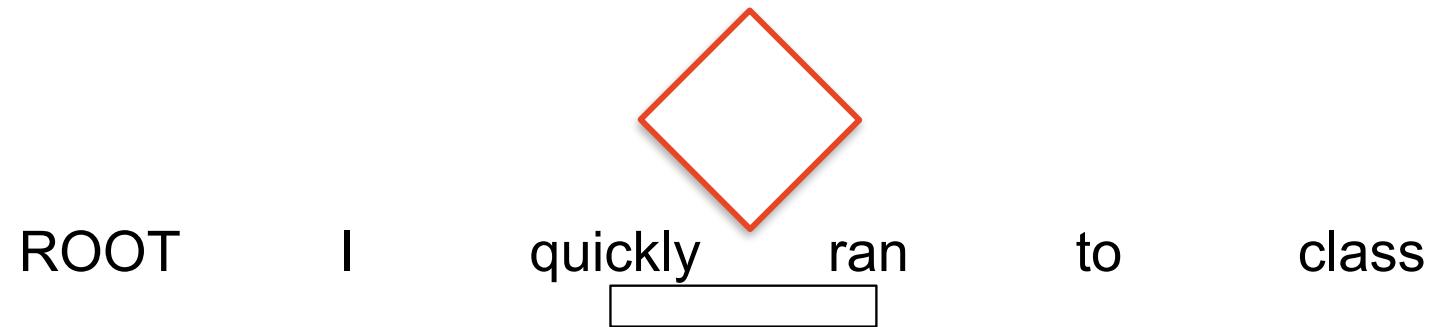


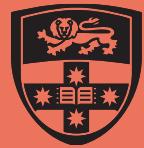


Can we go through the CYK algorithm for dependency parsing? /
Can you “Explain Like I’m 5” the CYK/CKY algorithm in dynamic
programming parsing in another way?



[menti.com 6574 1888](https://menti.com/65741888)

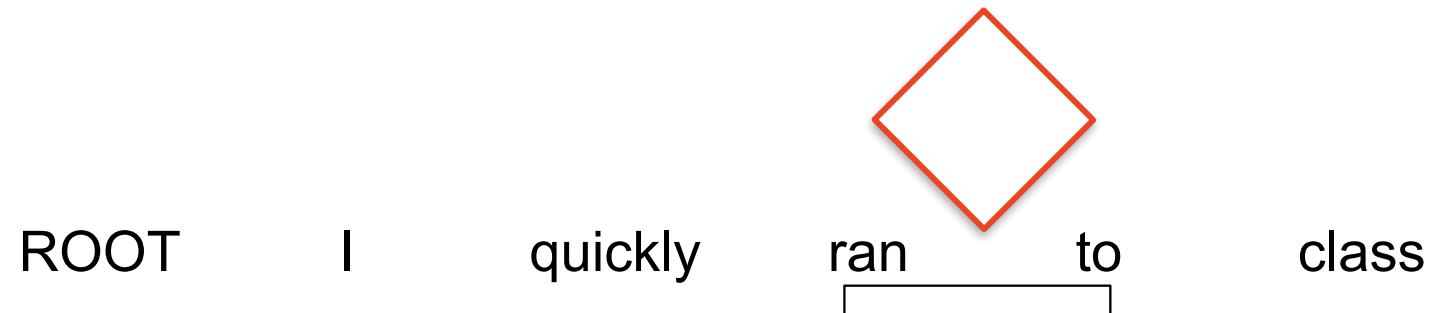




Can we go through the CYK algorithm for dependency parsing? /
Can you “Explain Like I’m 5” the CYK/CKY algorithm in dynamic
programming parsing in another way?



[menti.com 6574 1888](https://menti.com/65741888)

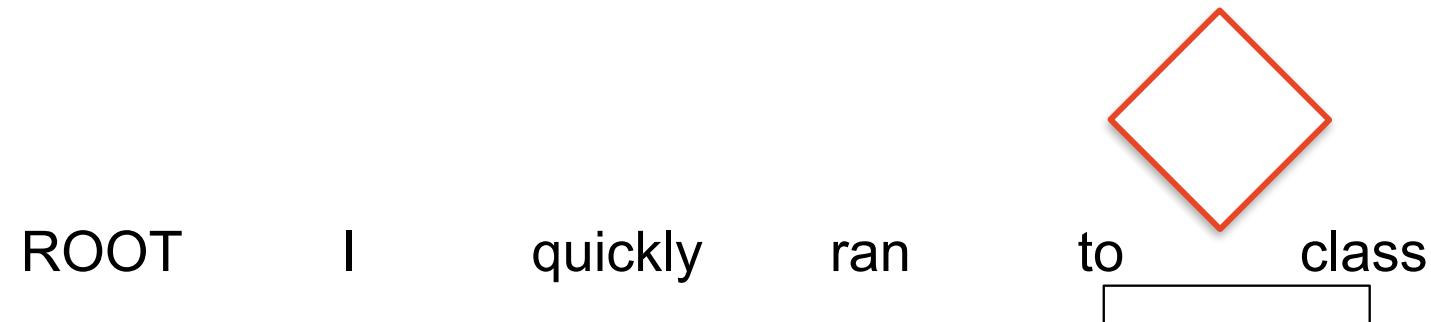


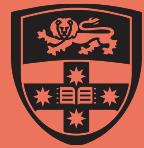


Can we go through the CYK algorithm for dependency parsing? /
Can you “Explain Like I’m 5” the CYK/CKY algorithm in dynamic
programming parsing in another way?



[menti.com 6574 1888](https://menti.com/65741888)

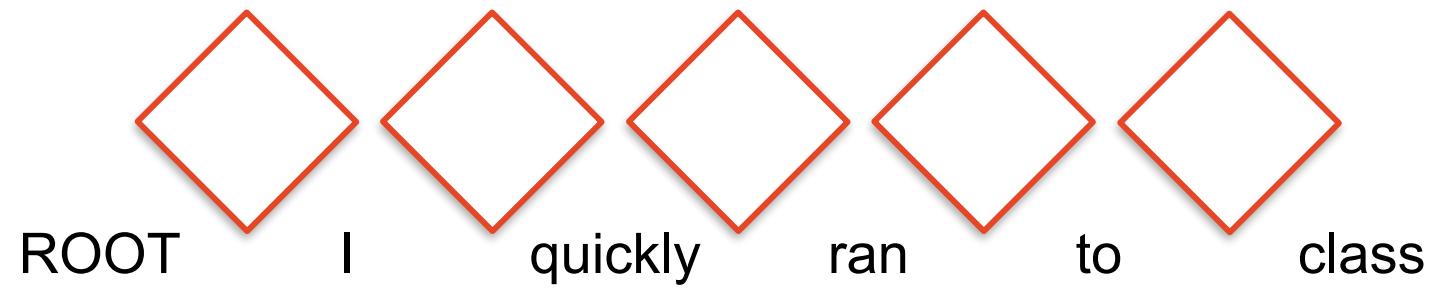


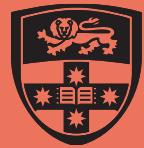


Can we go through the CYK algorithm for dependency parsing? /
Can you “Explain Like I’m 5” the CYK/CKY algorithm in dynamic
programming parsing in another way?



[menti.com 6574 1888](https://menti.com/65741888)

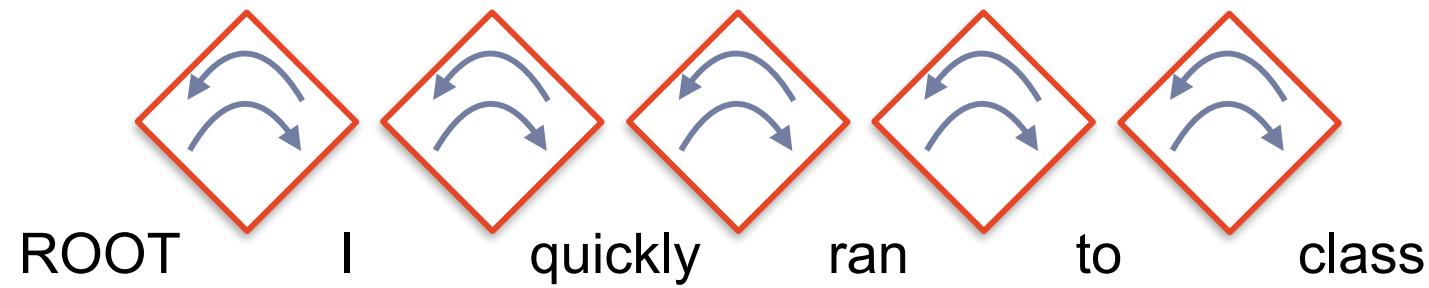


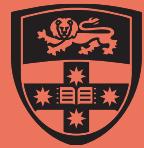


Can we go through the CYK algorithm for dependency parsing? /
Can you “Explain Like I’m 5” the CYK/CKY algorithm in dynamic
programming parsing in another way?

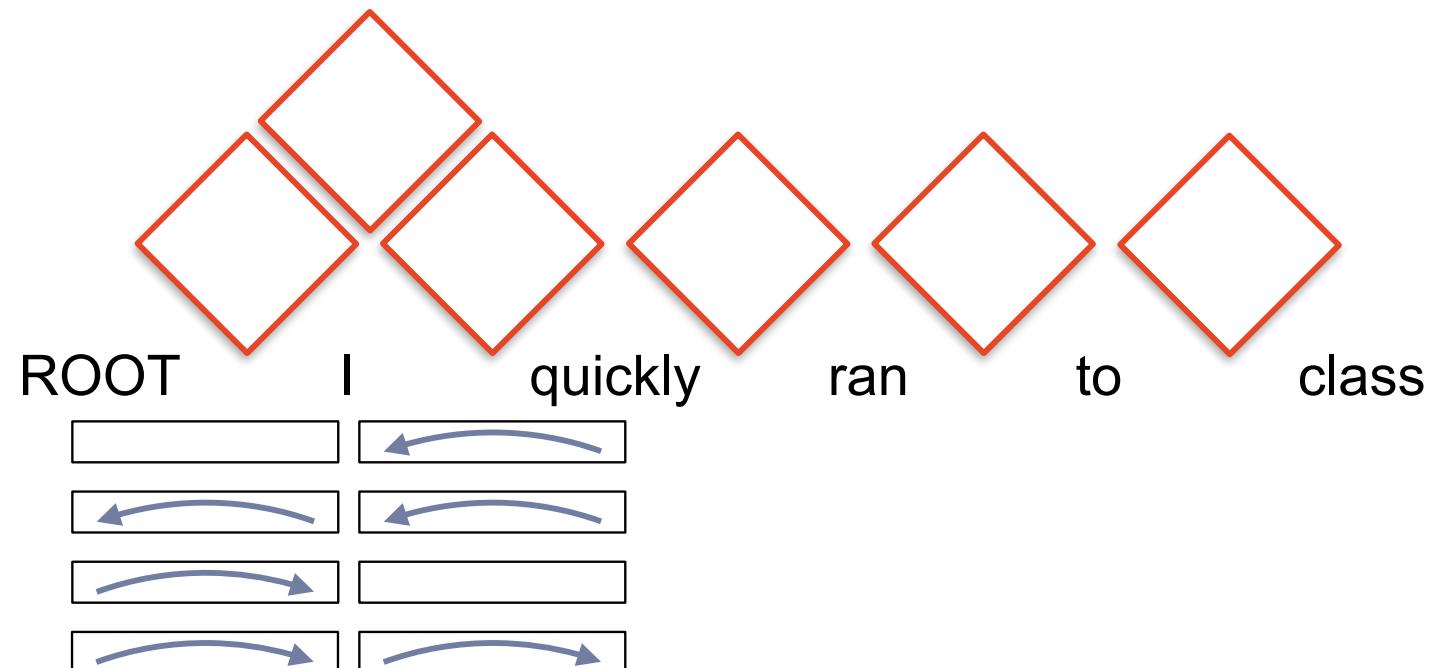


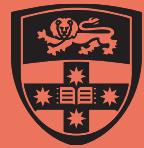
[menti.com 6574 1888](https://menti.com/65741888)





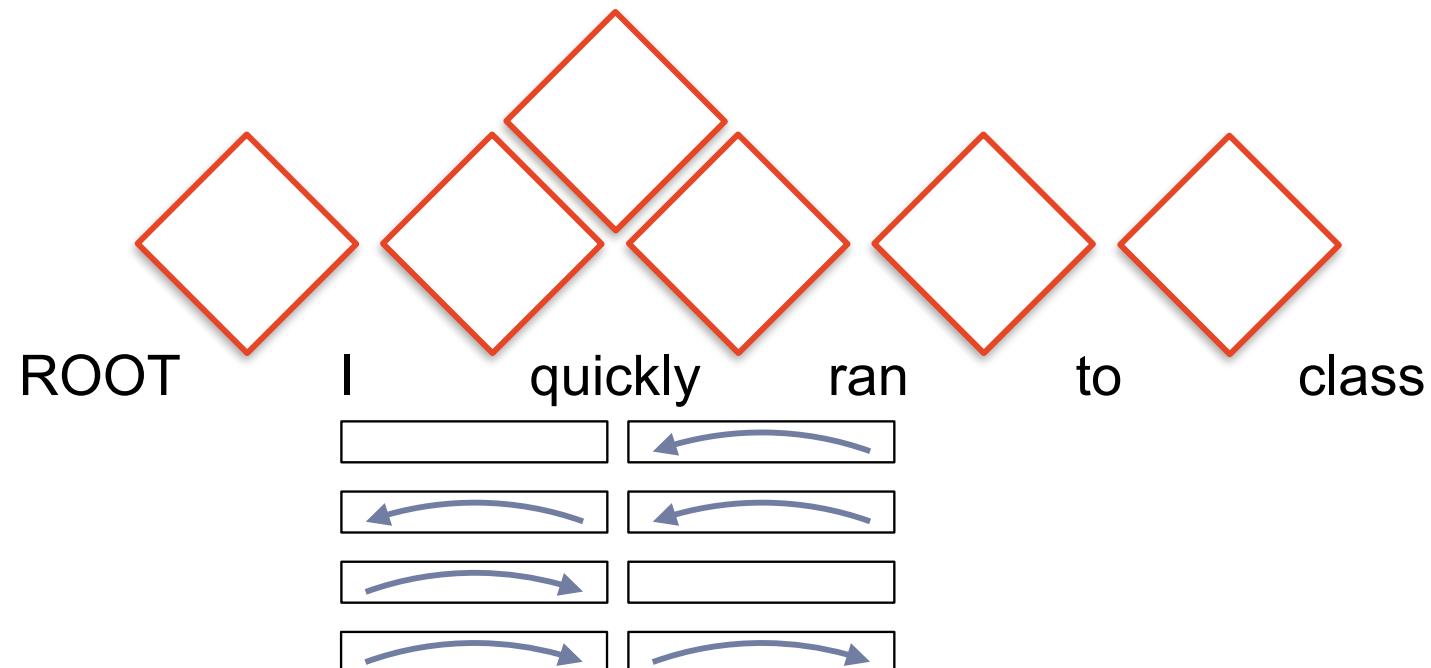
Can we go through the CYK algorithm for dependency parsing? /
Can you “Explain Like I’m 5” the CYK/CKY algorithm in dynamic
programming parsing in another way?

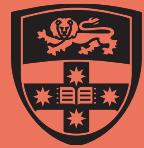




[menti.com 6574 1888](https://menti.com/65741888)

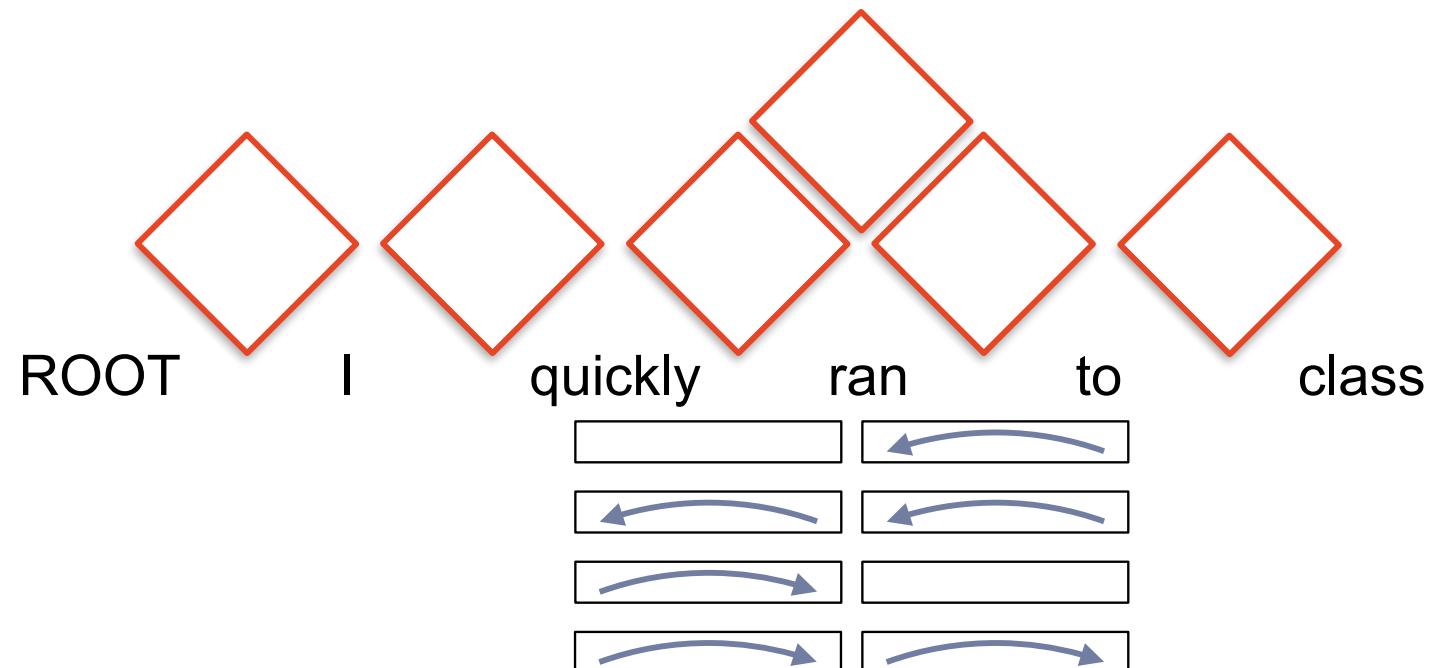
Can we go through the CYK algorithm for dependency parsing? /
Can you “Explain Like I’m 5” the CYK/CKY algorithm in dynamic
programming parsing in another way?

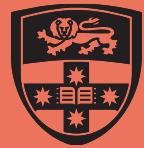




[menti.com 6574 1888](https://menti.com/65741888)

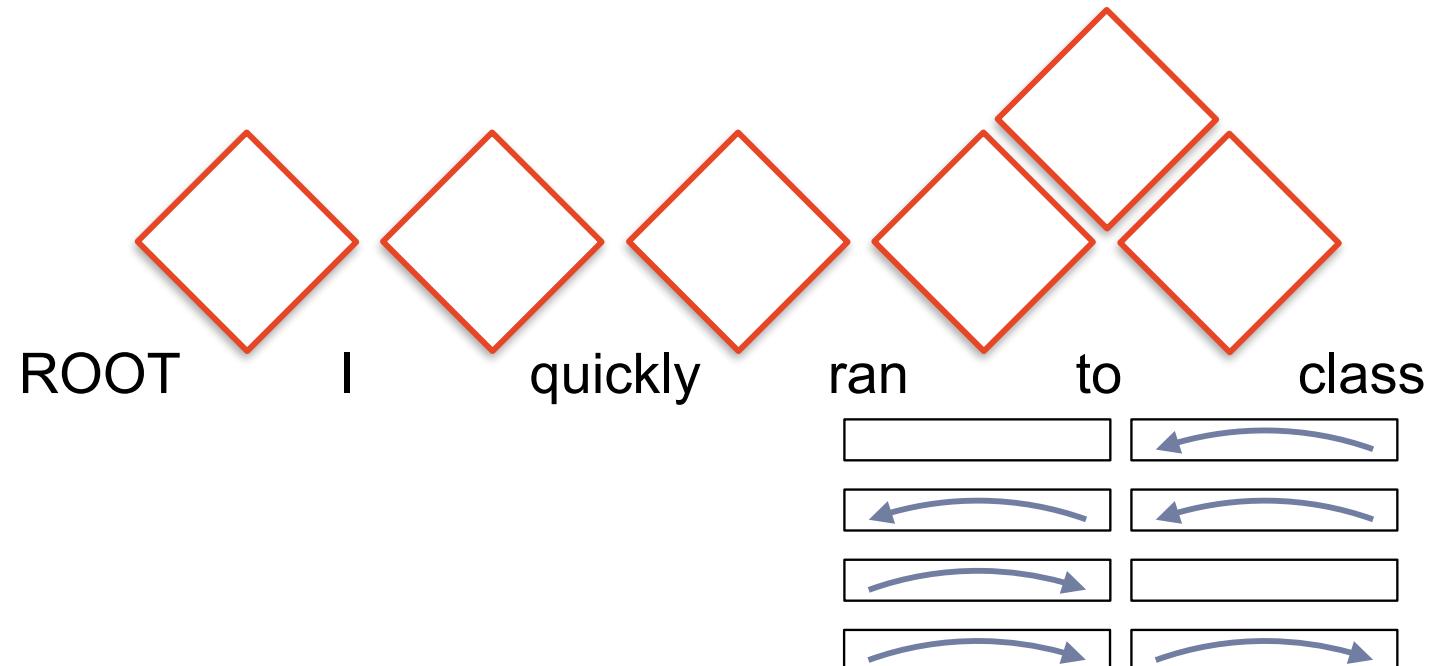
Can we go through the CYK algorithm for dependency parsing? /
Can you “Explain Like I’m 5” the CYK/CKY algorithm in dynamic
programming parsing in another way?

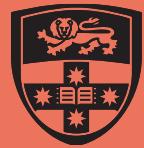




[menti.com 6574 1888](https://menti.com/65741888)

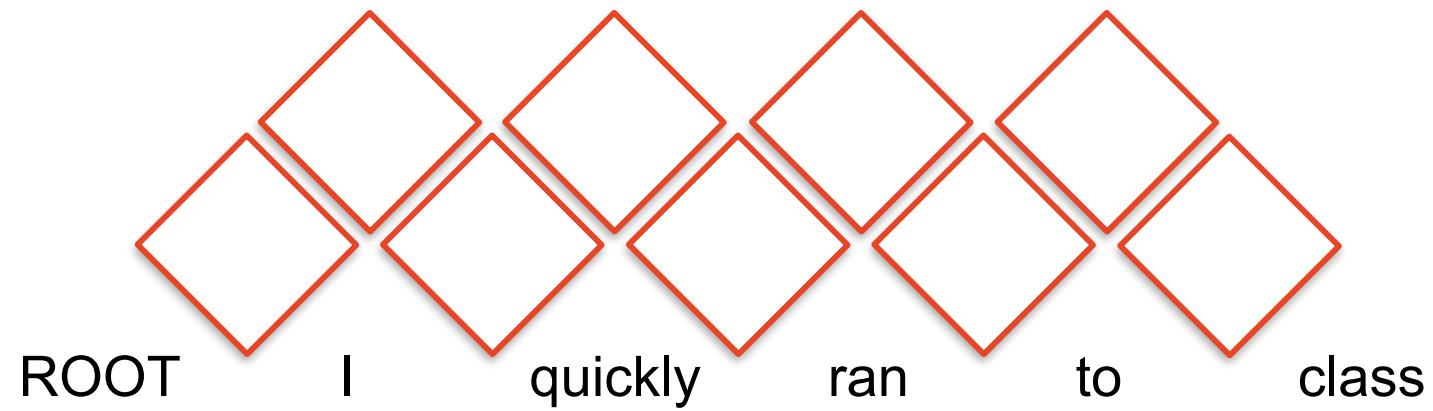
Can we go through the CYK algorithm for dependency parsing? /
Can you “Explain Like I’m 5” the CYK/CKY algorithm in dynamic
programming parsing in another way?

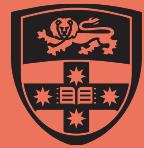




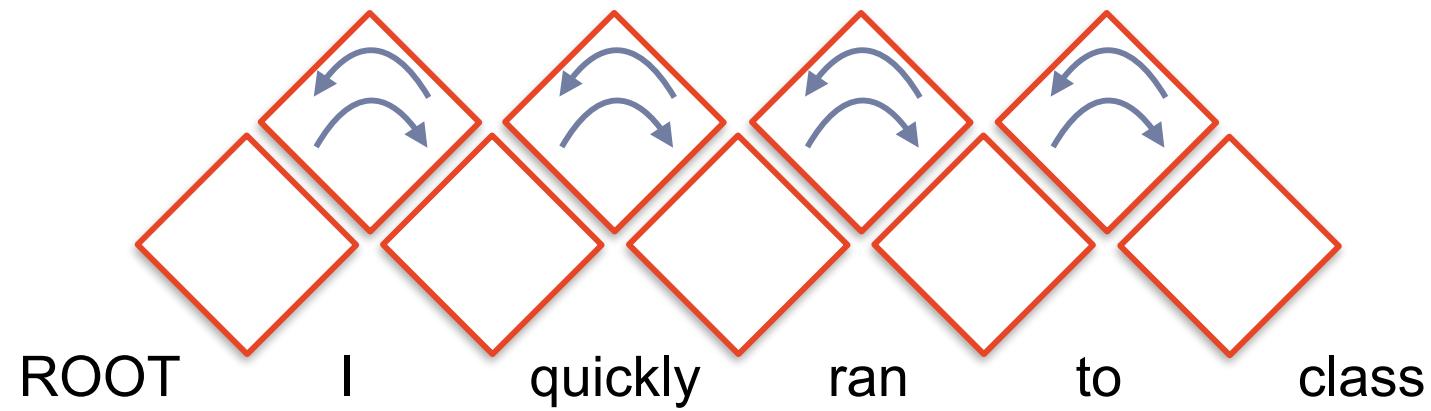
[menti.com 6574 1888](https://menti.com/65741888)

Can we go through the CYK algorithm for dependency parsing? /
Can you “Explain Like I’m 5” the CYK/CKY algorithm in dynamic
programming parsing in another way?





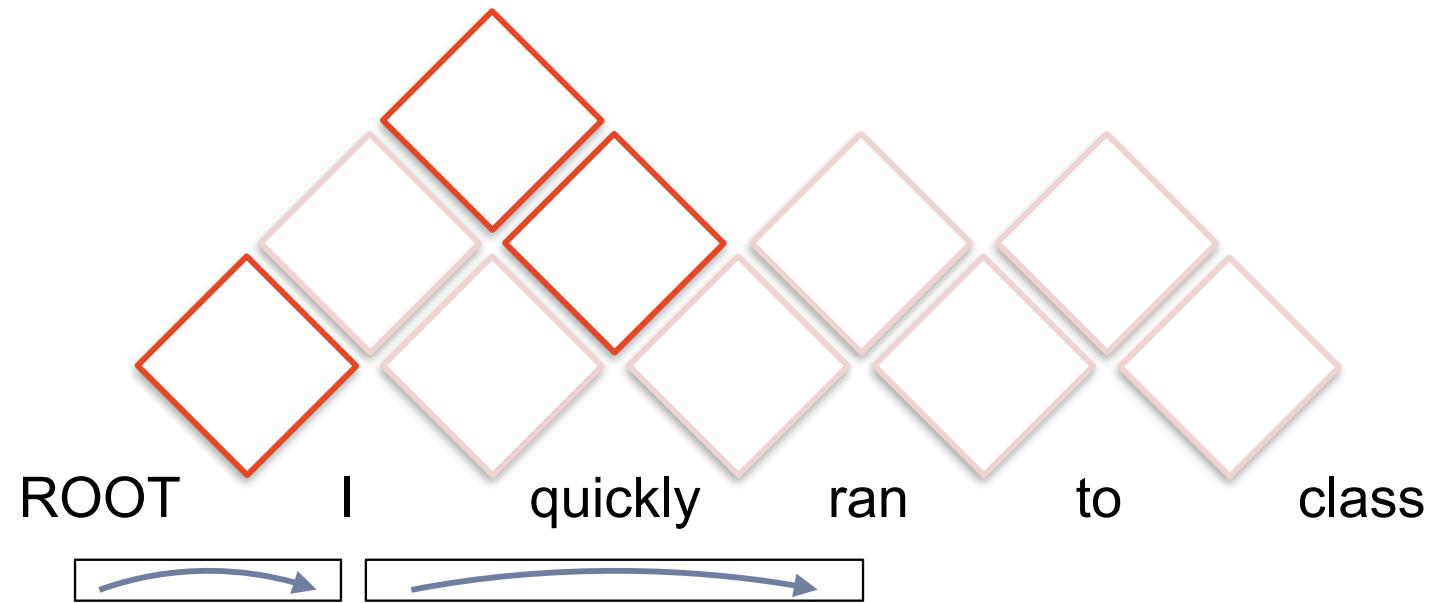
Can we go through the CYK algorithm for dependency parsing? /
Can you “Explain Like I’m 5” the CYK/CKY algorithm in dynamic
programming parsing in another way?

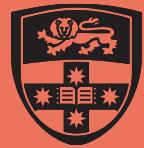




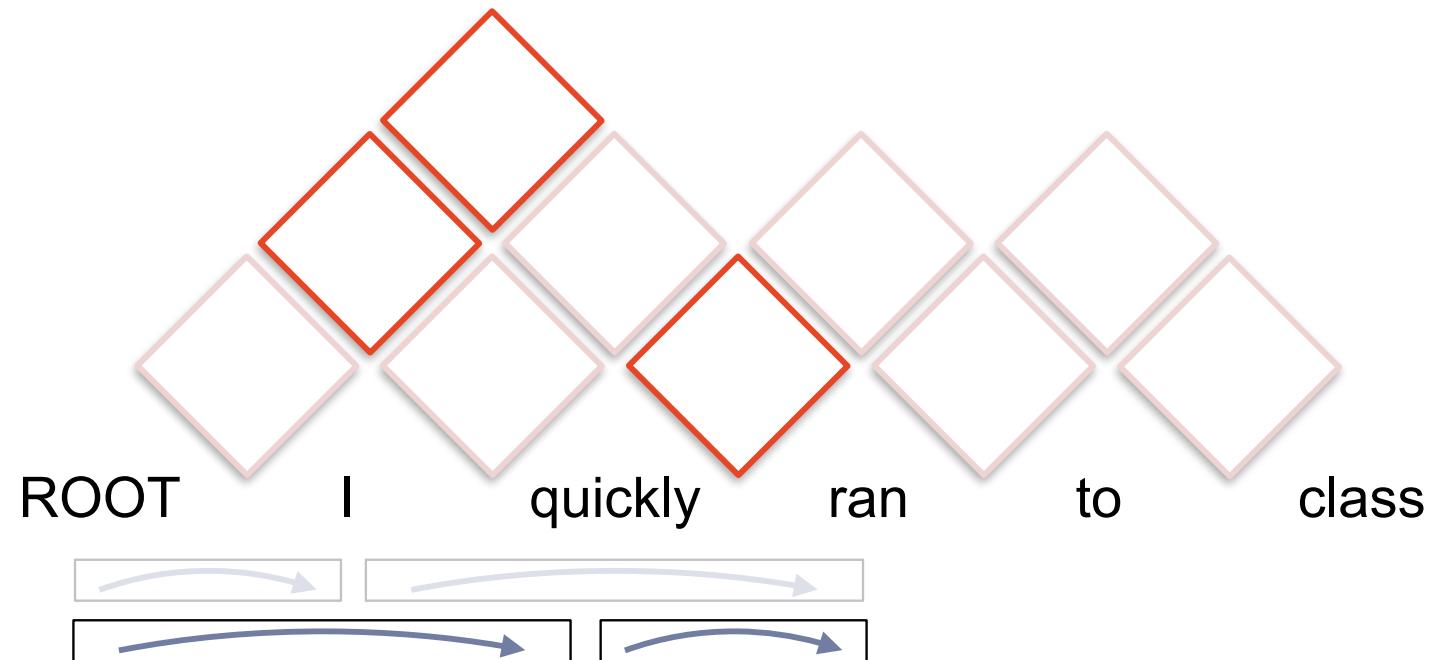
[menti.com 6574 1888](https://menti.com/65741888)

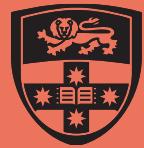
Can we go through the CYK algorithm for dependency parsing? /
Can you “Explain Like I’m 5” the CYK/CKY algorithm in dynamic
programming parsing in another way?



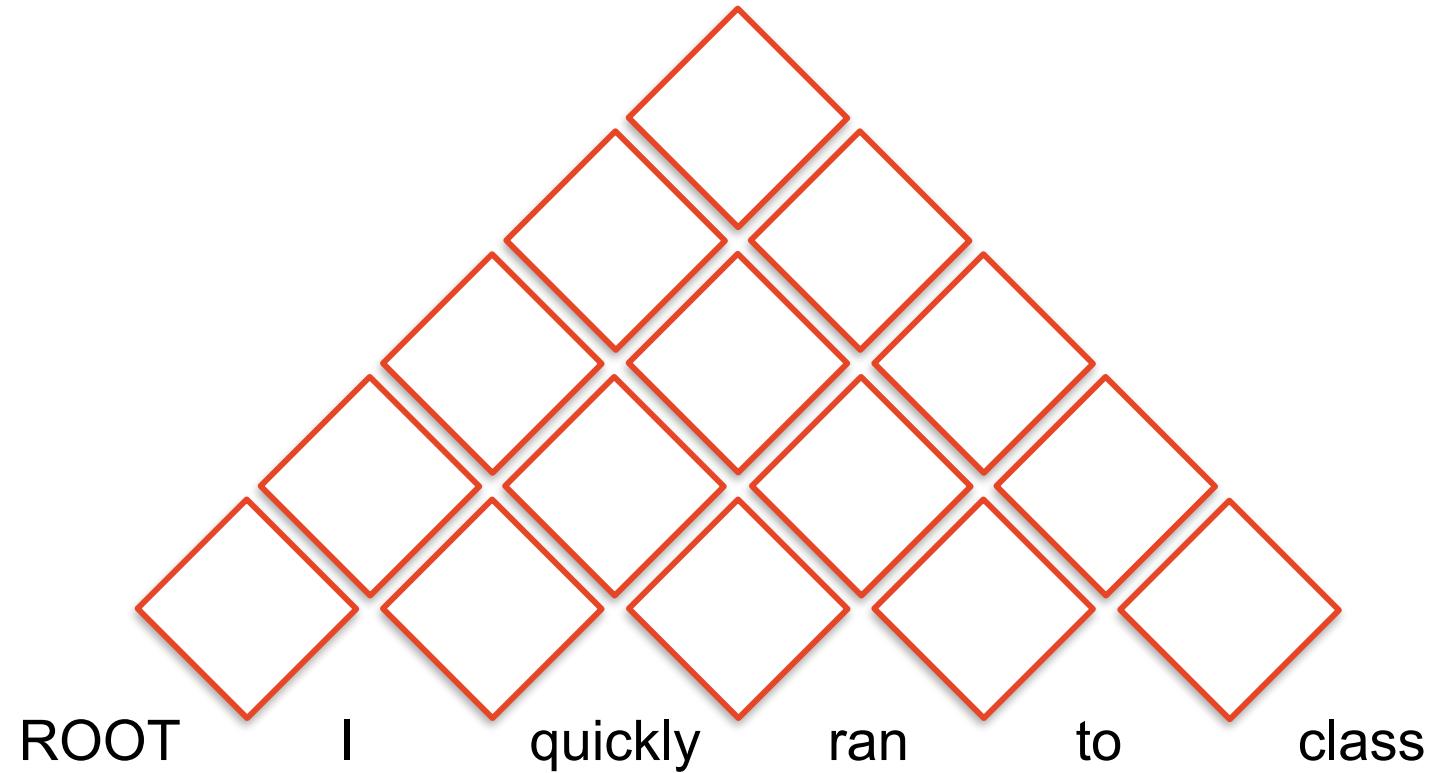


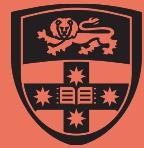
Can we go through the CYK algorithm for dependency parsing? /
Can you “Explain Like I’m 5” the CYK/CKY algorithm in dynamic
programming parsing in another way?



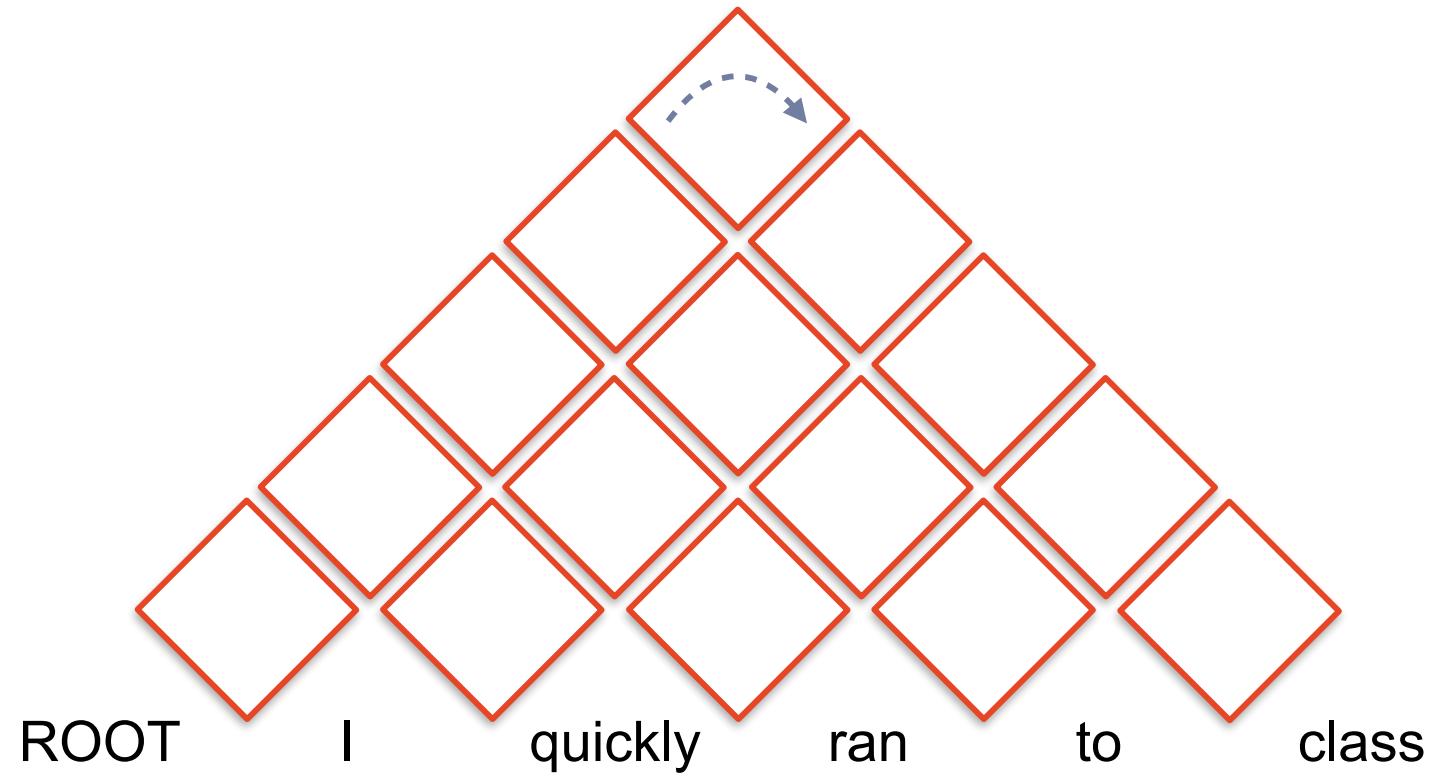


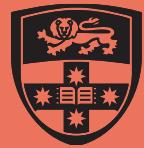
Can we go through the CYK algorithm for dependency parsing? /
Can you “Explain Like I’m 5” the CYK/CKY algorithm in dynamic
programming parsing in another way?



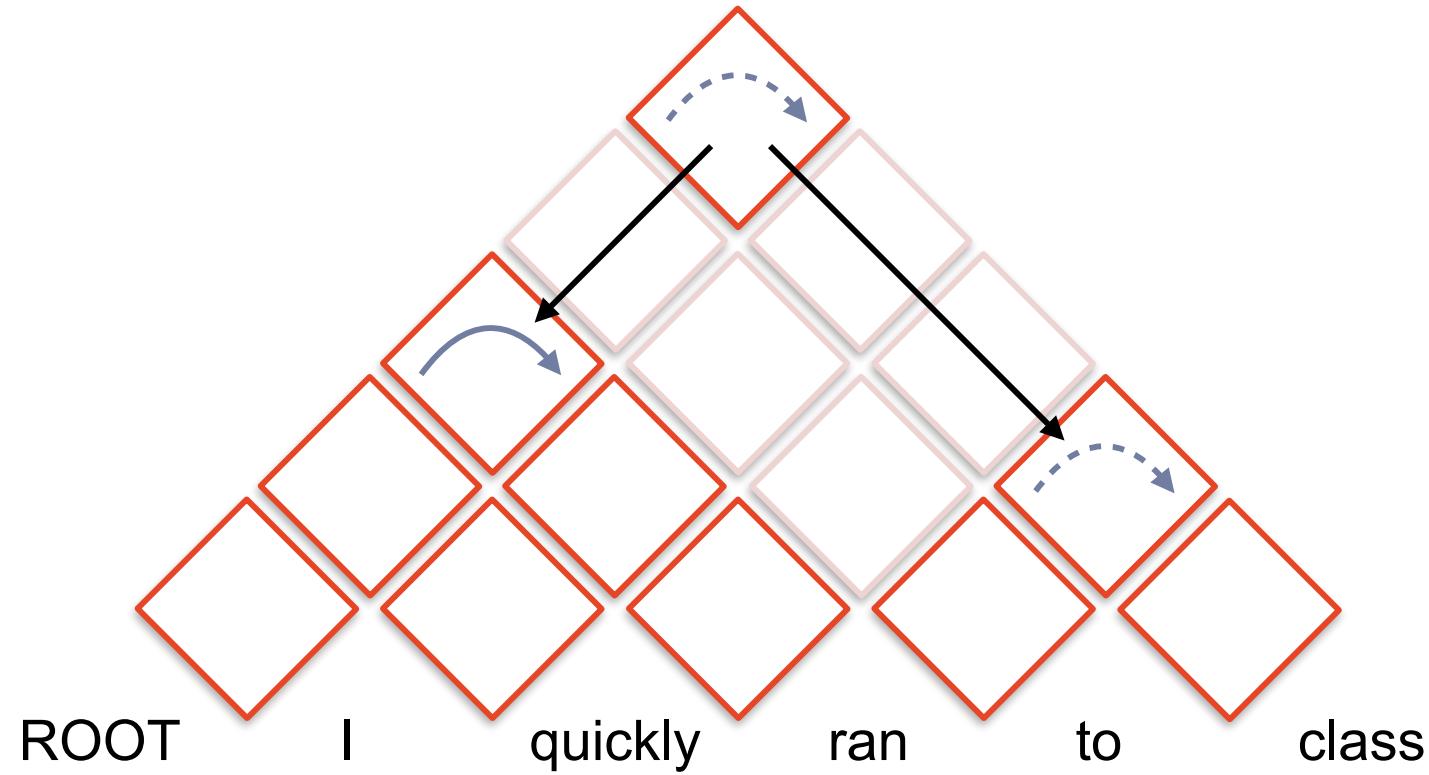


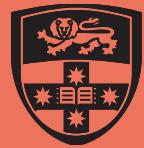
Can we go through the CYK algorithm for dependency parsing? /
Can you “Explain Like I’m 5” the CYK/CKY algorithm in dynamic
programming parsing in another way?





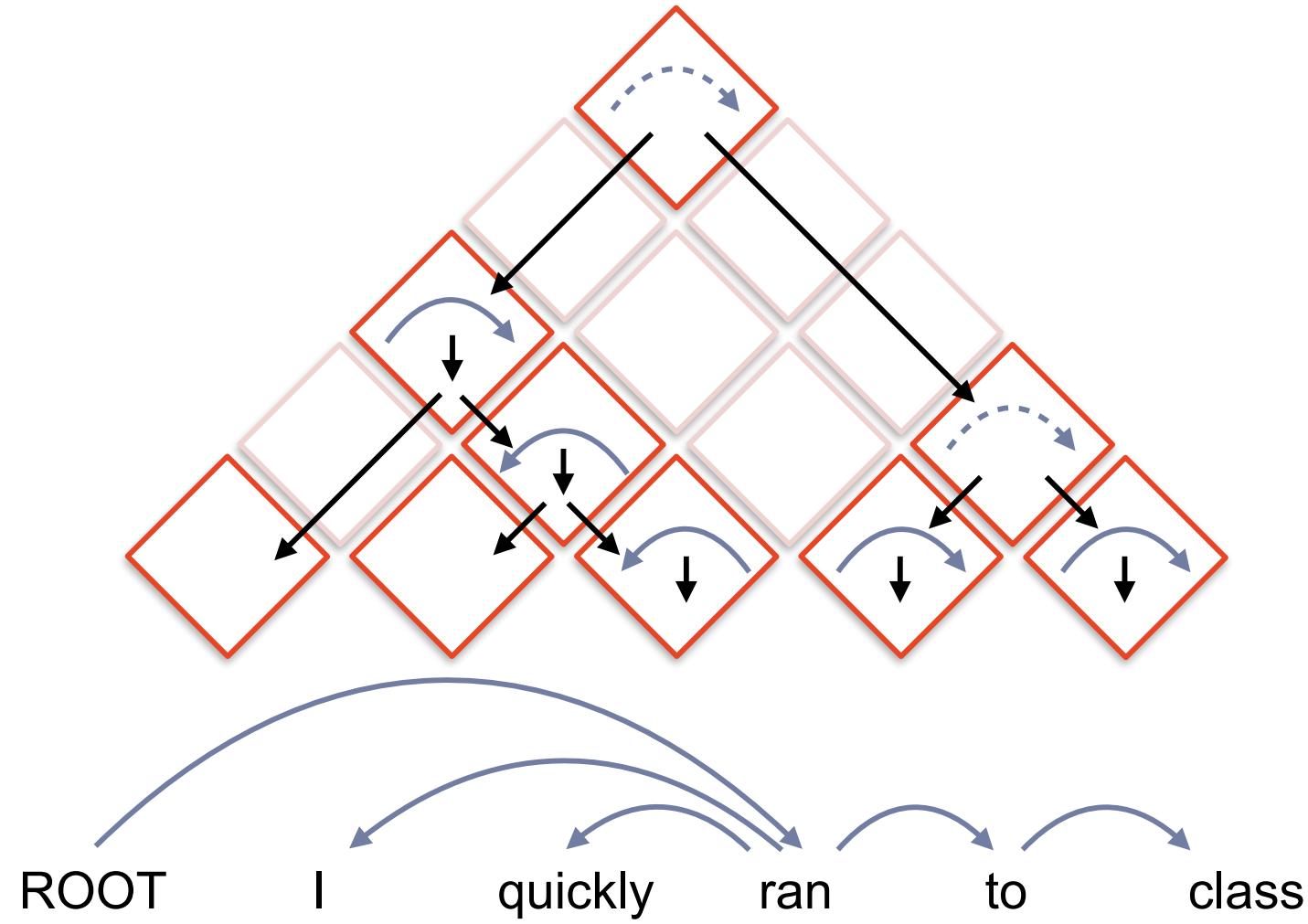
Can we go through the CYK algorithm for dependency parsing? /
Can you “Explain Like I’m 5” the CYK/CKY algorithm in dynamic
programming parsing in another way?

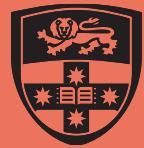




[menti.com 6574 1888](https://menti.com/65741888)

Can we go through the CYK algorithm for dependency parsing? /
Can you “Explain Like I’m 5” the CYK/CKY algorithm in dynamic
programming parsing in another way?





Bag of Words data storage methods

Vector

```
[0, 0, 0, 1, 0, 0, 0, 0, 0, 3, 1, 11, 0, ...]
```

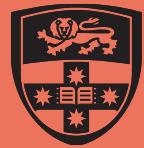
Sparse Vector

```
[(27, 10), (4, 1), (10, 3), (11, 1), (12, 11),  
(19, 12), (54, 6)]
```

Map

```
{27:10, 4:1, 10:3, 11:1, 12:11, 19:12, 54:6}
```

	Size	Optimal time to compare
Vector	$ \text{Vocab} $	Slowest $ \text{Vocab} $
Sparse Vector	$ \text{Unique tokens} $	$ \text{Unique tokens in shorter sequence} $
Map	Map uses slightly more	



General method for calculating the time and space complexity of algorithms so that we don't need to memorise them

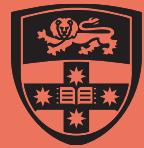
Time - Think about how many steps are needed as a function of input

Space - Think about how storage changes as a function of input



[menti.com 6574 1888](https://menti.com/65741888)

This is not a major focus of this unit



TF-IDF

$\text{count}(t, d)$ = how often token t occurs in document d

N = number of documents

df_t = how many documents token t occurs in

Term
Frequency

$$tf_{t,d} = \begin{cases} 1 + \log_{10} \text{count}(t, d) & \text{if } \text{count}(t, d) > 0 \\ 0 & \text{otherwise} \end{cases}$$

Inverse
Document
Frequency

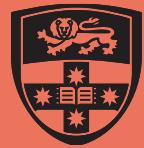
$$idf_t = \log_{10} \left(\frac{N}{df_t} \right)$$

We use log to rescale so that the first occurrence is more significant (going 0 to 1 matters more than 100 to 101)

TF-IDF

$$w_t = tf_{t,d} \times idf_t$$





Part of Speech tagging



NN NN VB DT NN



NN VBZ IN DT NN

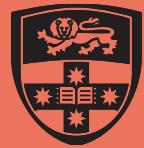
Time

flies

like

an

arrow



Named Entity Recognition

Person	Other	Other	Other
Jonathan	really	likes	chocolate

Identifying proper nouns:

Words	IO Label	BIO Label	BIOES Label
Jane	I-PER	B-PER	B-PER
Villanueva	I-PER	I-PER	E-PER
of	O	O	O
United	I-ORG	B-ORG	B-ORG
Airlines	I-ORG	I-ORG	I-ORG
Holding	I-ORG	I-ORG	E-ORG
discussed	O	O	O
the	O	O	O
Chicago	I-LOC	B-LOC	S-LOC
route	O	O	O
.	O	O	O



[menti.com 6574 1888](https://menti.com/65741888)



Coreference Resolution

Victoria Chen

Megabucks Banking

Lotsabucks

her pay

Victoria Chen, CFO of Megabucks

Banking, saw her pay jump to \$2.3

million, as the 38-year-old became

the company's president. It is widely

known that she came to Megabucks

from rival Lotsabucks.



[menti.com 6574 1888](https://menti.com/65741888)



My
Survey

menti.com
6574 1888



Kayak



University
Survey

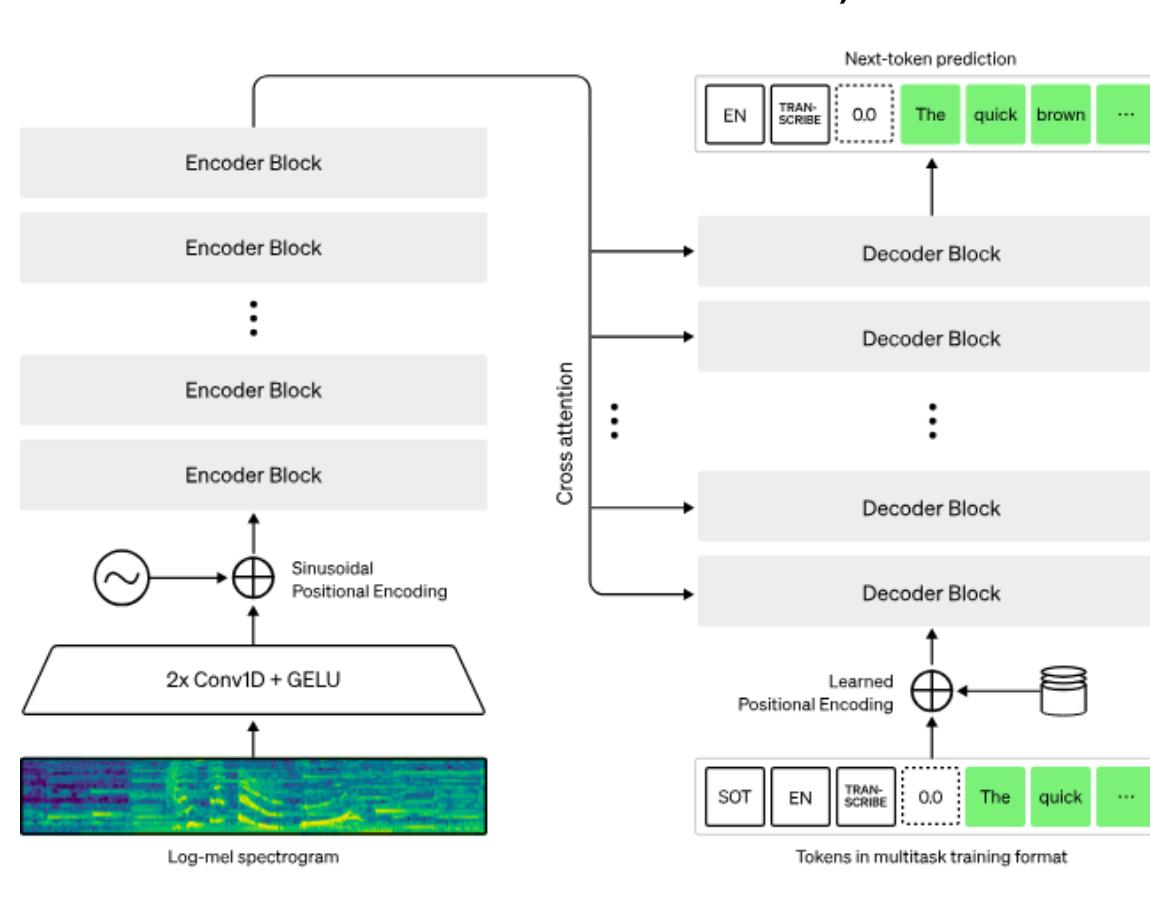
[Man, there's future *everywhere*.]



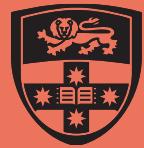
Source: <https://xkcd.com/2390/>



ASR automatic speech recognition, the mechanism behind whisper, the mechanism behind openAI realtime API (respond voice in seconds one no like traditional ASR)



“Under the hood, the Realtime API lets you create a persistent WebSocket connection to exchange messages with GPT-4o.”



Some depth and breadth in the whole quantization space for model training/serving. Breadth e.g. KV cache or AWQ.

KV cache

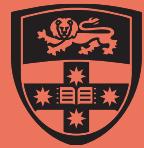
Store Keys and Values from attention calculations and avoid re-computation where possible

AWQ = Activation-aware Weight Quantization

Similar idea to LLM.int8(), but looking at activations rather than weights, i.e., rather than the static model weights, think about the values while processing an input



[menti.com 6574 1888](https://menti.com/65741888)



What are the 5 most important ideas in NLP?

Representations matter

Context matters

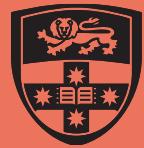
Data quantity and quality matter

Language is incredibly diverse and flexible

Bias comes in many forms that are difficult to detect



[menti.com 6574 1888](https://menti.com/65741888)

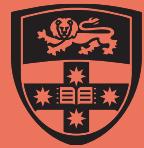


How to build and fine tune a LLM for your own task/industry,
including practical issue including multiGPU training, renting GPU,
hyper parameters tuning.

Some general thoughts....



[menti.com 6574 1888](https://menti.com/65741888)



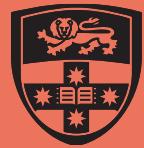
COMP 4446 / 5046
Lecture 13, 2025

What are the cutting edge questions or areas of research in NLP?



[menti.com 6574 1888](https://menti.com/65741888)

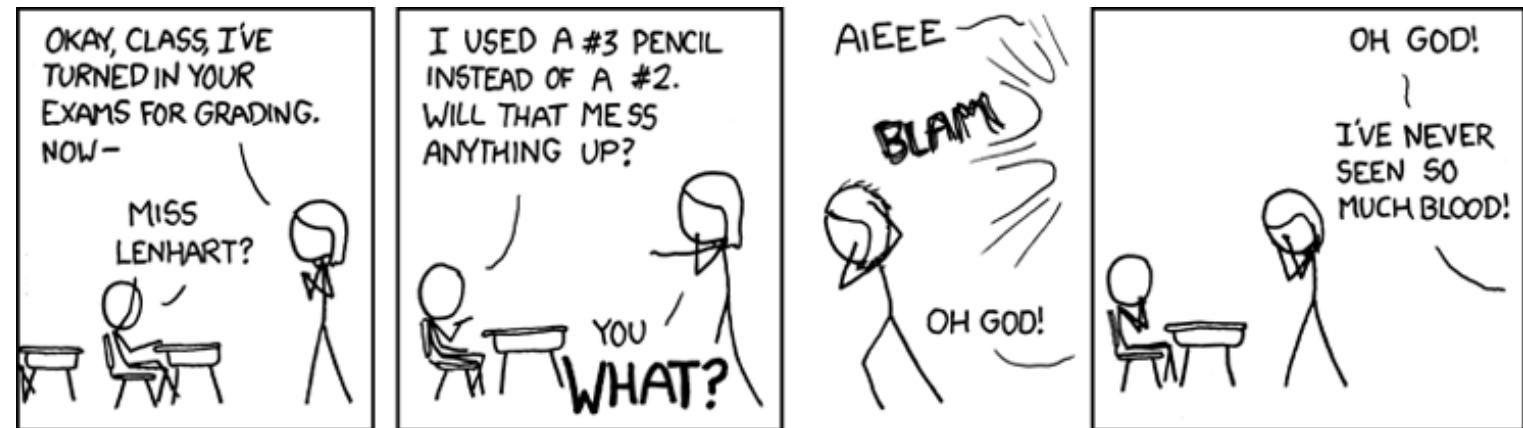
<https://westoahu.hawaii.edu/library/2020/09/07/clean-air-for-blue-skies/>



Muddy Card - <https://saipll.shinyapps.io/student-interface/> or on Ed



Thank you and
good luck on the exam!



[Also, after all the warnings about filling in the bubbles completely, I spent like 30 seconds on each one.]

Source: <https://xkcd.com/499/>



menti.com 6574 1888