

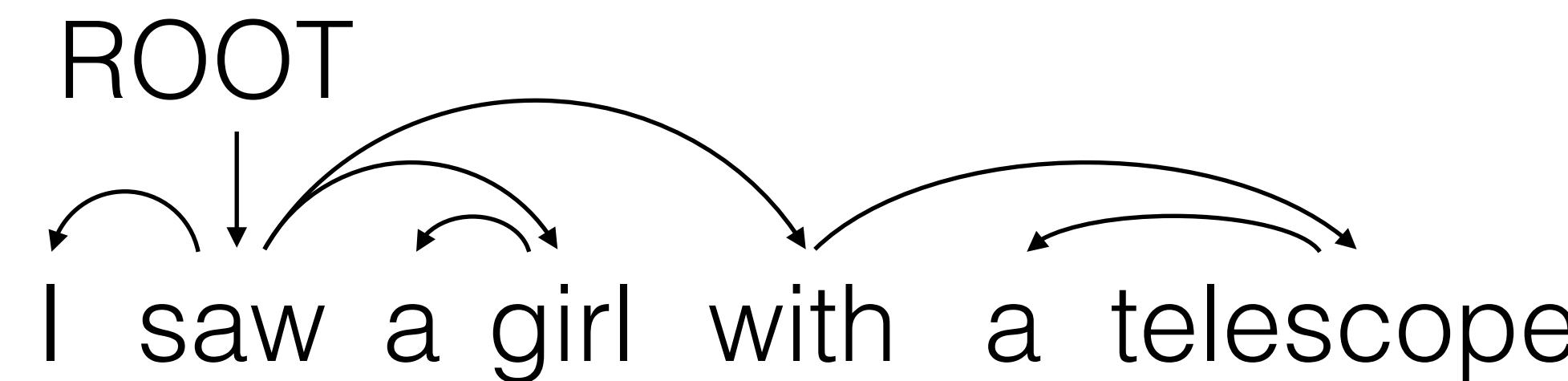
Dependency Parsing

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

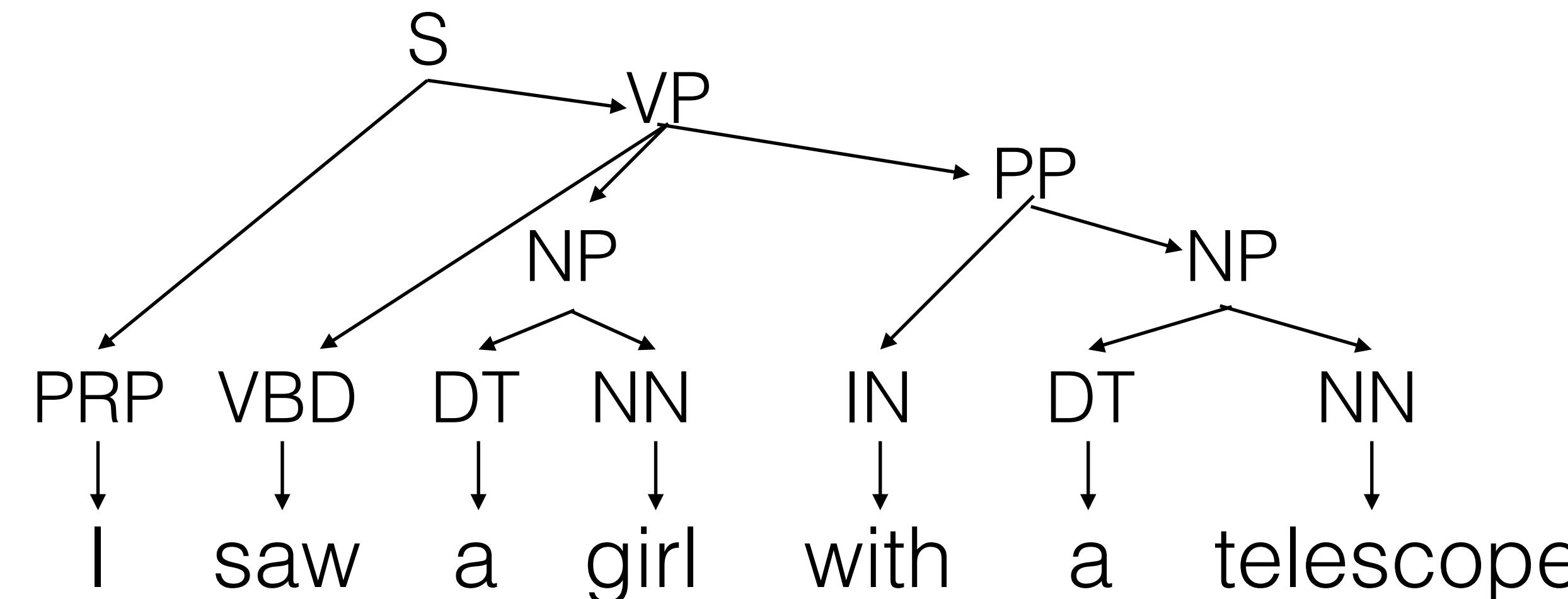


Two Types of Linguistic Structure

- **Dependency:** focus on relations between words

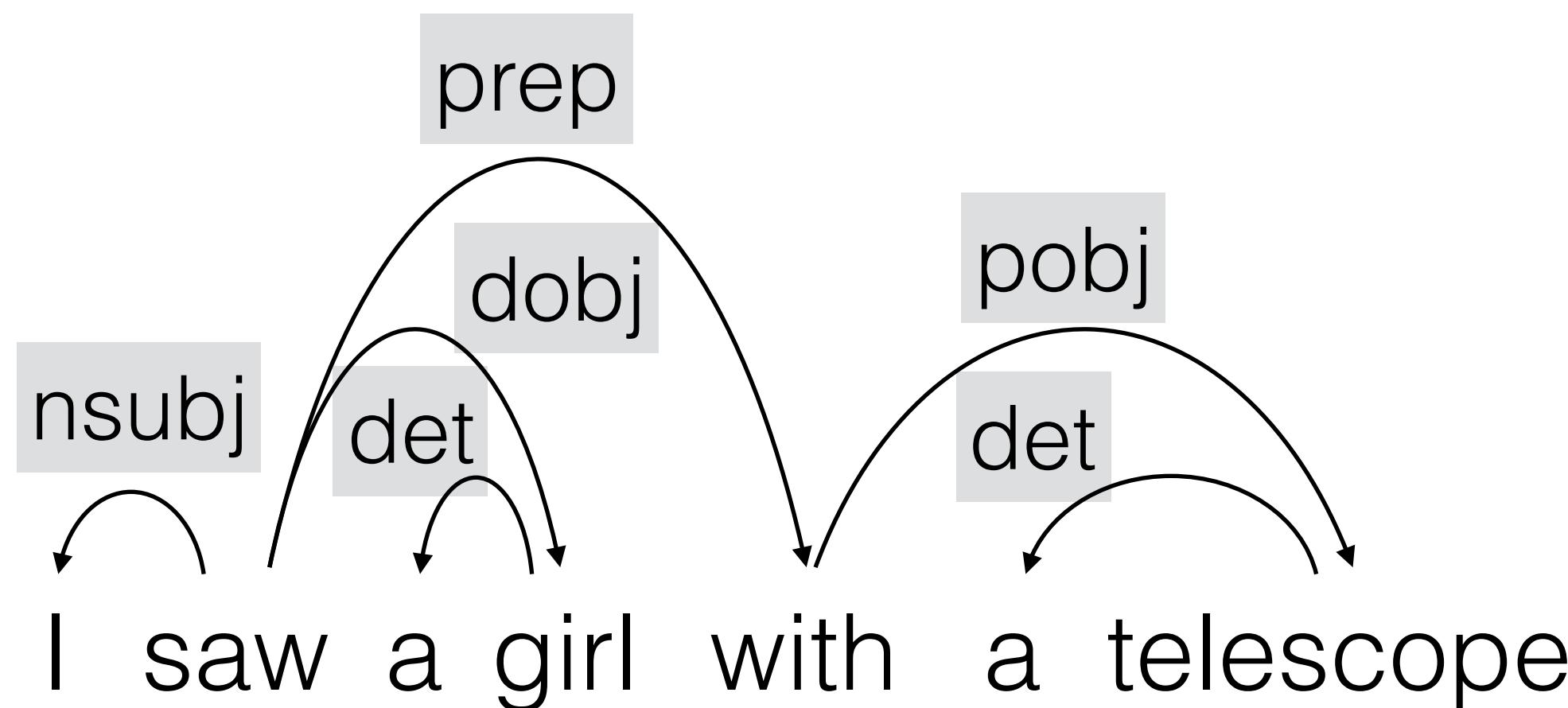


- **Phrase structure:** focus on the structure of the sentence



Why Dependencies?

- Dependencies are often good for semantic tasks, as related words are close in the tree
- It is also possible to create labeled dependencies, that explicitly show the relationship between words
- Easier to annotate and understand

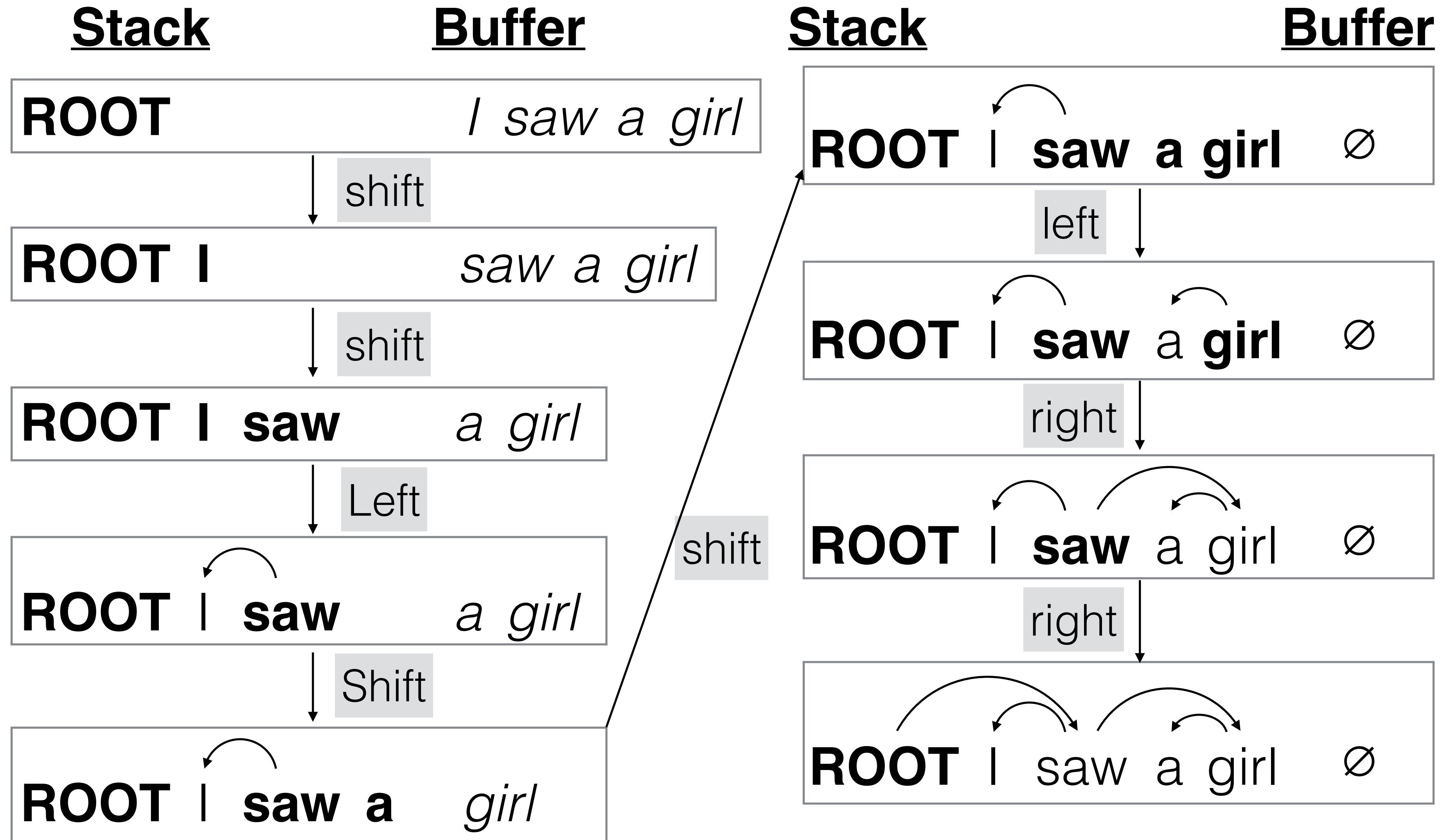


Arc Standard Shift-Reduce Parsing

(Yamada & Matsumoto 2003, Nivre 2003)

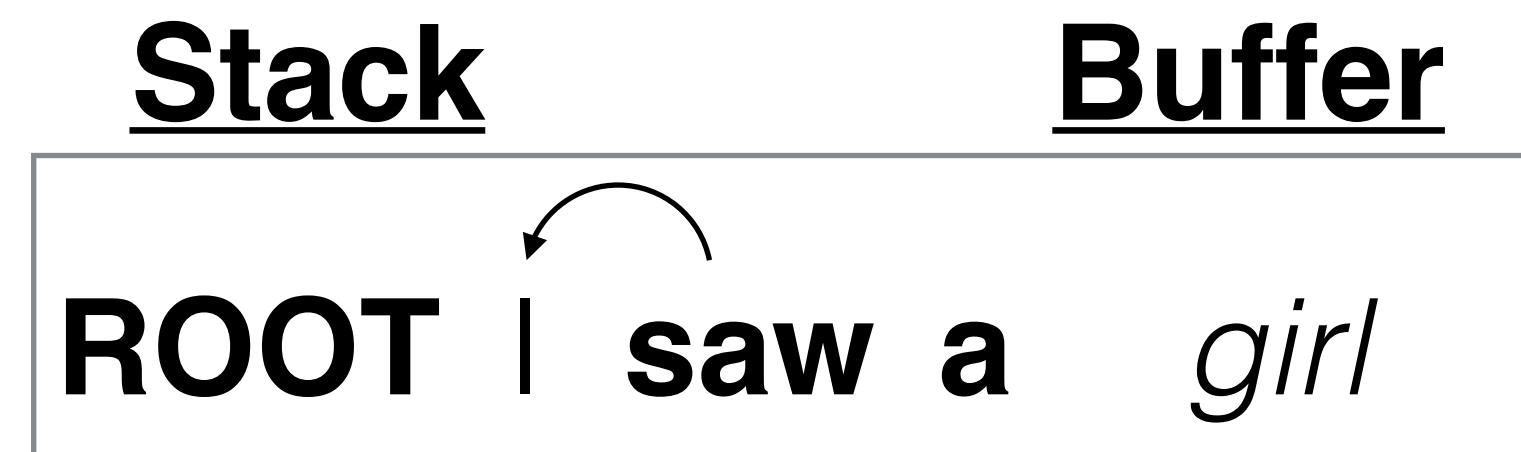
- Process words one-by-one left-to-right
- Two data structures
 - **Buffer:** of unprocessed words
 - **Stack:** of partially processed words
- At each point choose
 - **shift:** move one word from queue to stack
 - **reduce left:** top word on stack is head of second word
 - **reduce right:** second word on stack is head of top word
- Learn how to choose each action with a classifier

Shift Reduce Example

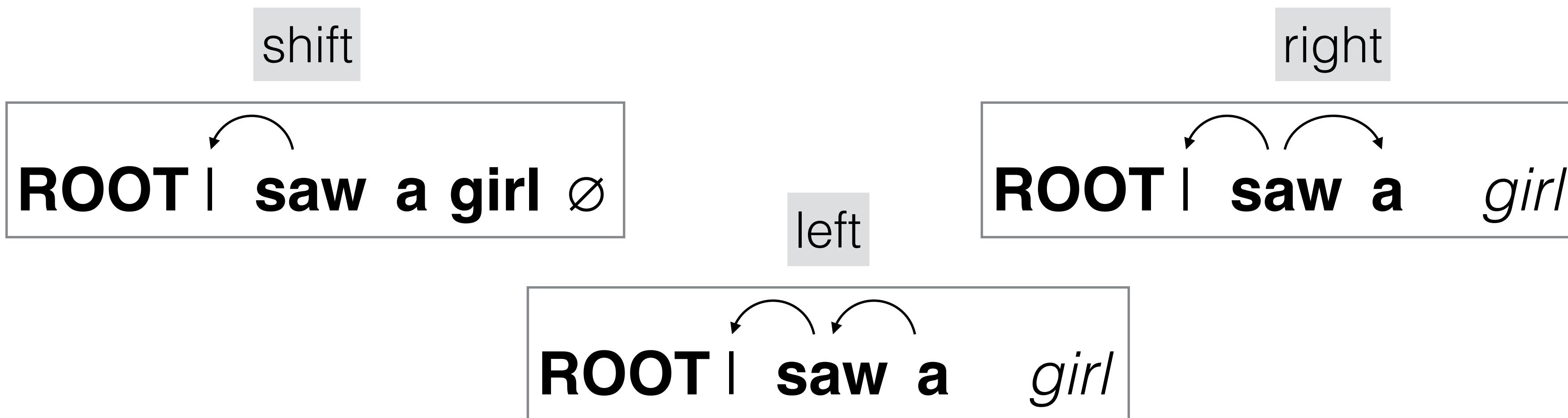


Classification for Shift-reduce

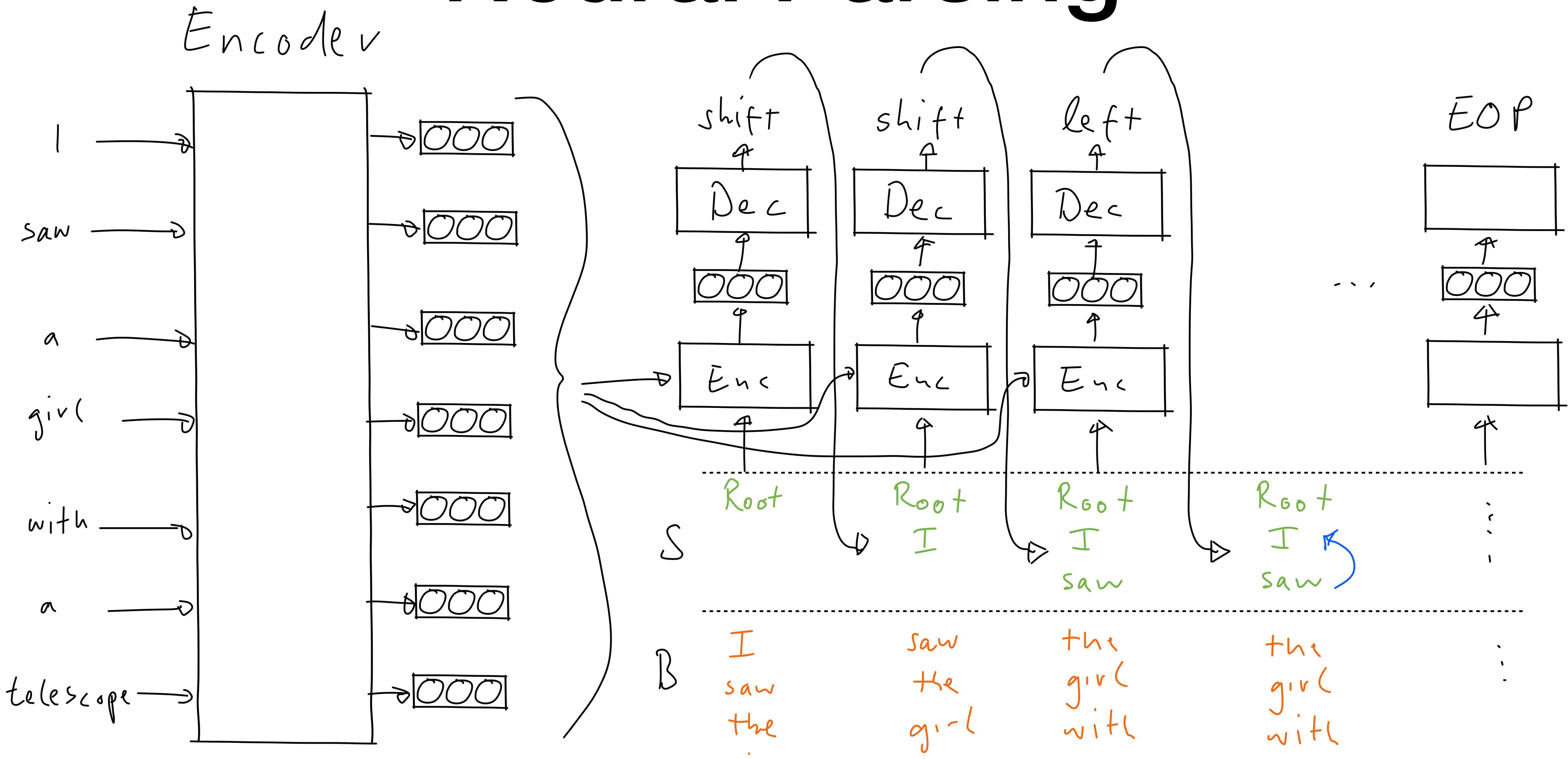
- Given a **configuration**



- Which **action** do we choose?



Neural Parsing

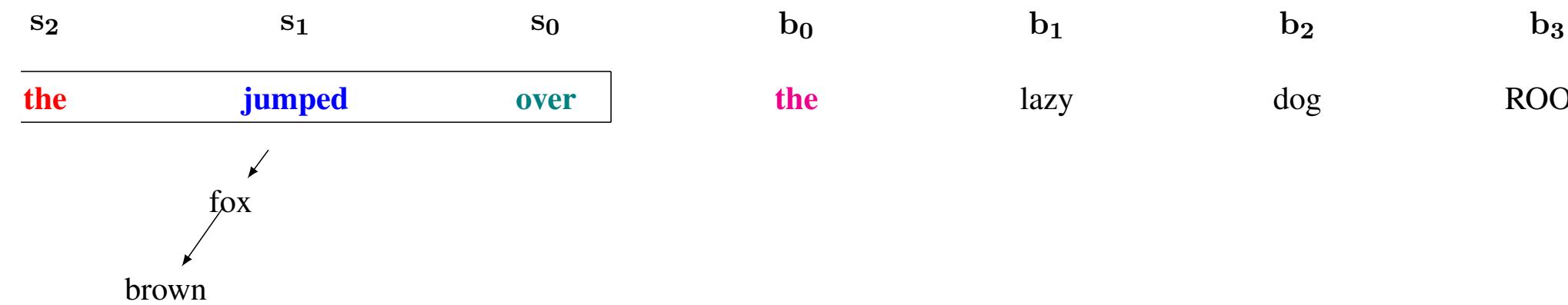


Neural Transition Based Parser

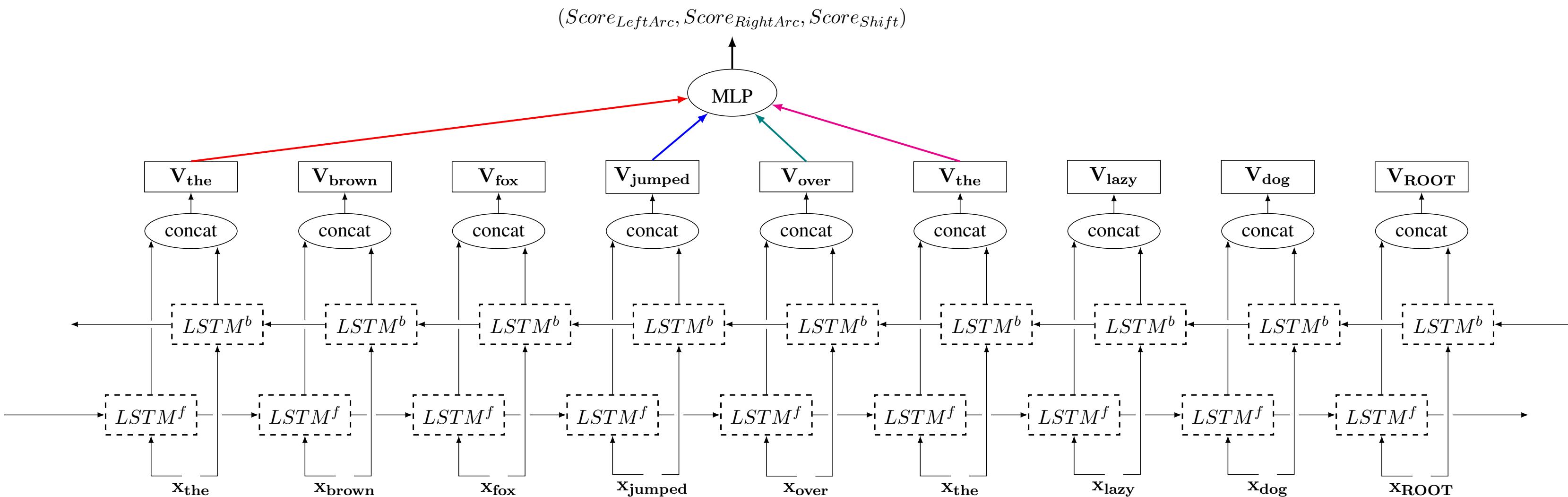
Figure here

Action Encoder-Decoder

Configuration:



Scoring:



LISA

Linguistically-Informed Self-Attention for Semantic Role Labeling



Emma
Strubell¹



Patrick
Verga¹



Daniel
Andor²



David
Weiss²

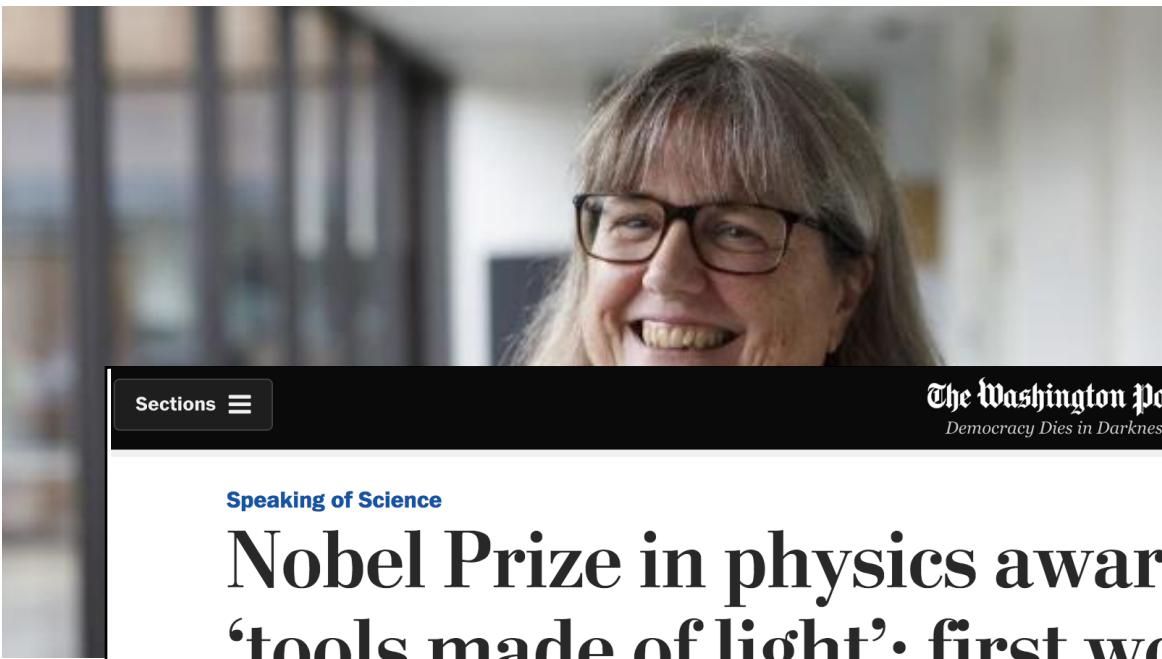


Andrew
McCallum¹

Want fast, accurate, robust NLP

Nobel laureate Donna Strickland: 'I see myself as a scientist, not a woman'

For Just the Third Time in 117 Years, a Woman Wins the Nobel Prize in Physics



The 2018 Nobel laureate in physics, Donna Strickland, has won the award for her work on lasers. She is the first woman to win the prize in physics since 1963.

Speaking of Science

Nobel Prize in physics awarded for ‘tools made of light’; first woman in 55 years honored

By Sarah Kaplan
October 2, 2018

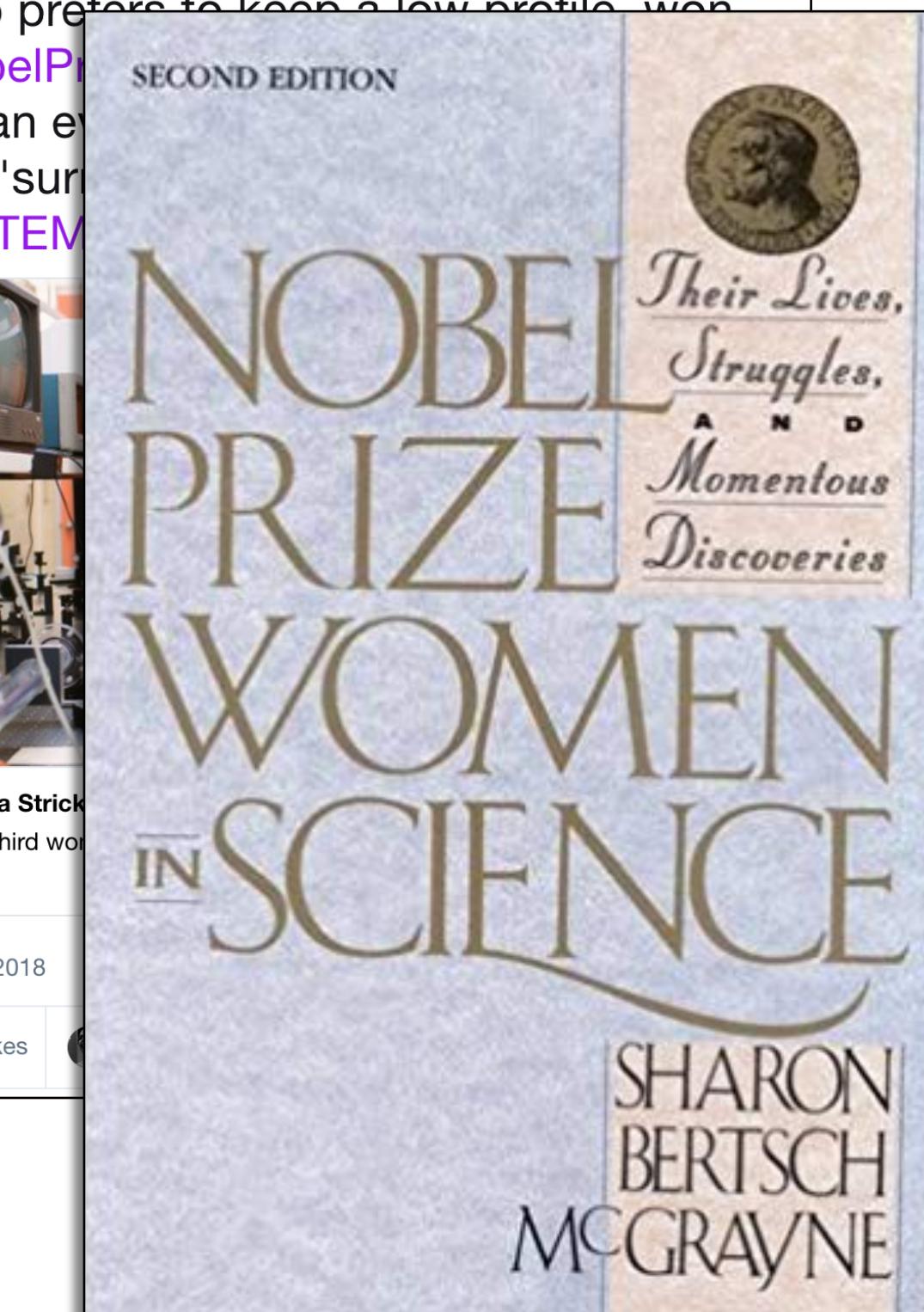
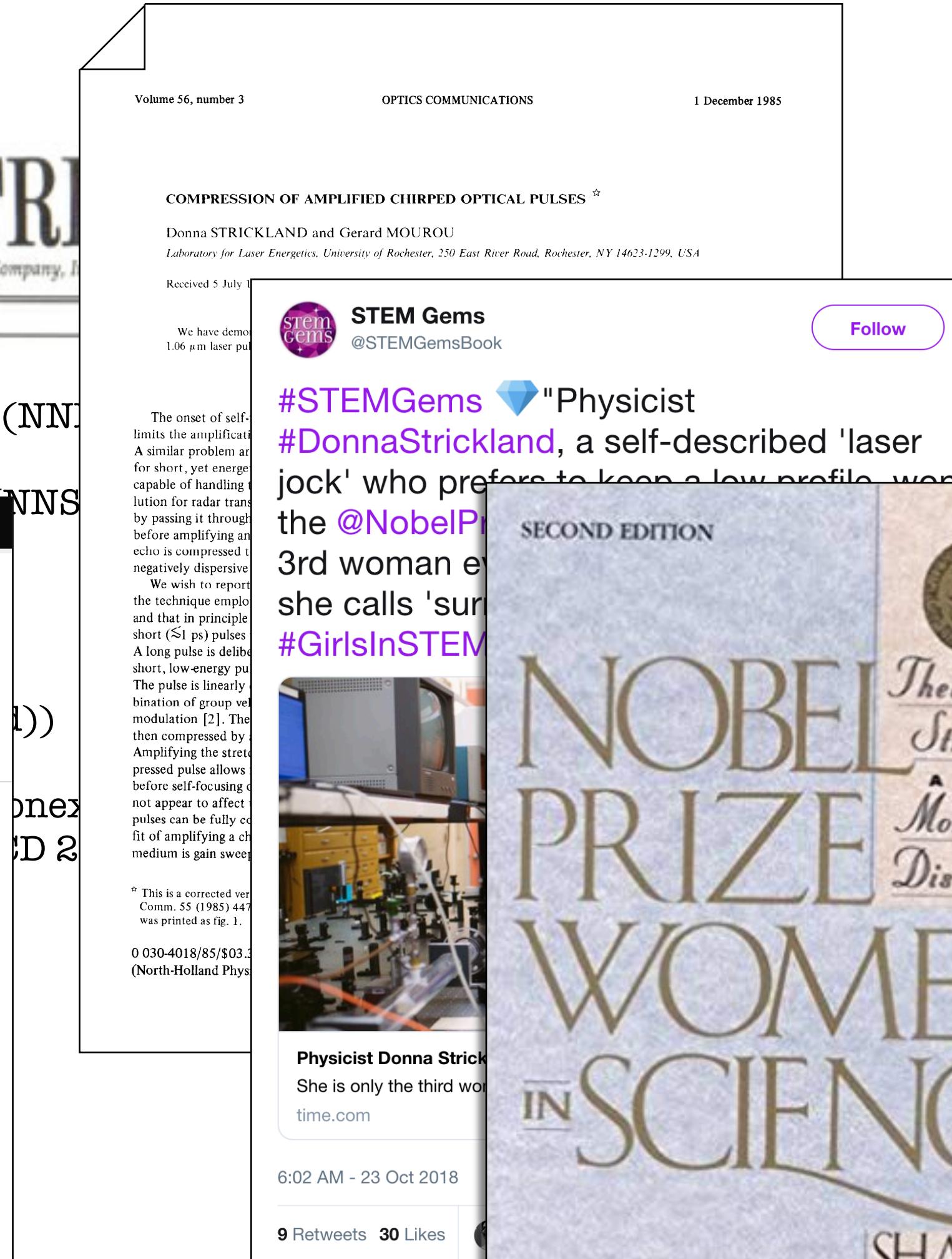
The 2018 Nobel Prize in physics was awarded Tuesday to Arthur Ashkin, Gérard Mourou and Donna Strickland for their pioneering work to turn lasers into powerful tools.

Ashkin, a researcher at Bell Laboratories in New Jersey, invented “optical tweezers” — focused beams of light that can be used to grab particles, atoms and even living cells and are now widely used to study the machinery of life.

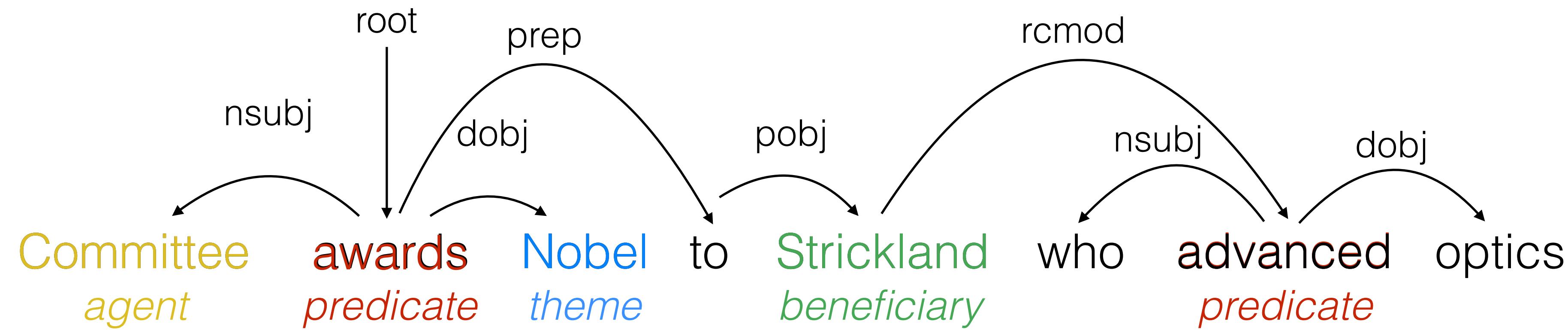
Mourou, of École Polytechnique in France and the University of Michigan, and Strickland, of the University of Waterloo in Canada, “paved the way” for the most powerful lasers ever created by humans via a technique that stretches and then amplifies the light beam.

“Billions of people make daily use of optical disk drive, laser printers and optical scanners ... millions undergo laser surgery,” Nobel committee member Olga Botner said. “The laser is truly one of the many examples of how a so-called blue sky discovery in a fundamental science eventually may transform our daily lives.”

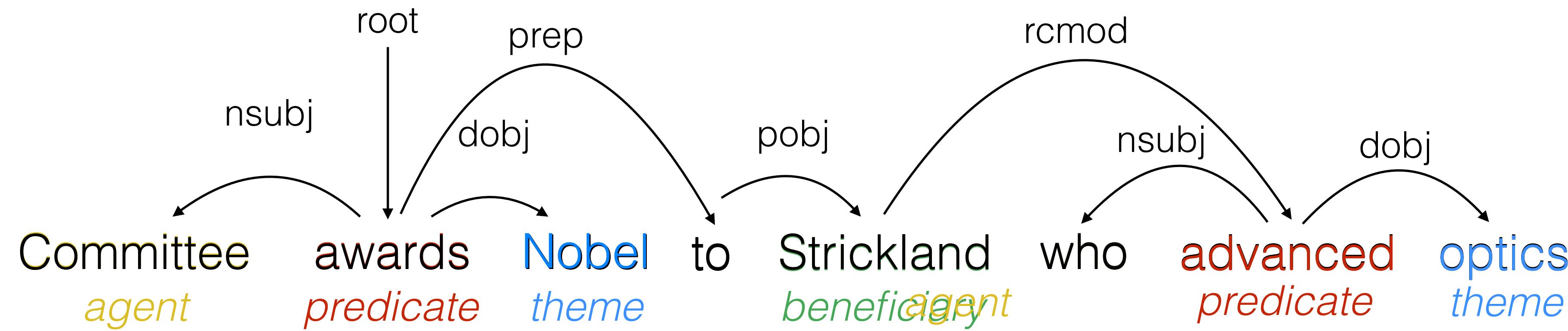
Strickland is the first woman to be awarded the physics prize since 1963, when Maria Goeppert-Mayer was recognized for her work on the structure of atomic nuclei. Marie Curie won the physics prize in 1903 and the chemistry Nobel Prize in 1911.



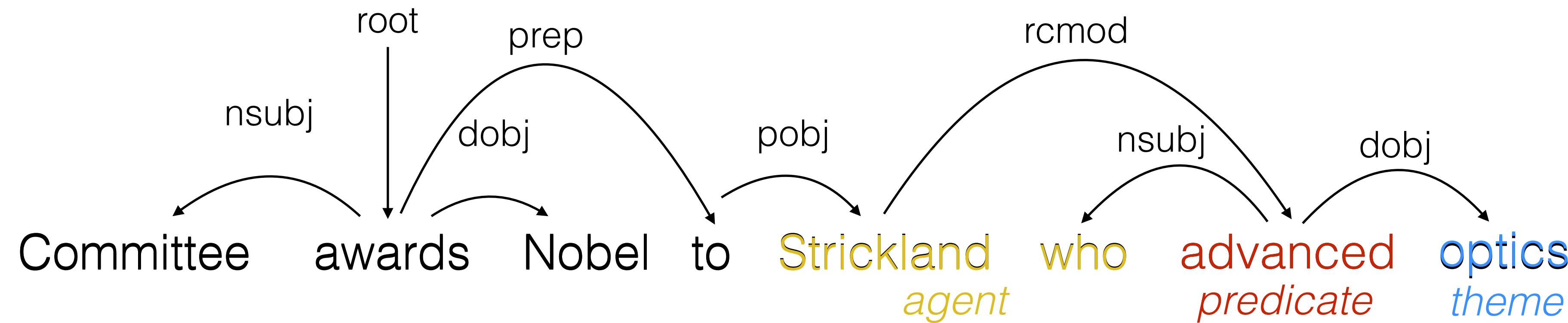
SRL: Who did what to whom?



SRL: Who did what to whom?

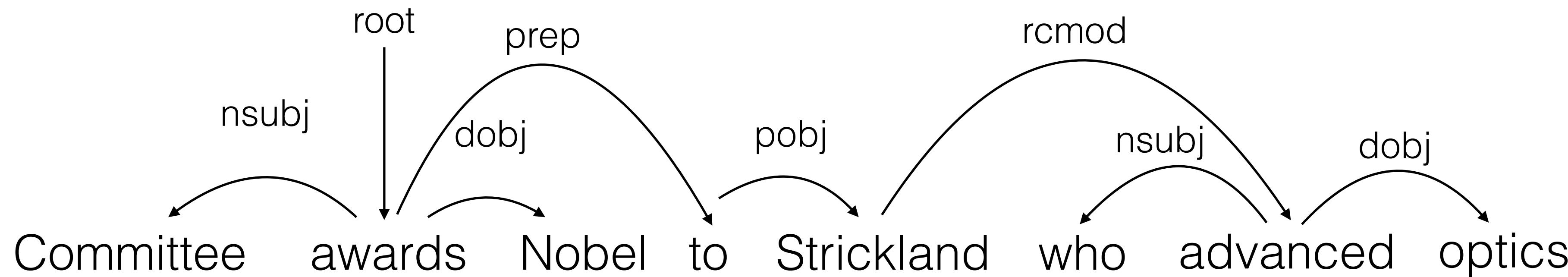


SRL: Who did what to whom?



Committee awards Nobel to Strickland who advanced optics
agent predicate theme beneficiary

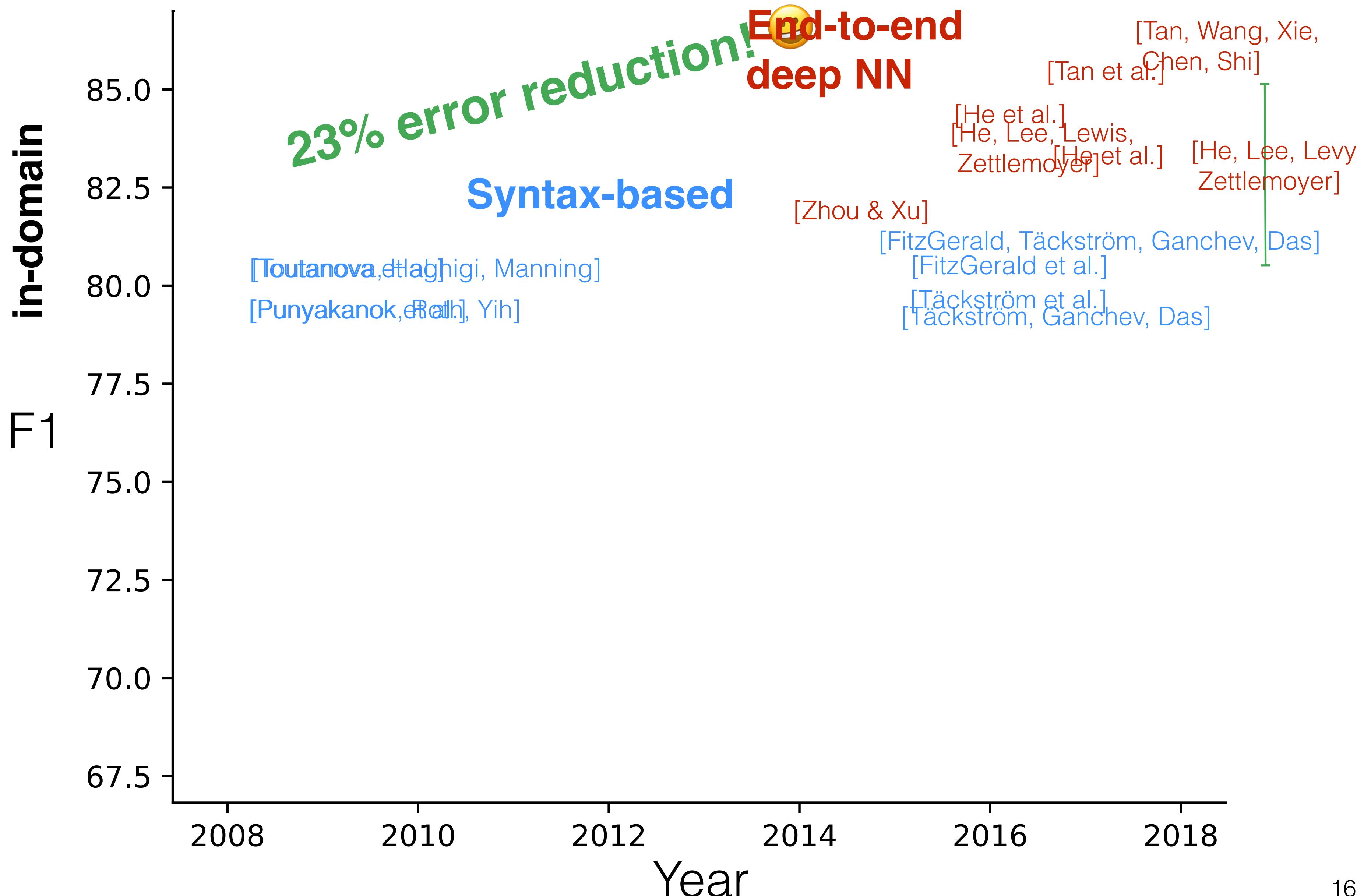
PropBank SRL: Who did what to whom?



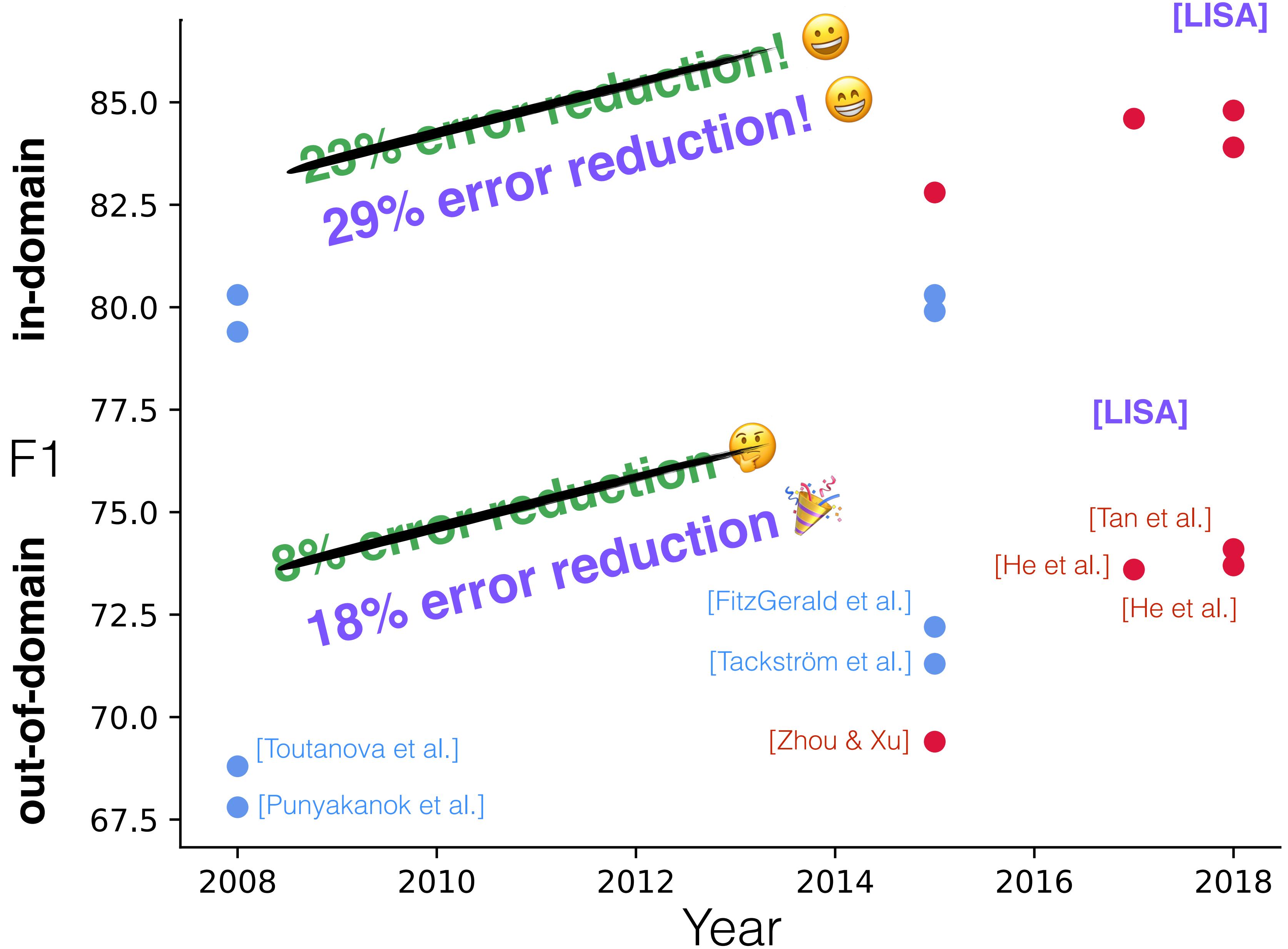
Committee awards Nobel to Strickland who advanced optics
ARG₀ R-ARG₀ predicate tARG₀

Committee Agent awards predicate Nobel tARG₀ to Strickland beneficiary who advanced optics

10 years of PropBank SRL

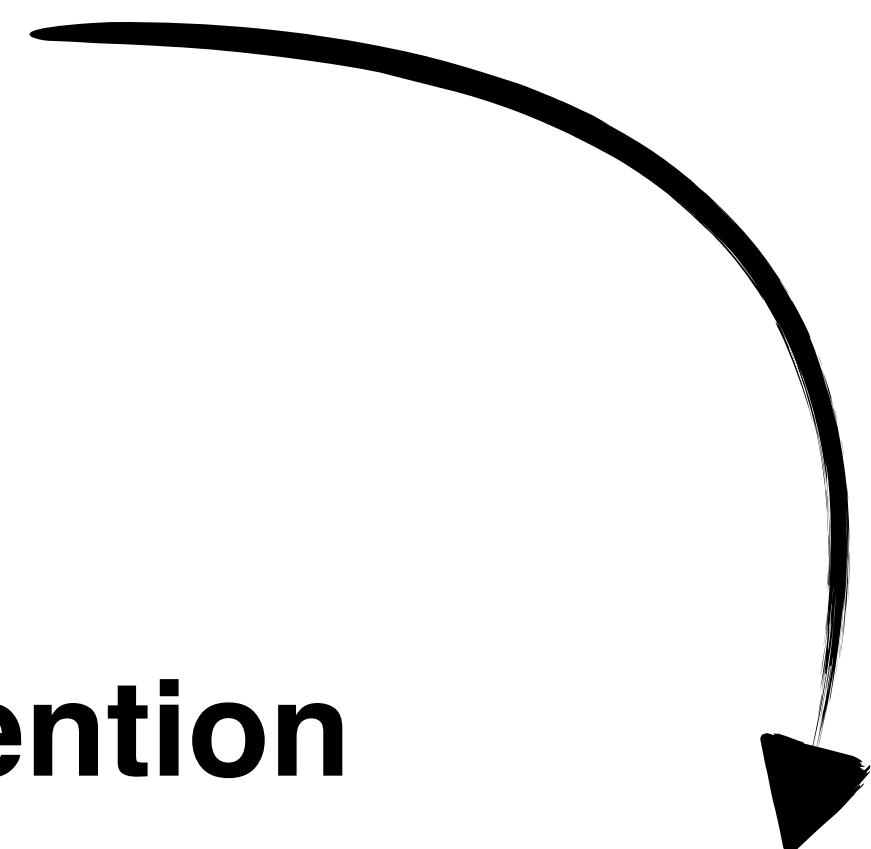
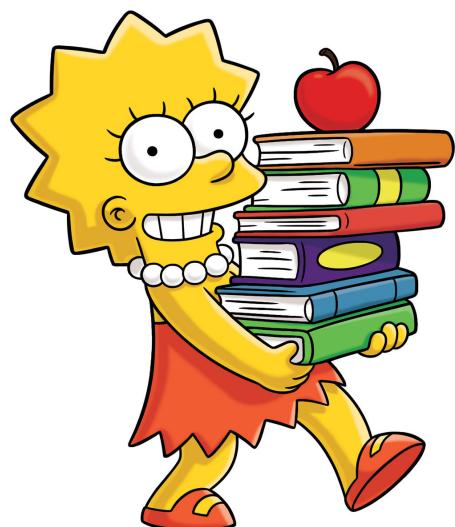


10 years of PropBank SRL



Linguistically-Informed Self-Attention

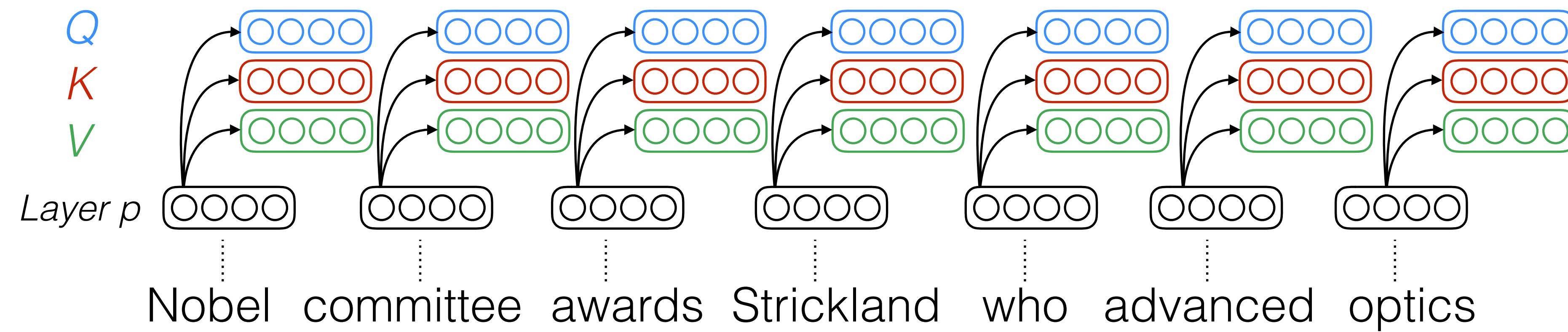
- **Multi-task learning,**
 - Part-of-speech tagging
 - Labeled dependency parsing
 - Predicate detection
 - Semantic role spans & labeling
- **Syntactically-informed self-attention**
 - Multi-head self-attention supervised by **syntax**



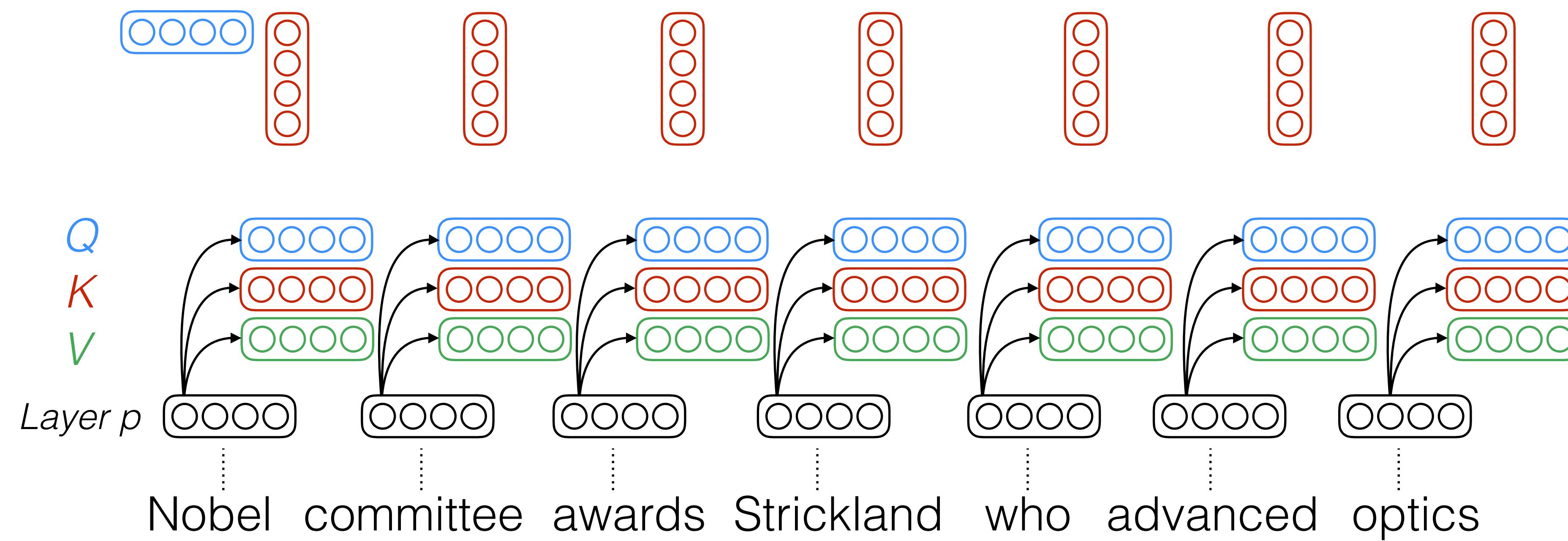
Outline

- Want fast, accurate, robust NLU
- PropBank SRL: Who did what to whom?
- 10 years of PropBank SRL
- LISA: Linguistically-informed self attention
 - Multi-head self-attention [Vaswani et al. 2017]
 - Syntactically-informed self-attention
 - Multi-task learning, single-pass inference
 - Experimental results & error analysis

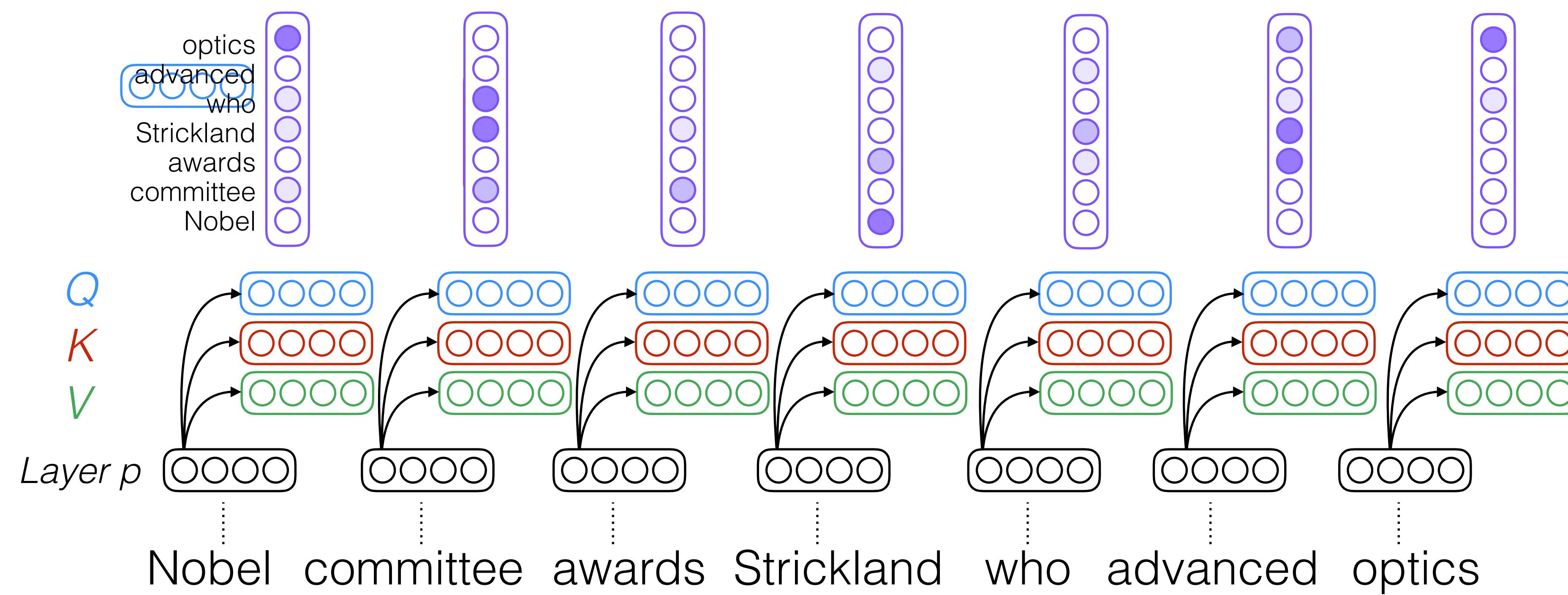
Self-attention



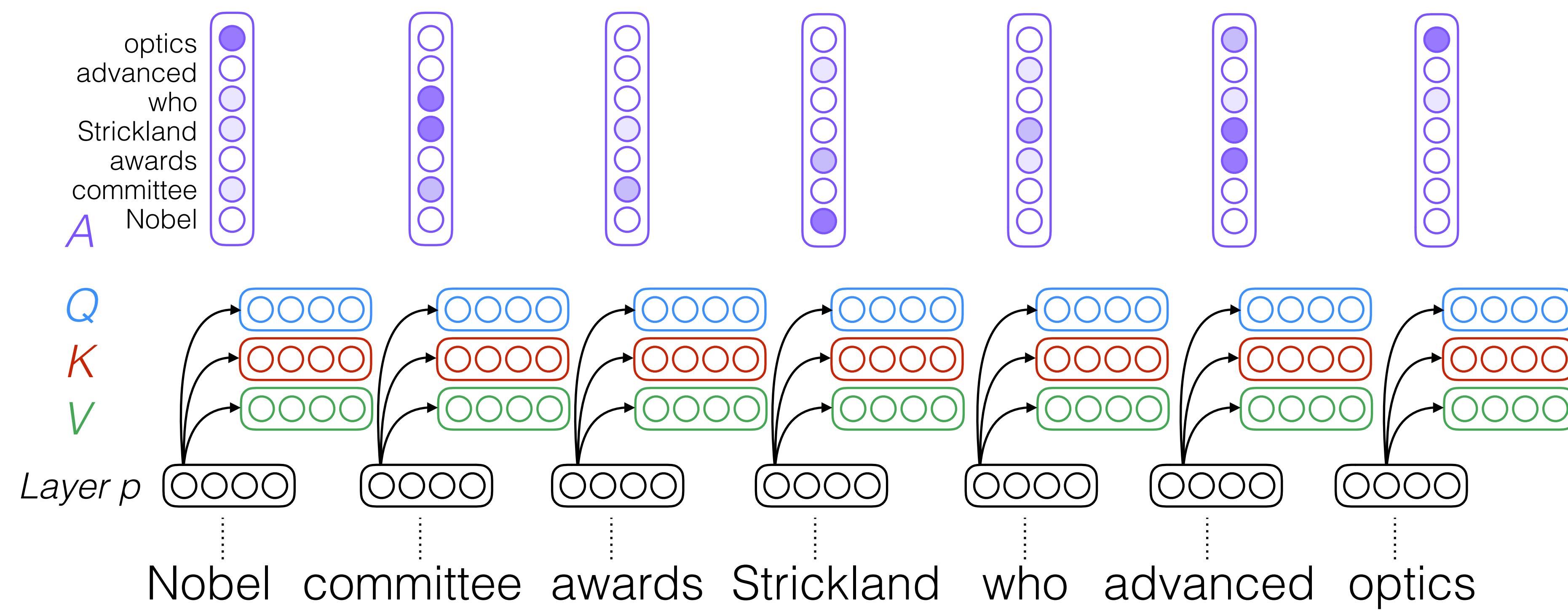
Self-attention



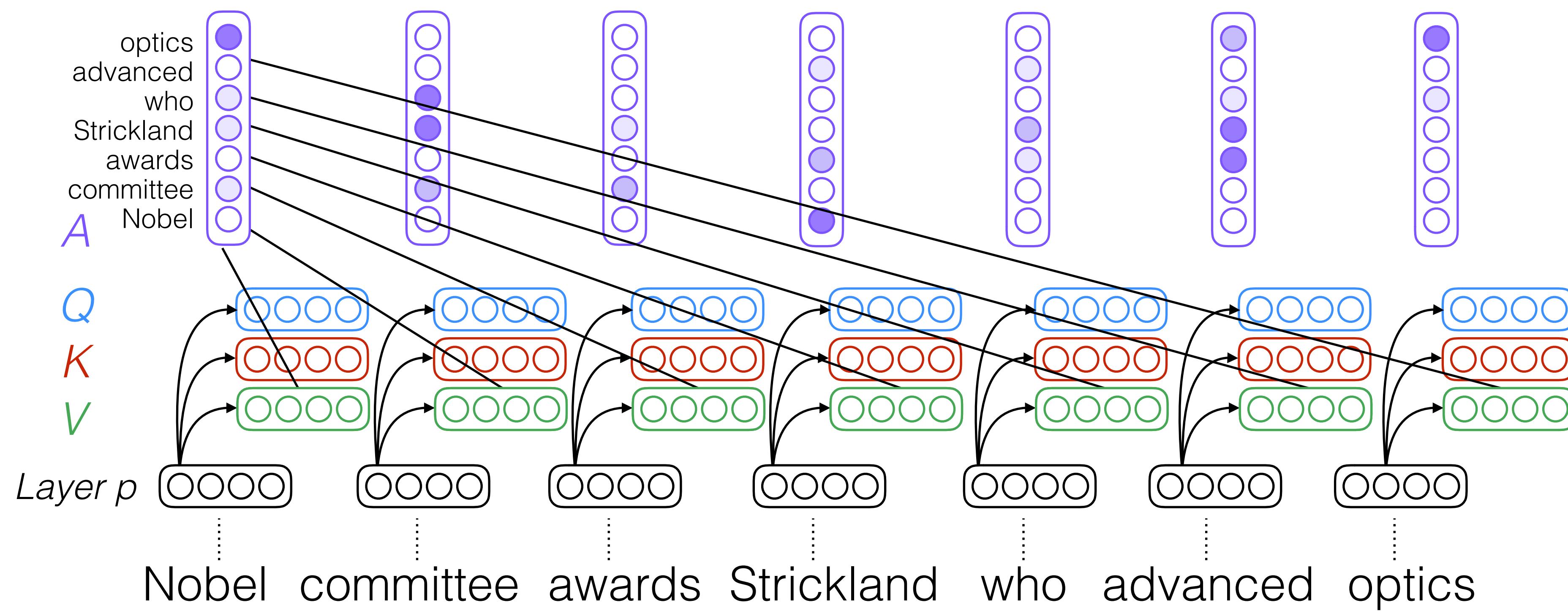
Self-attention



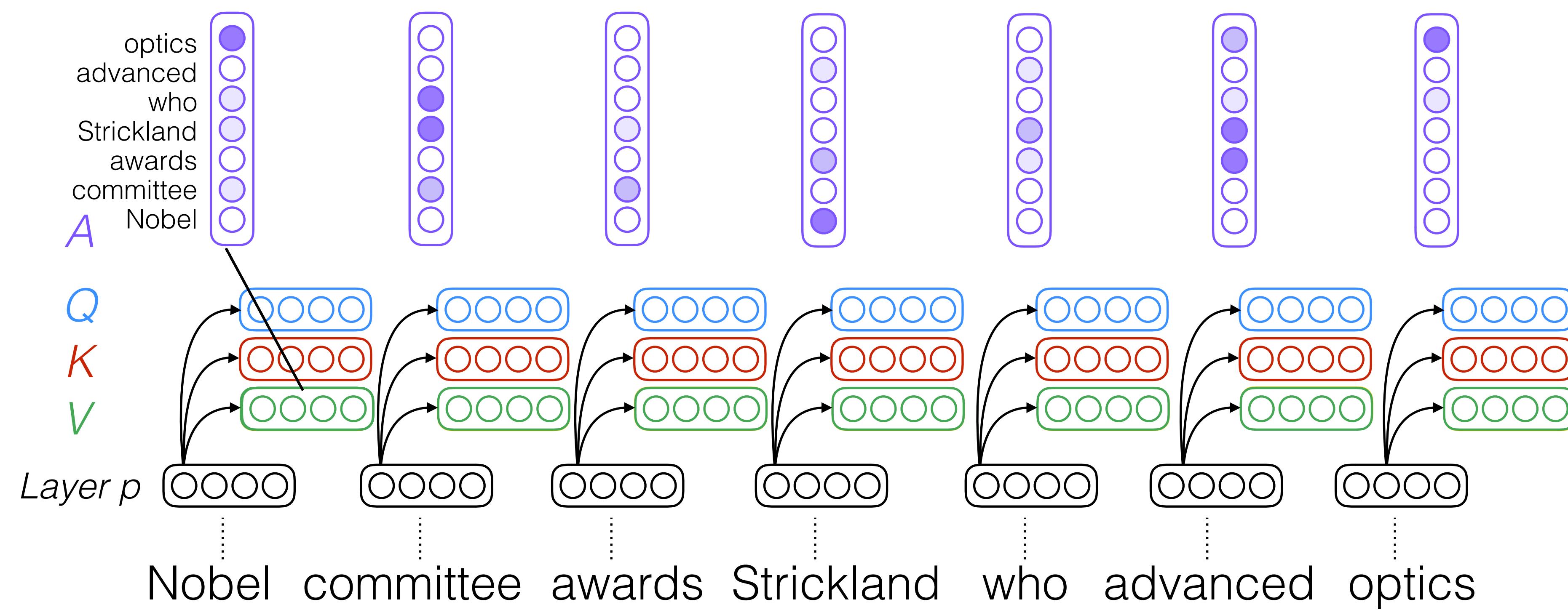
Self-attention



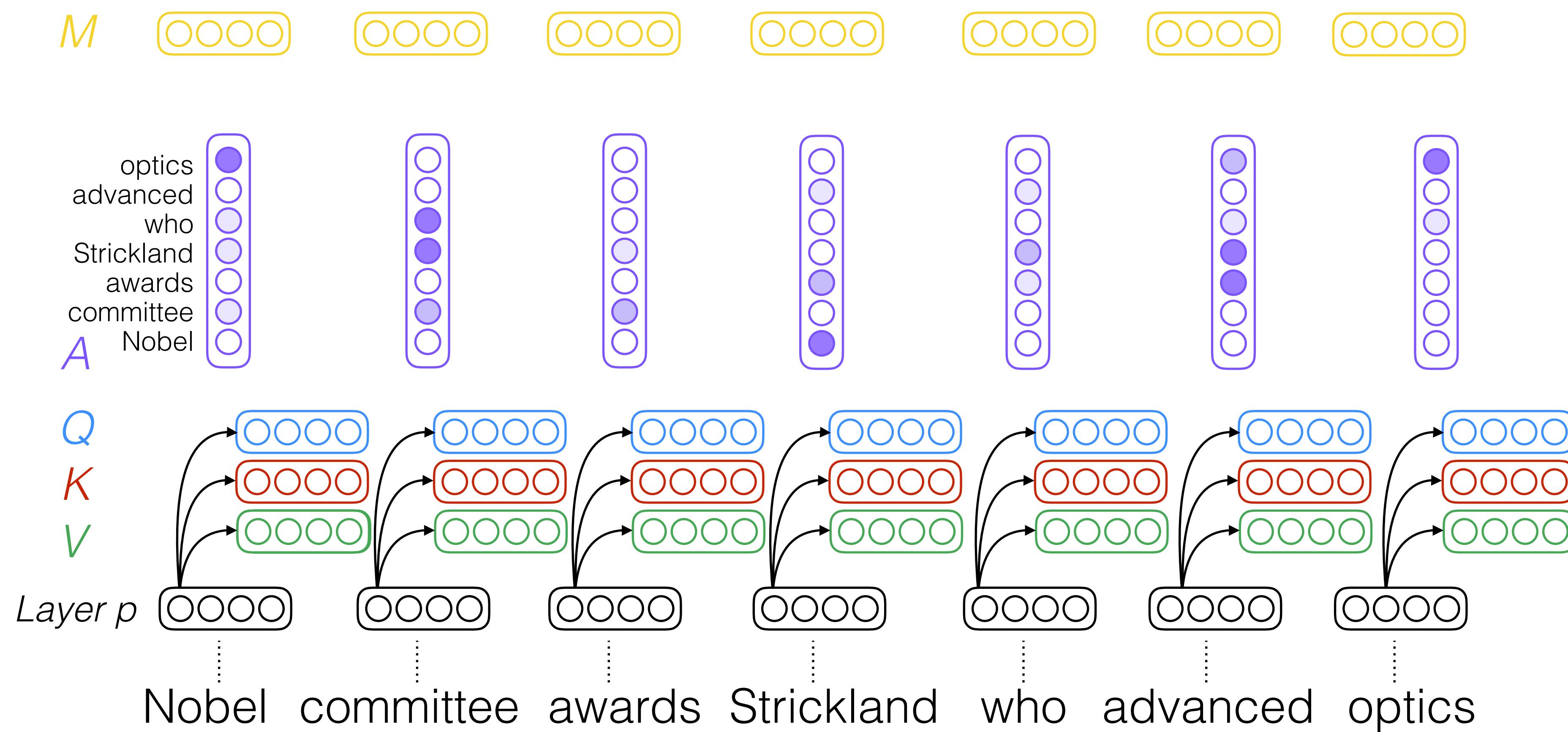
Self-attention



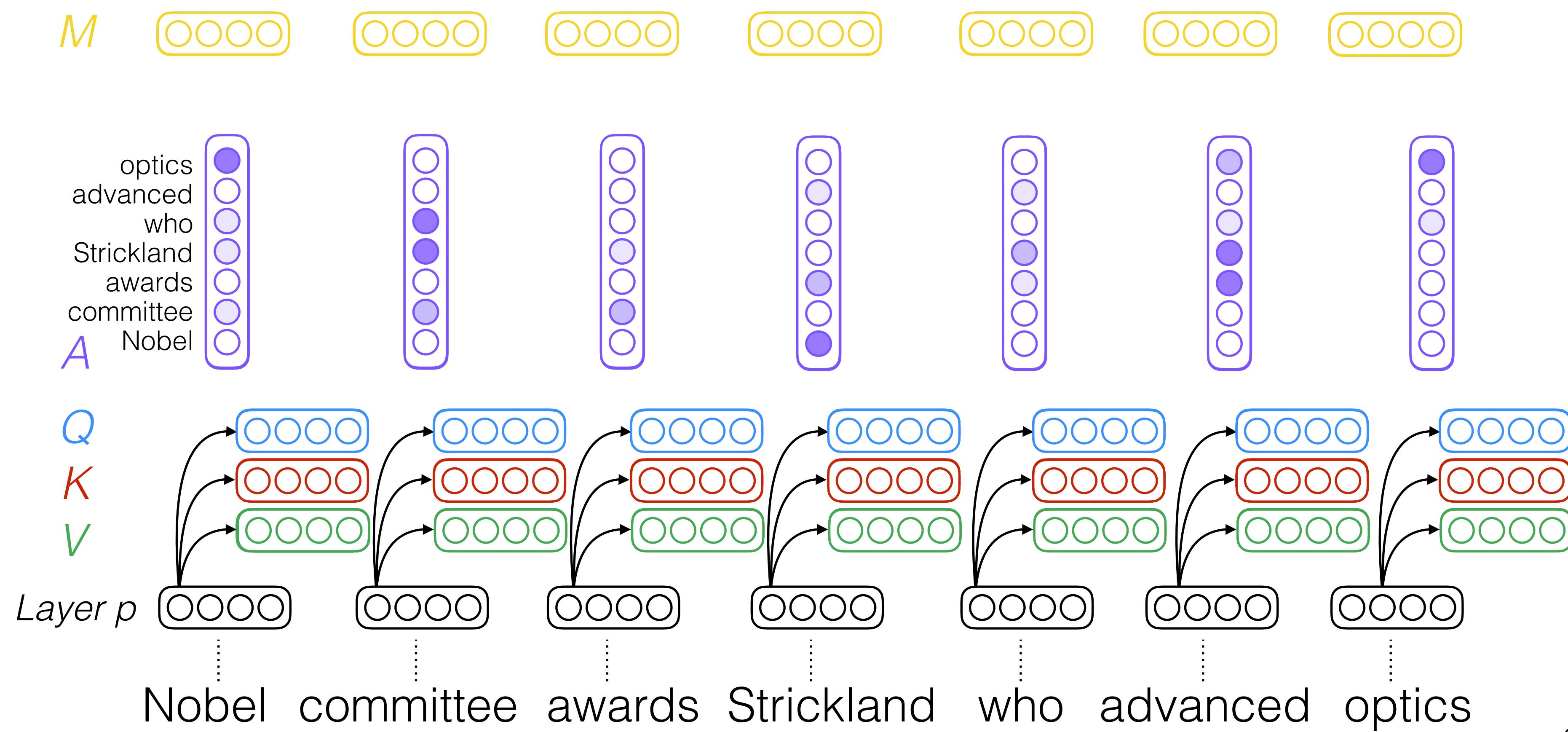
Self-attention



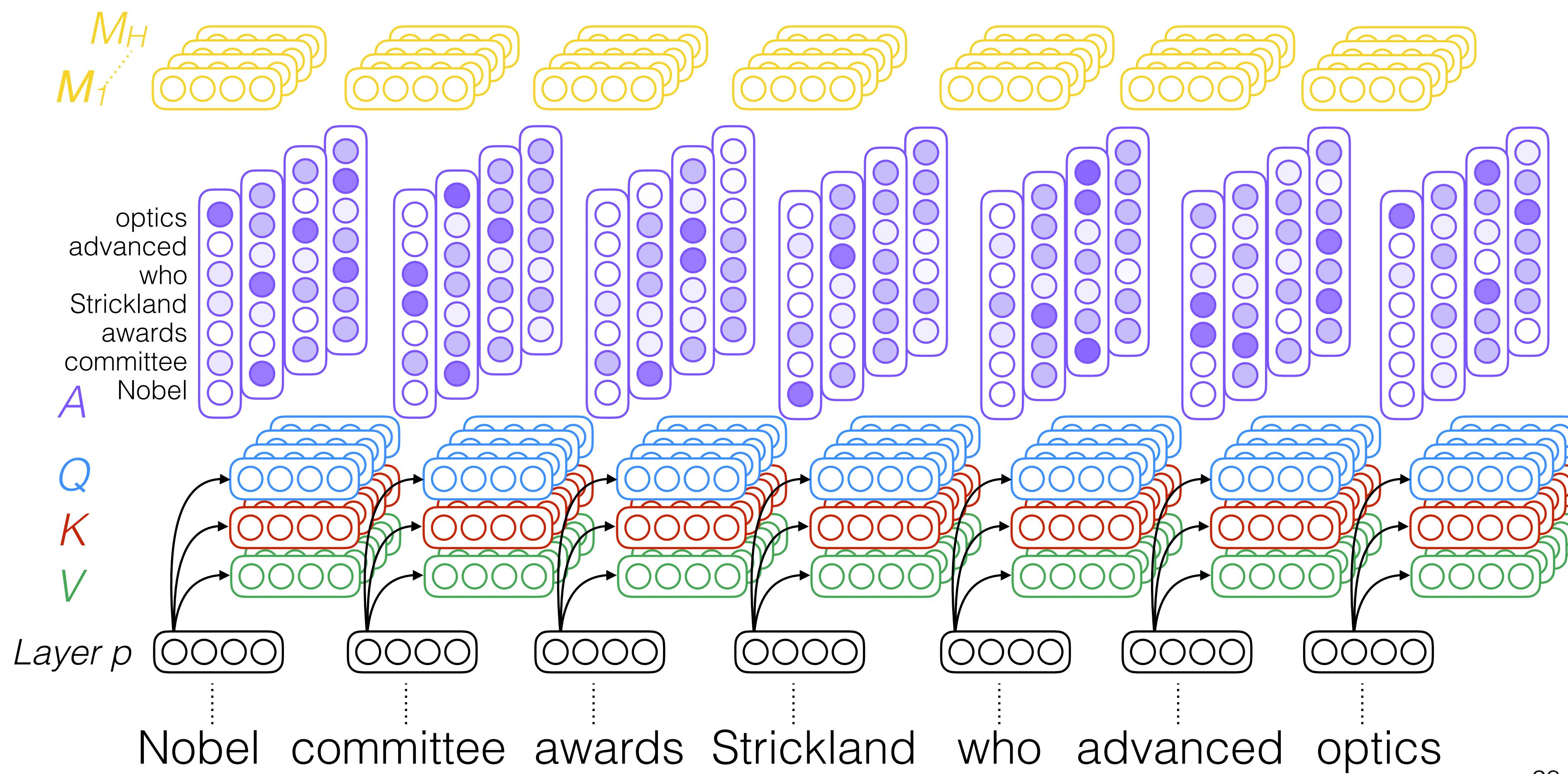
Self-attention



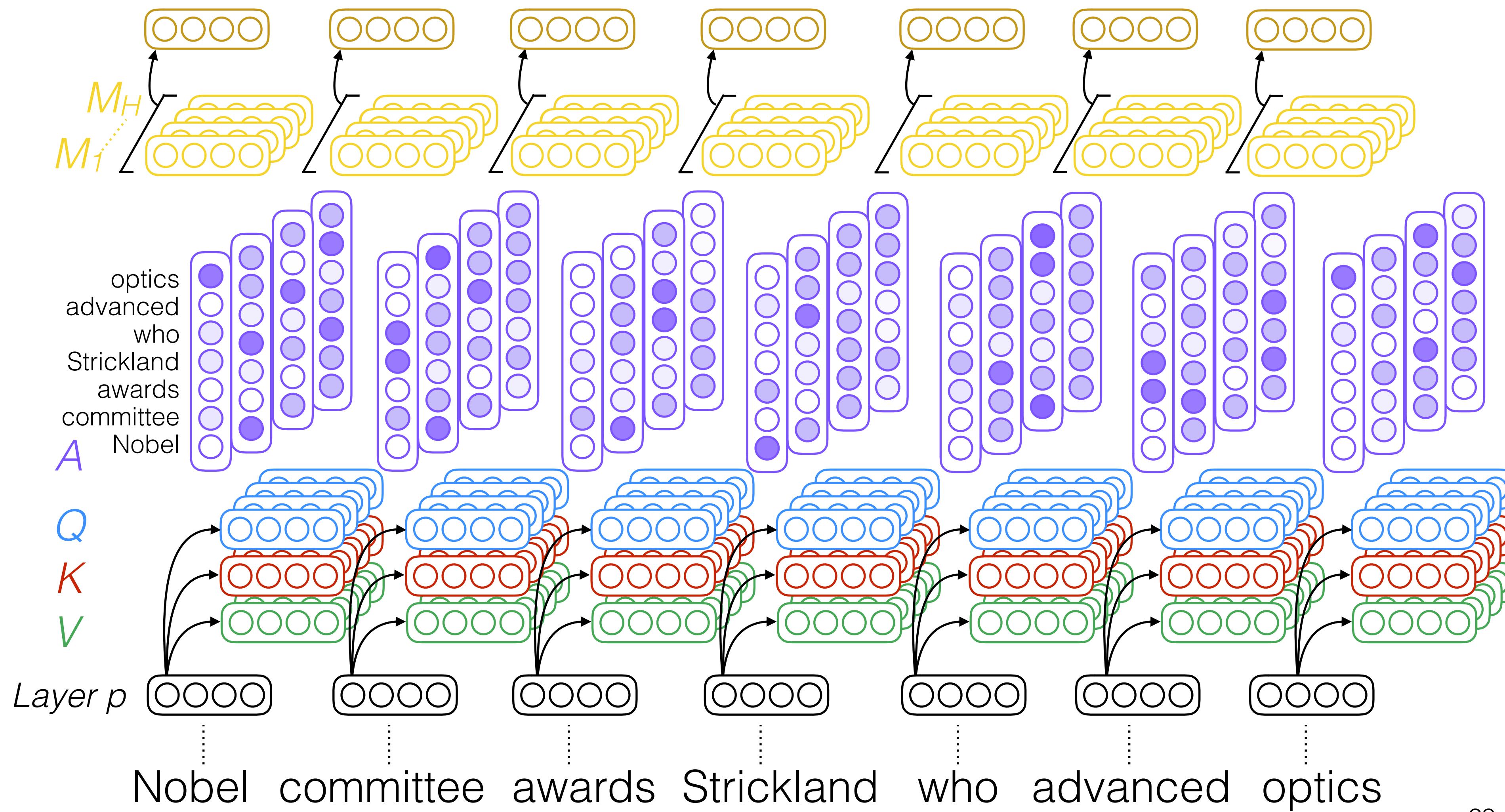
Self-attention



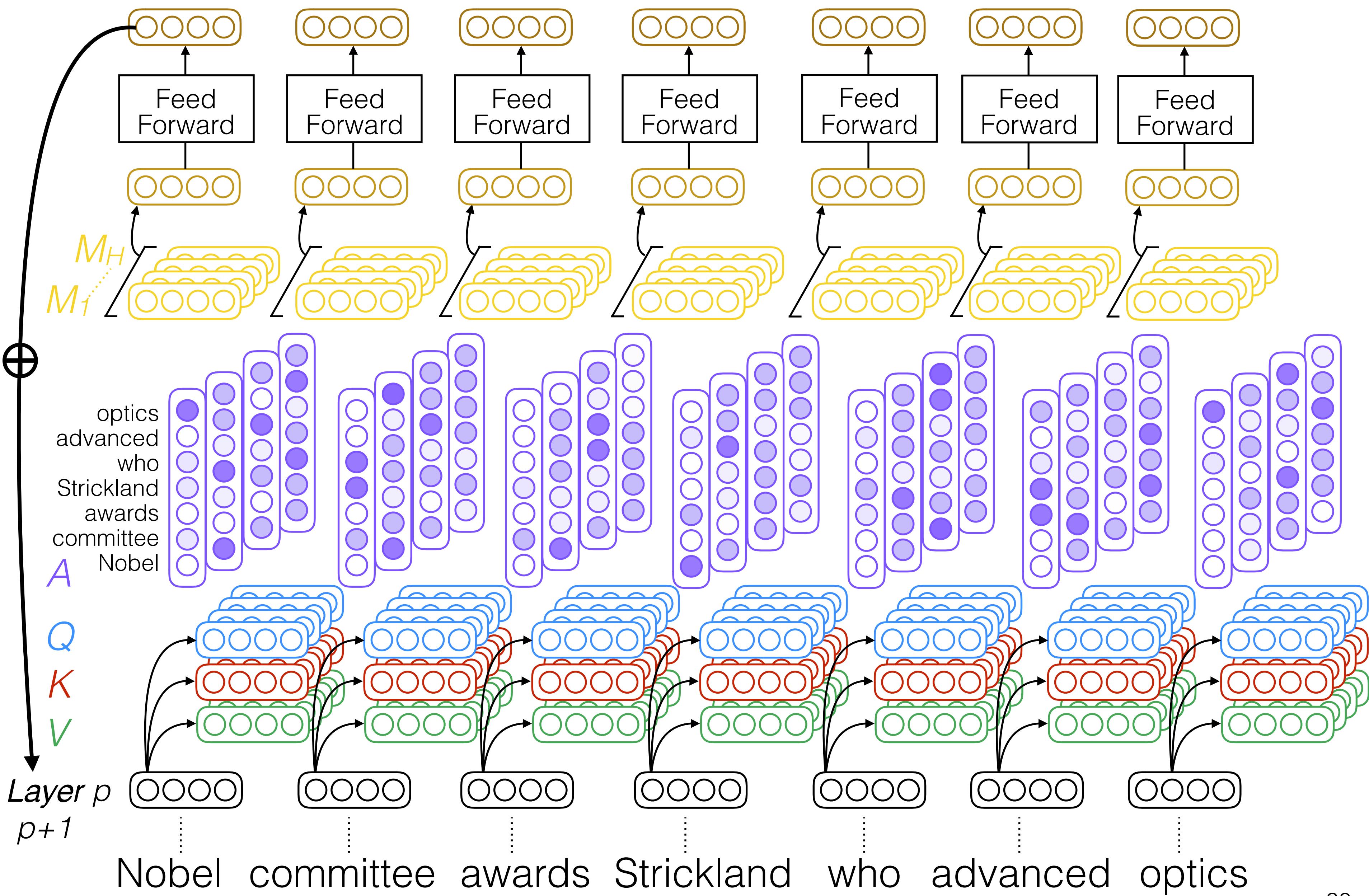
Multi-head self-attention



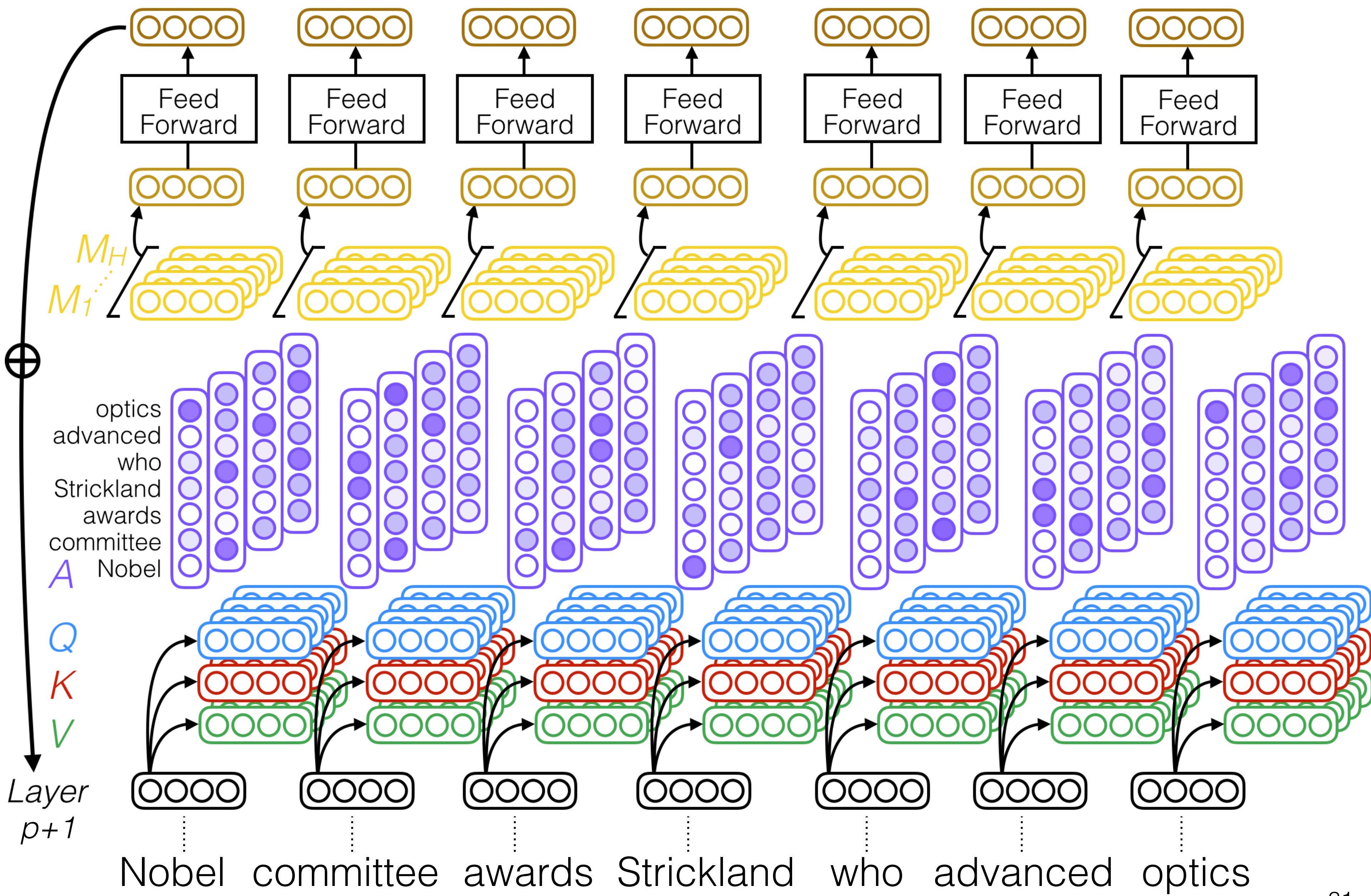
Multi-head self-attention



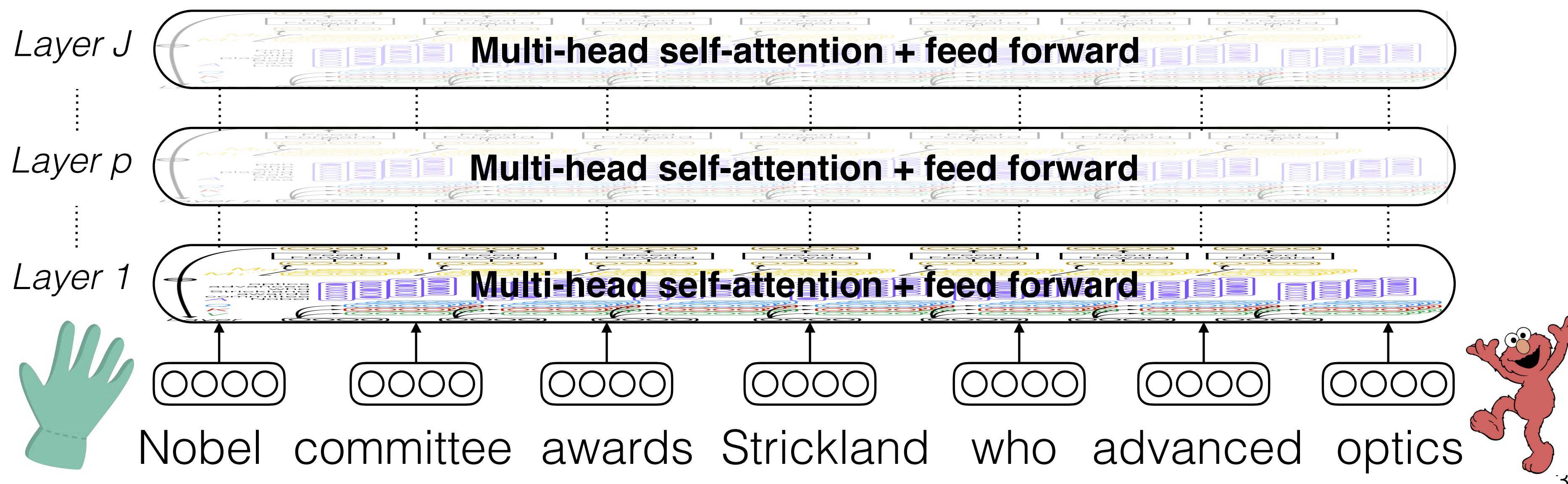
Multi-head self-attention



Multi-head self-attention



Multi-head self-attention



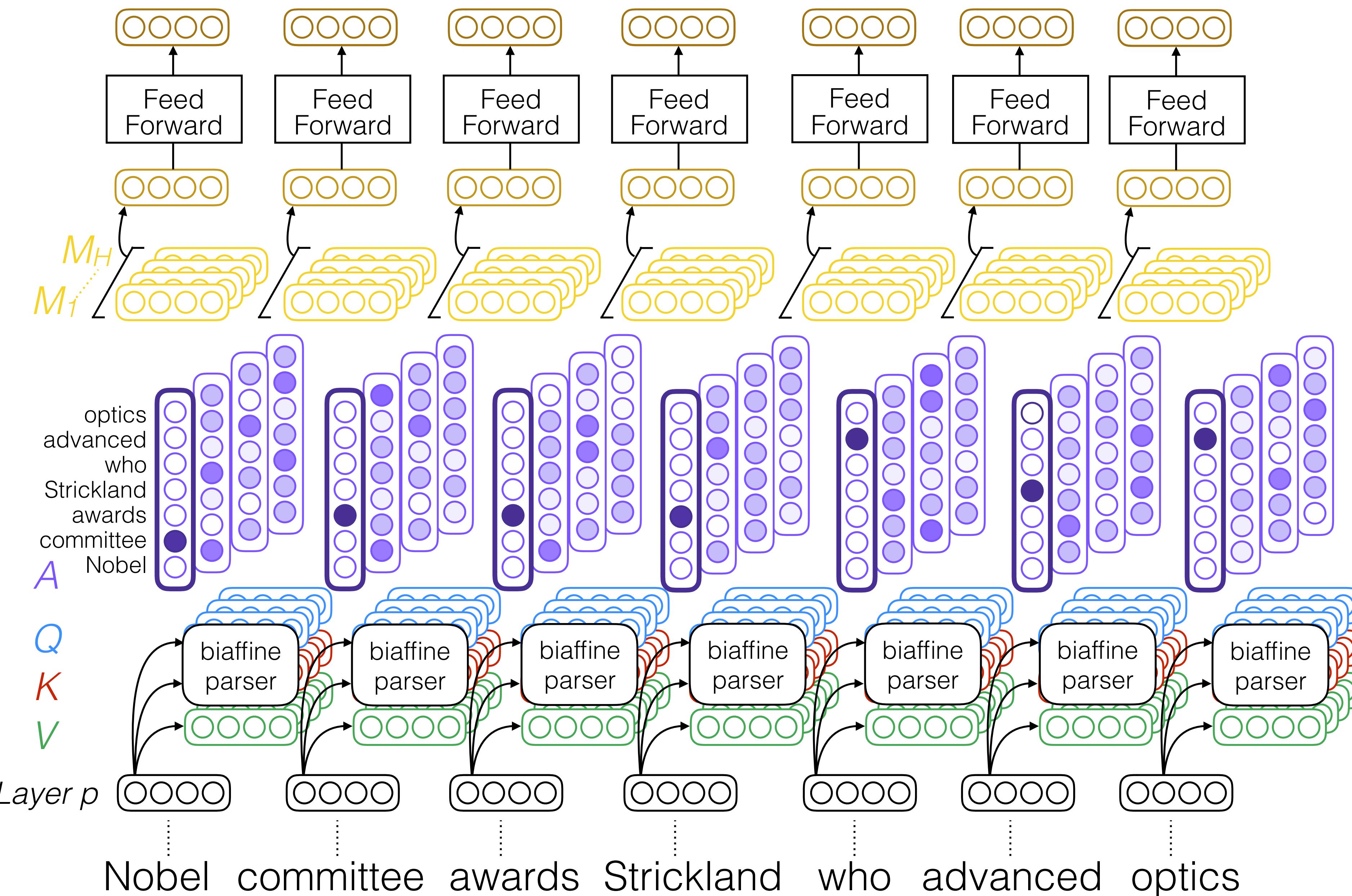
Outline

- Want fast, accurate, robust NLU
- PropBank SRL: Who did what to whom?
- 10 years of PropBank SRL
- LISA: Linguistically-informed self attention
 - Multi-head self-attention [Vaswani et al. 2017]
 - Syntactically-informed self-attention
 - Multi-task learning, single-pass inference
 - Experimental results & error analysis

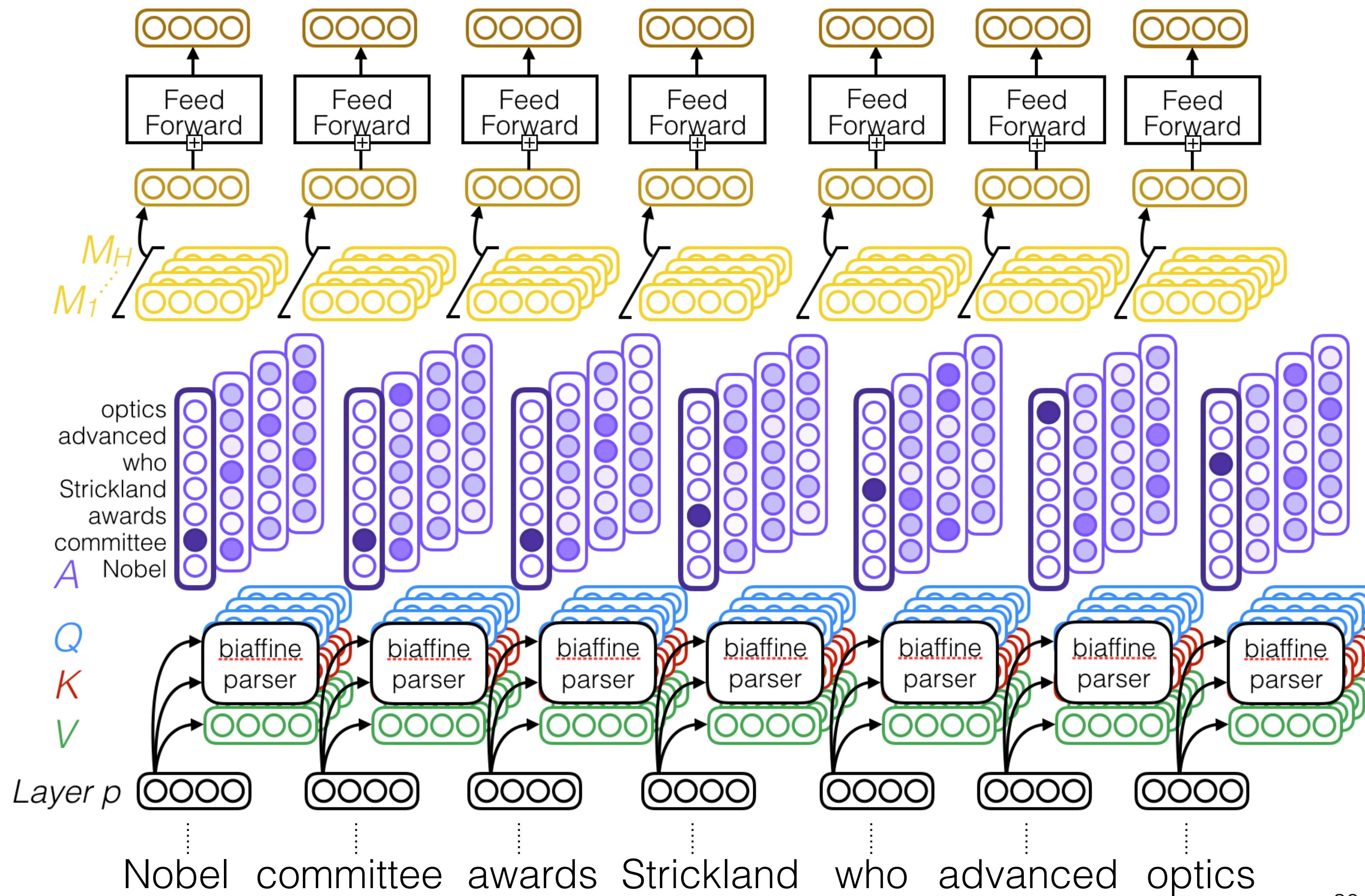
How to incorporate syntax?

- Multi-task learning [Caruana 1993; Collobert et al. 2011]:
 - Overfits to training domain like single-task end-to-end NN.
 - Must re-train SRL model to leverage new (improved) syntax.
- Dependency path embeddings [Roth & Lapata 2016];
Graph CNN over parse [Marcheggiani & Titov 2017]
 - Restricted context: path to predicate or fixed-width window.
- Syntactically-informed self-attention
 - In one head, token attends to its likely syntactic parent(s).
 - Global context: In next layer, tokens observe all other parents.
 - At test time: can use own predicted parse, *OR*
supply syntax to improve SRL model without re-training.

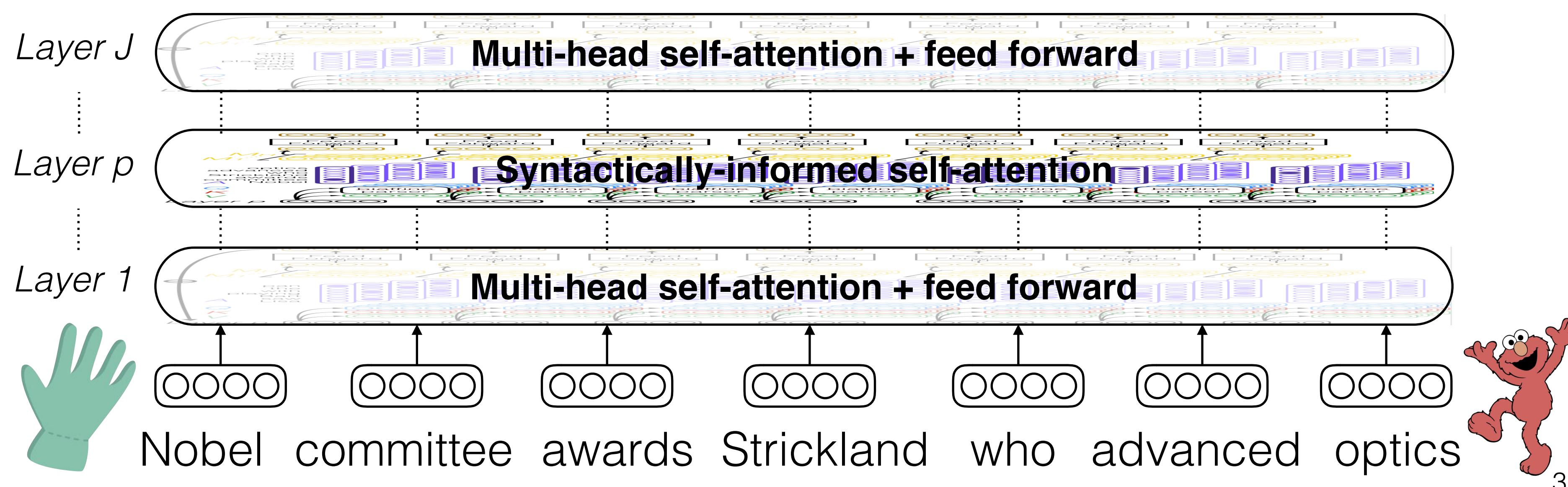
Syntactically-informed self-attention



Syntactically-informed self-attention



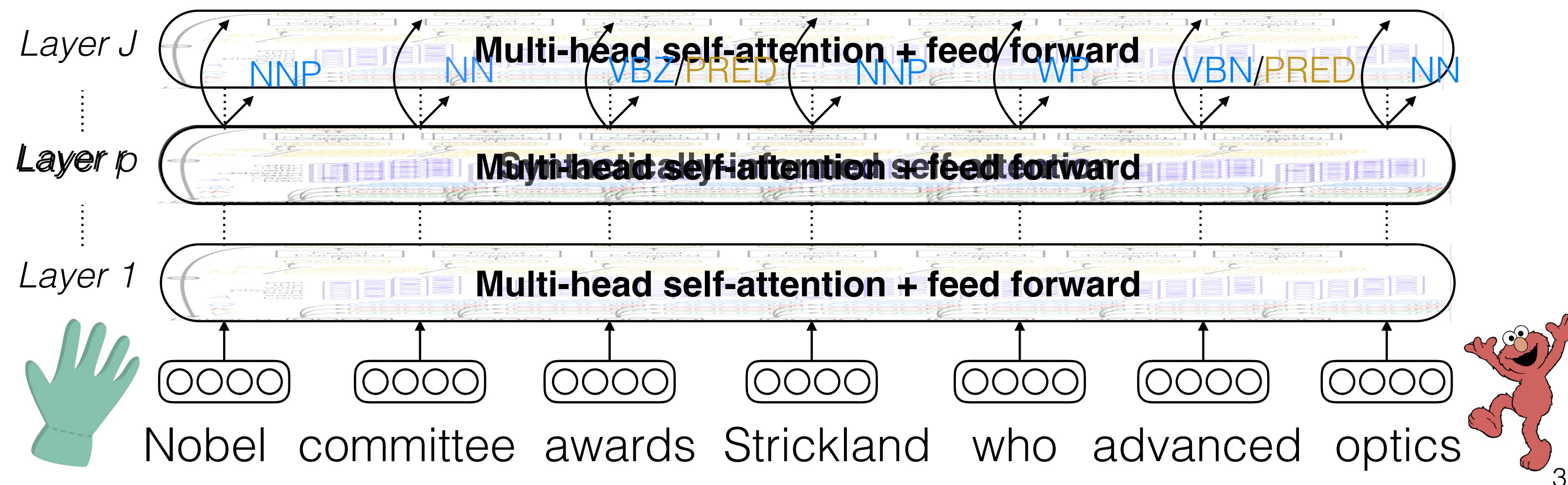
Syntactically-informed self-attention



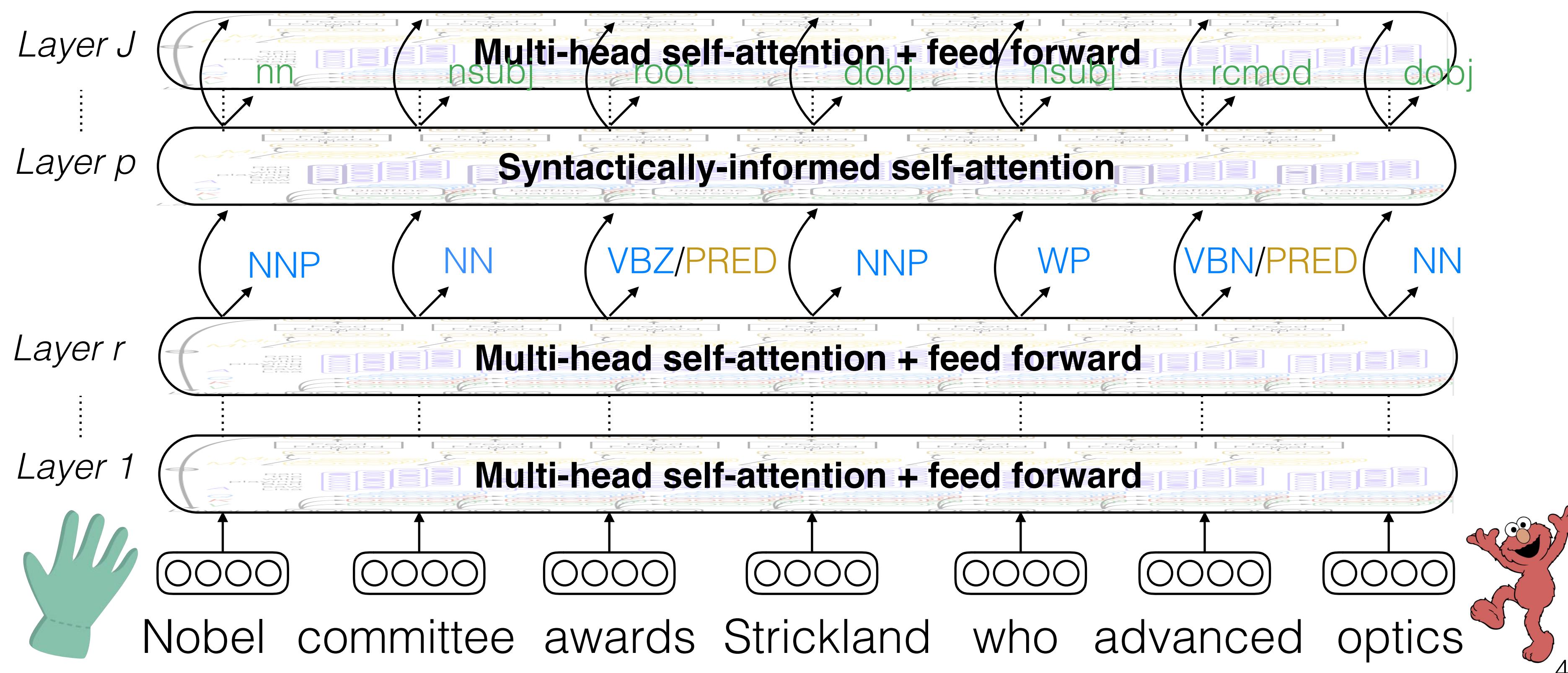
Outline

- Want fast, accurate, robust NLU
- PropBank SRL: Who did what to whom?
- 10 years of PropBank SRL
- LISA: Linguistically-informed self attention
 - Multi-head self-attention
 - Syntactically-informed self-attention
- Multi-task learning, single-pass inference
- Experimental results & error analysis

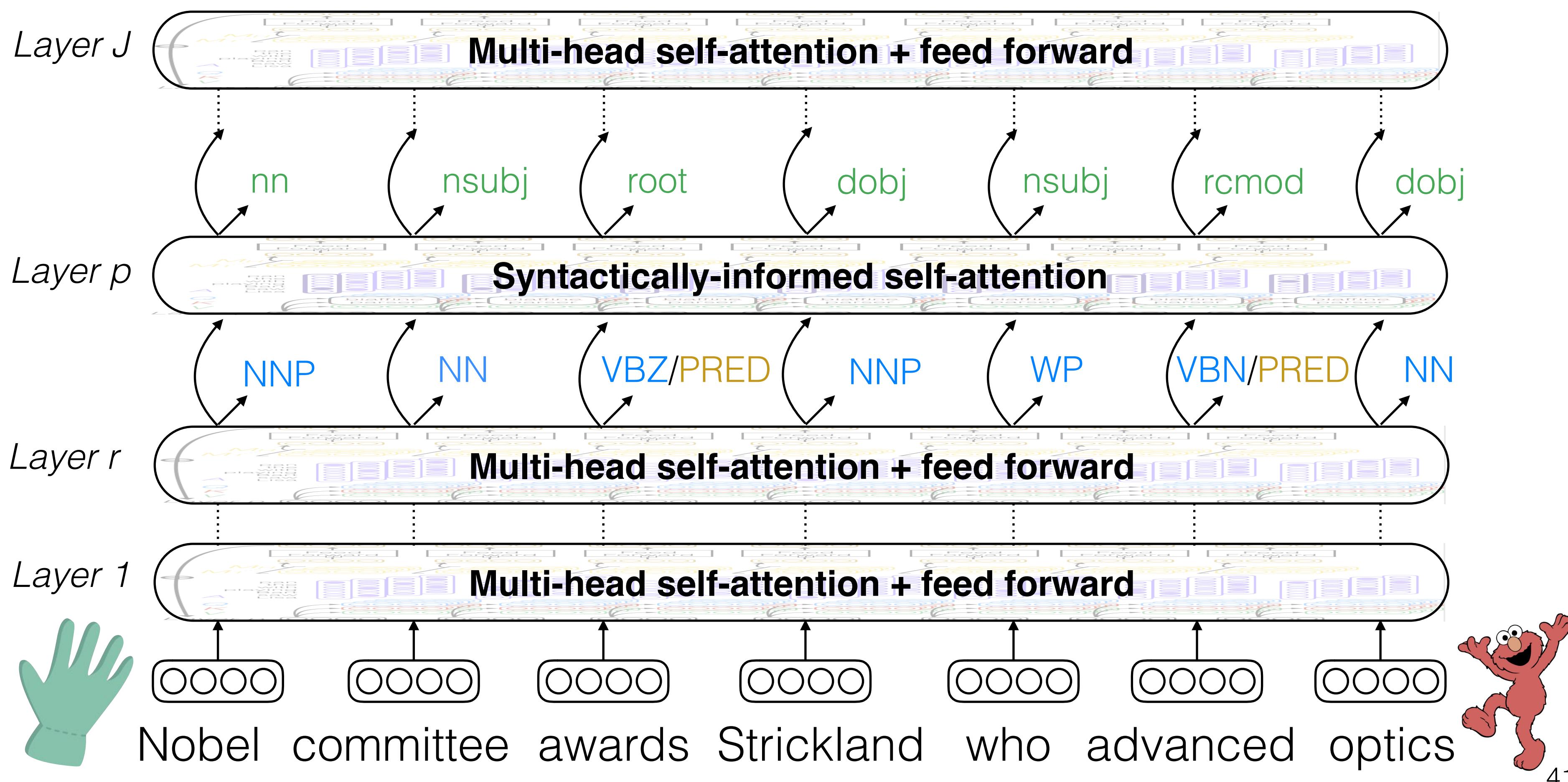
LISA: Linguistically-Informed Self-Attention



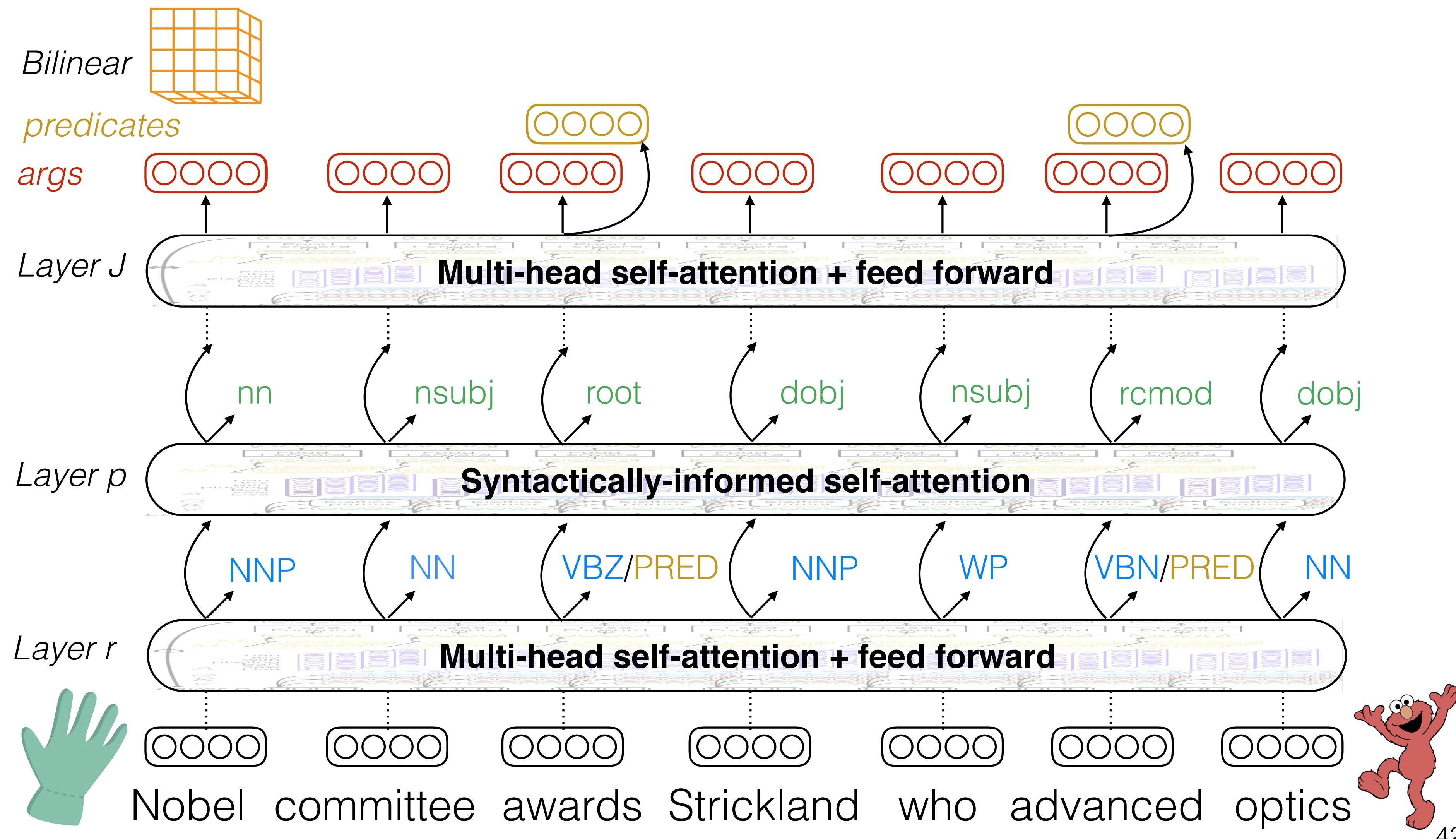
LISA: Linguistically-Informed Self-Attention



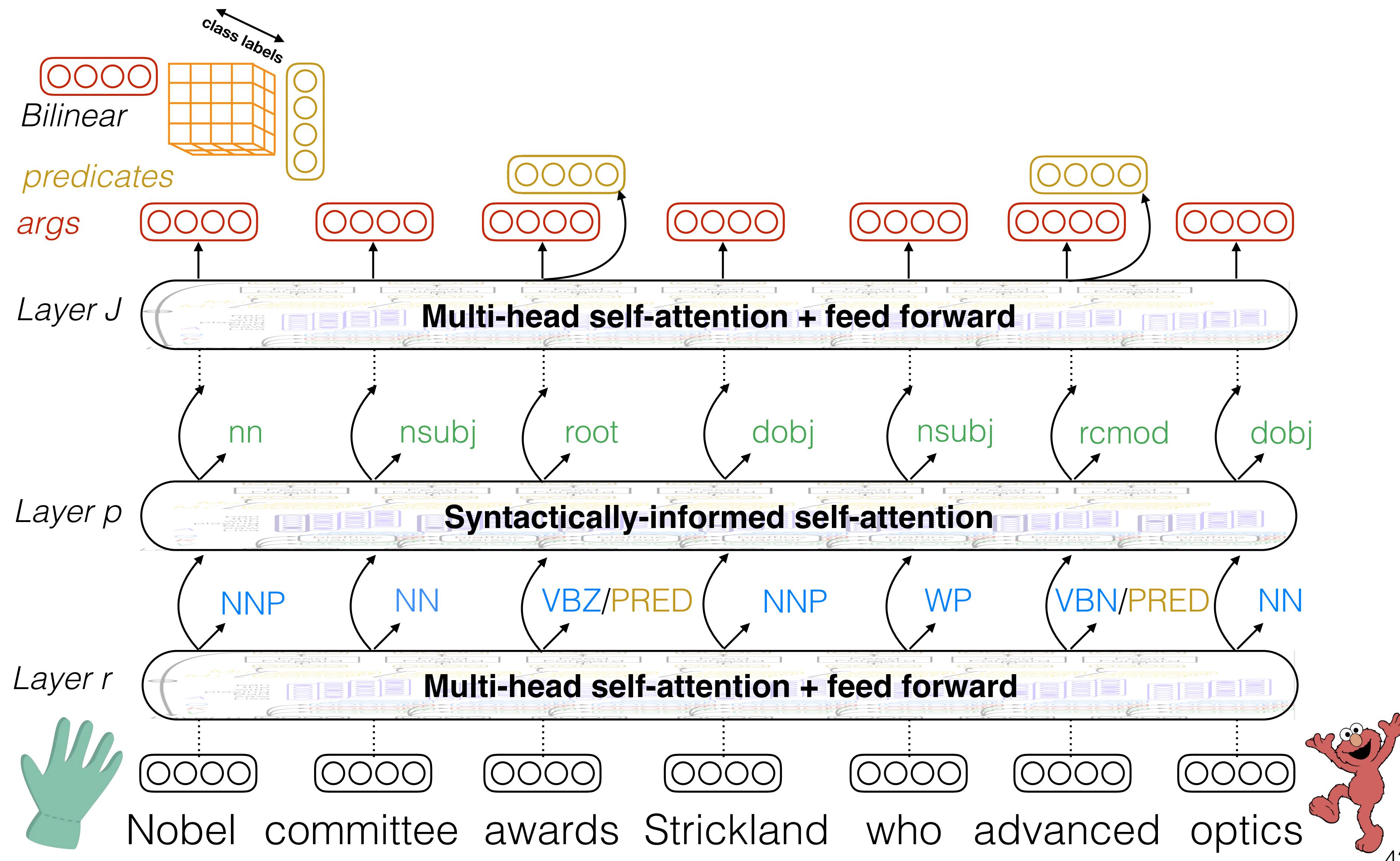
LISA: Linguistically-Informed Self-Attention



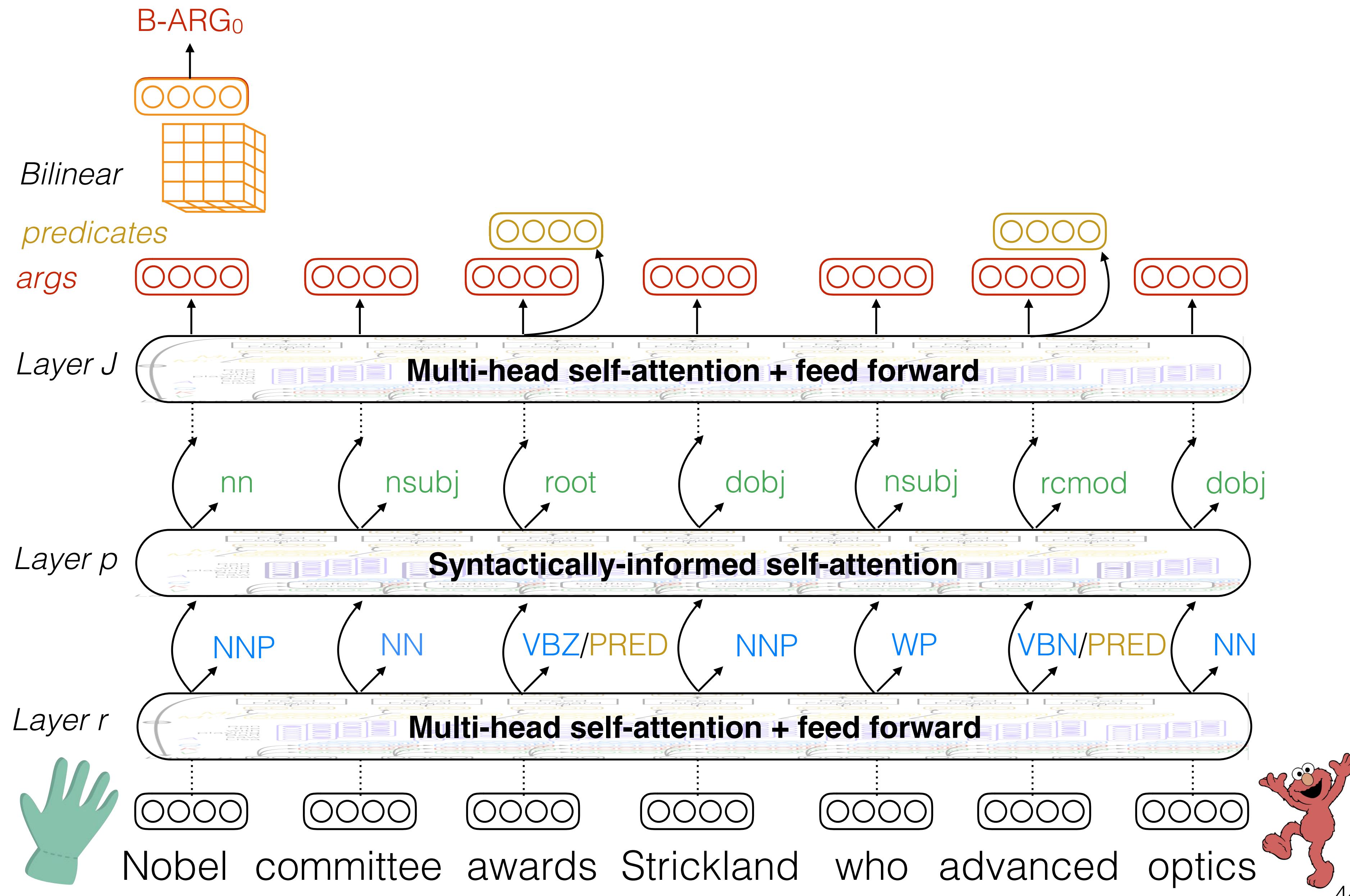
LISA: Linguistically-Informed Self-Attention



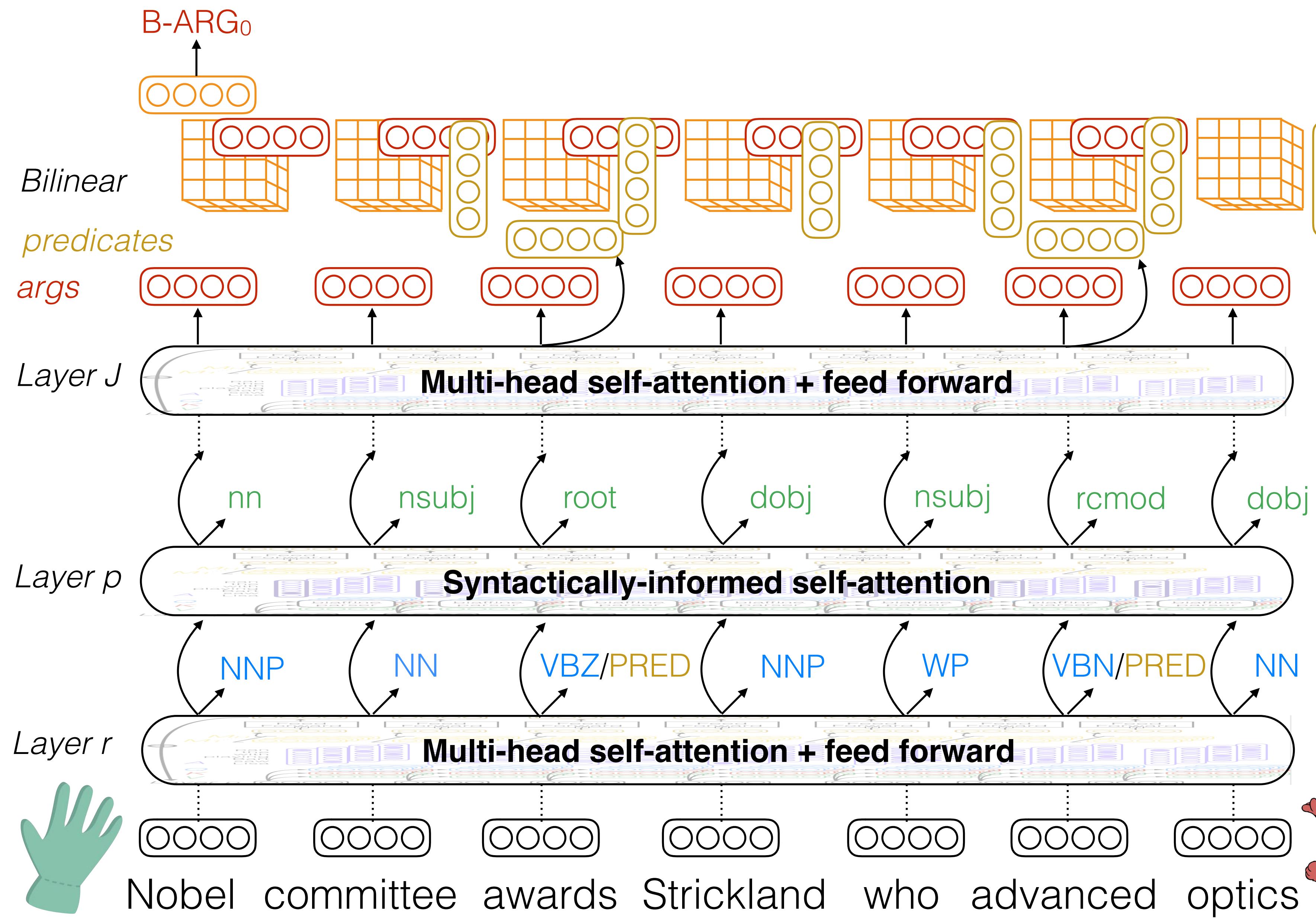
LISA: Linguistically-Informed Self-Attention



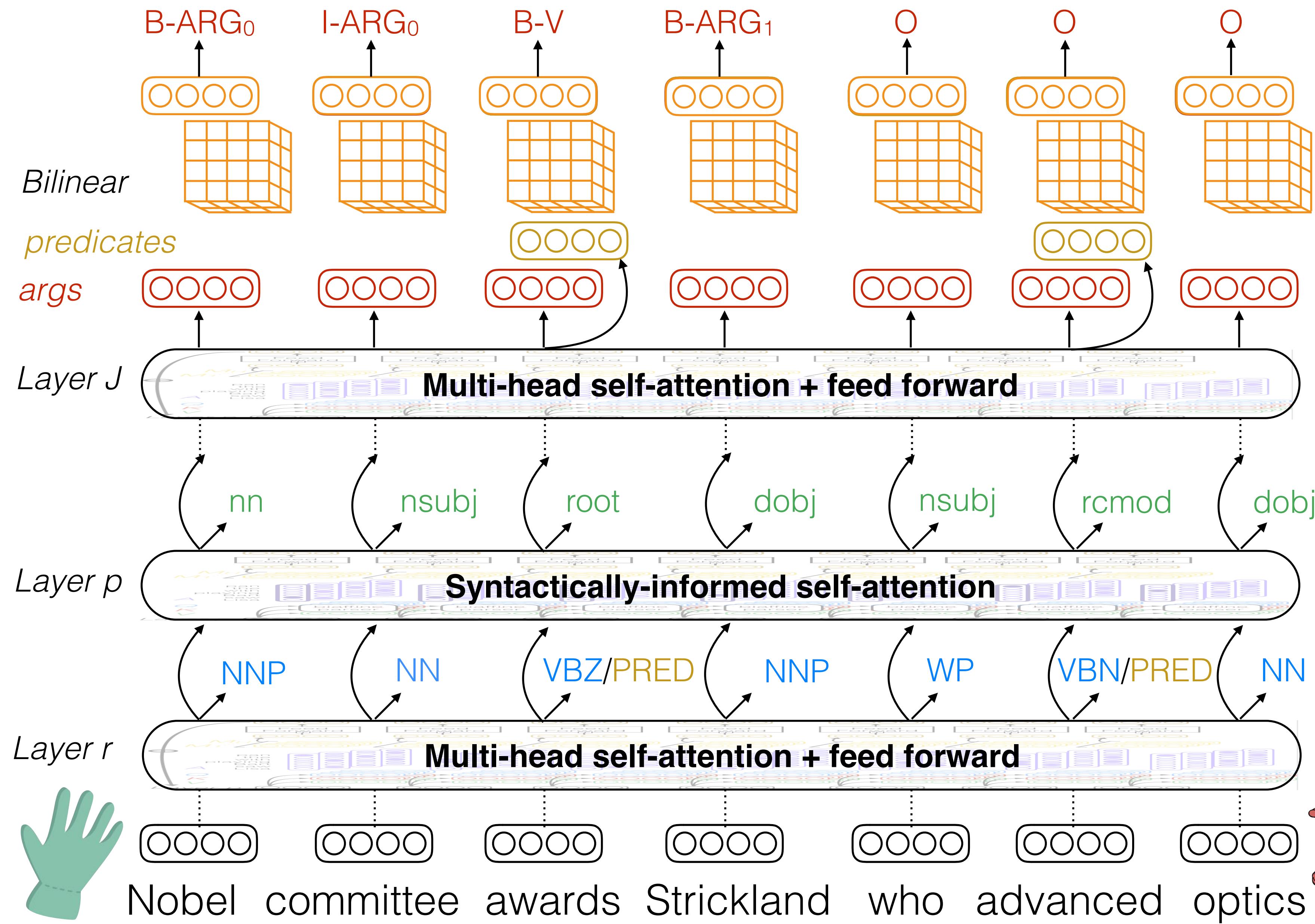
LISA: Linguistically-Informed Self-Attention



LISA: Linguistically-Informed Self-Attention

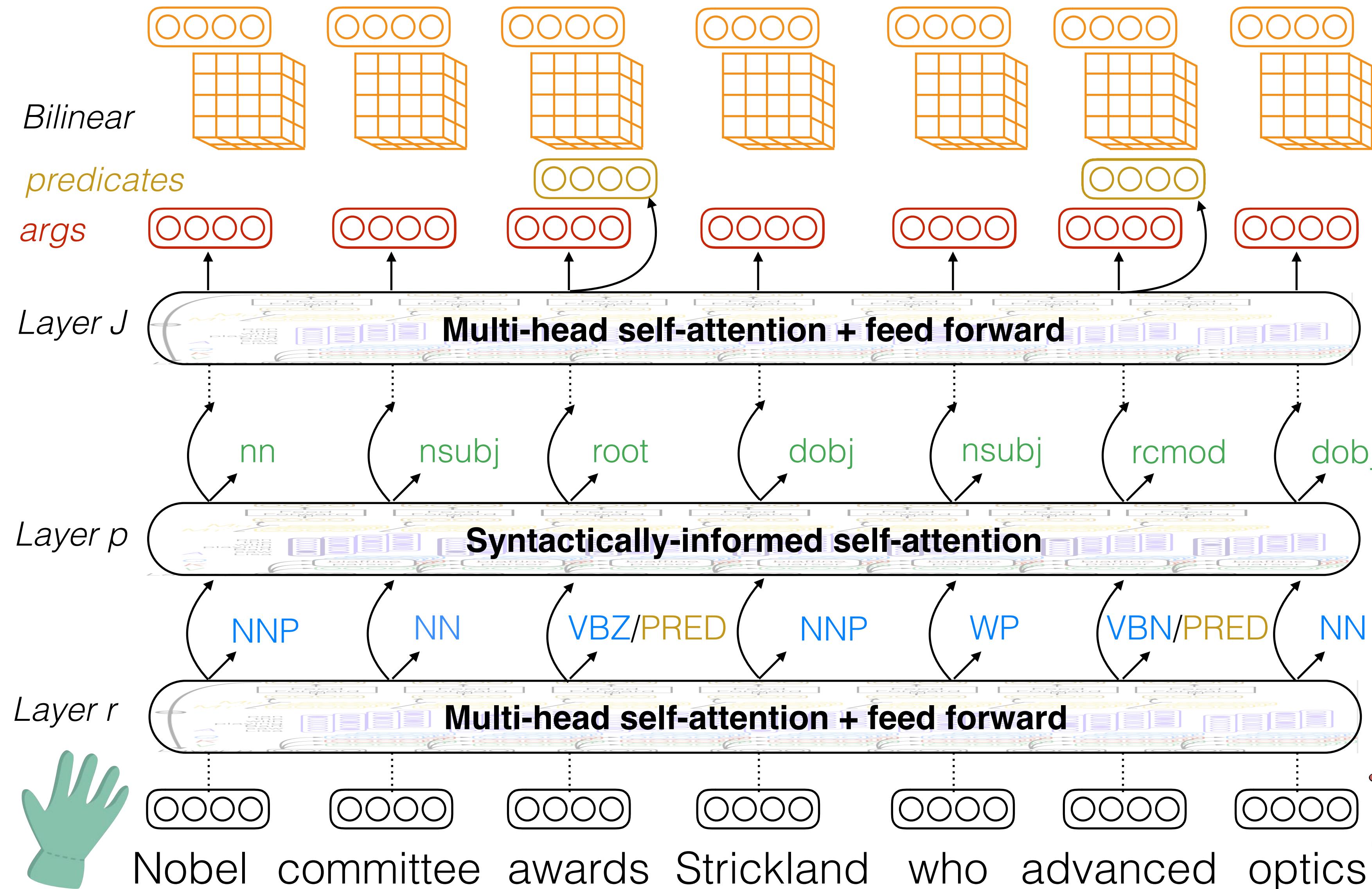


LISA: Linguistically-Informed Self-Attention



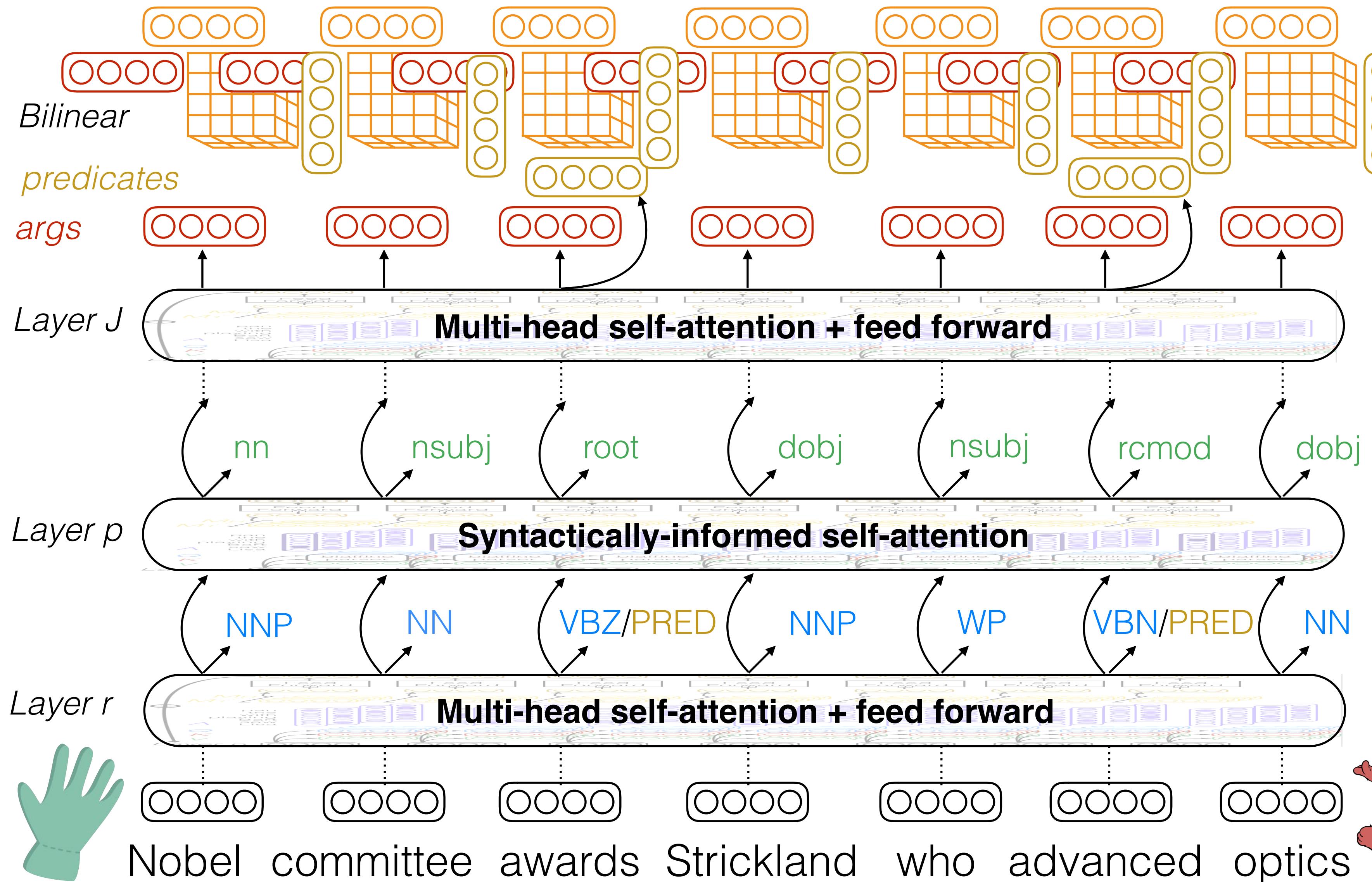
LISA: Linguistically-Informed Self-Attention

B-ARG₀ I-ARG₀ B-V B-ARG₁ O O O

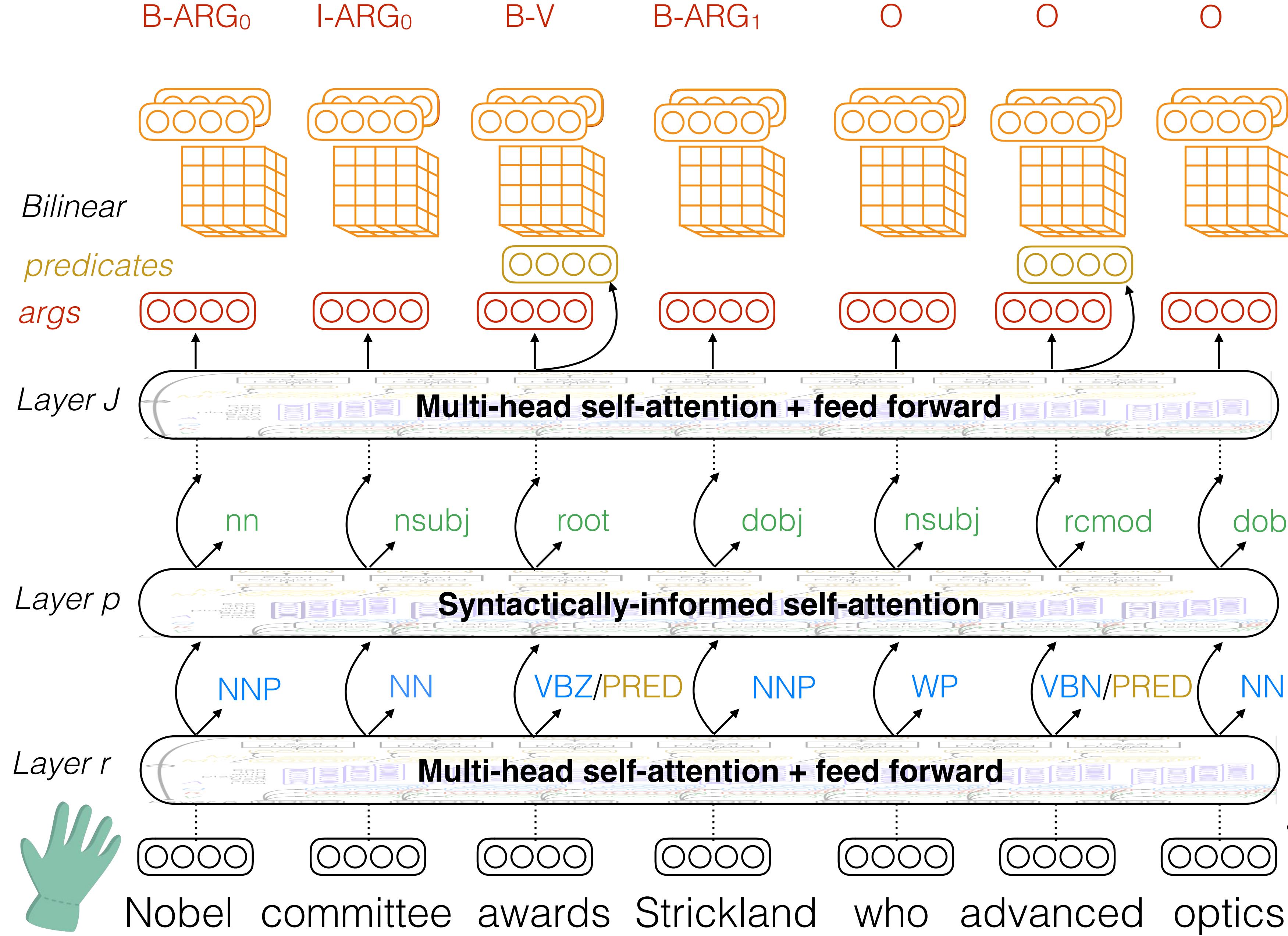


LISA: Linguistically-Informed Self-Attention

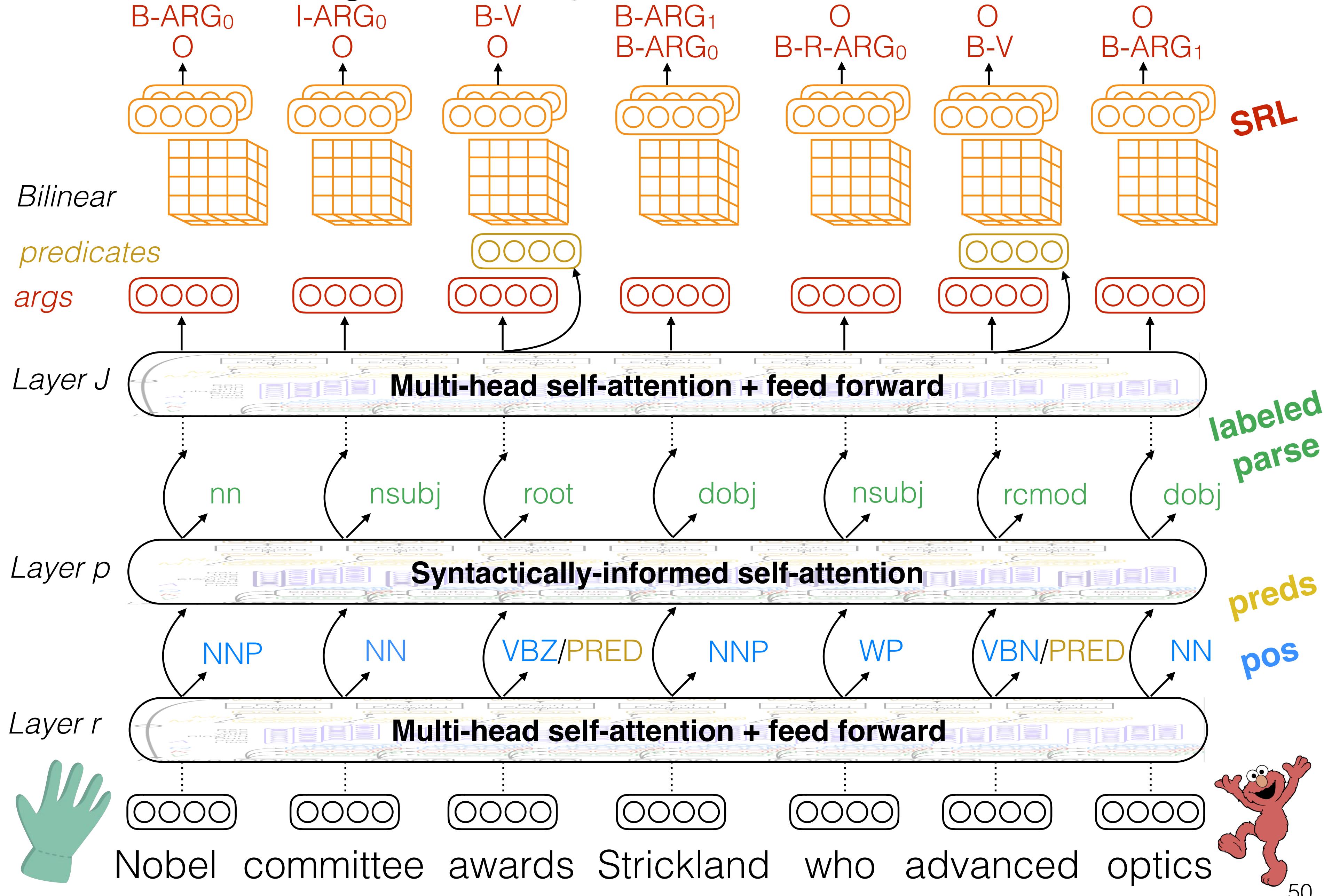
B-ARG₀ I-ARG₀ B-V B-ARG₁ O O O



LISA: Linguistically-Informed Self-Attention



LISA: Linguistically-Informed Self-Attention



Outline

- Want fast, accurate, robust NLU
- PropBank SRL: Who did what to whom?
- 10 years of PropBank SRL
- LISA: Linguistically-informed self attention
 - Multi-head self-attention
 - Syntactically-informed self-attention
- Multi-task learning, single-pass inference
- Experimental results & error analysis

Experimental results

	CoNLL-2005	CoNLL-2012
domains	Train, dev: news Test: news, novels	Train, dev, test: 7 domains (news, telephone, bible, ...)
word embeddings	GloVe [Pennington et al. 2014] ELMo [Peters et al. 2018]	GloVe [Pennington et al. 2014] ELMo [Peters et al. 2018]
predicates	predicted; gold	predicted
baselines	He et al. 2017 He et al. 2018 Tan et al. 2018	He et al. 2018
our models	SA LISA LISA+D&M, +Gold	SA LISA LISA+D&M, +Gold

Experimental results: CoNLL-2005

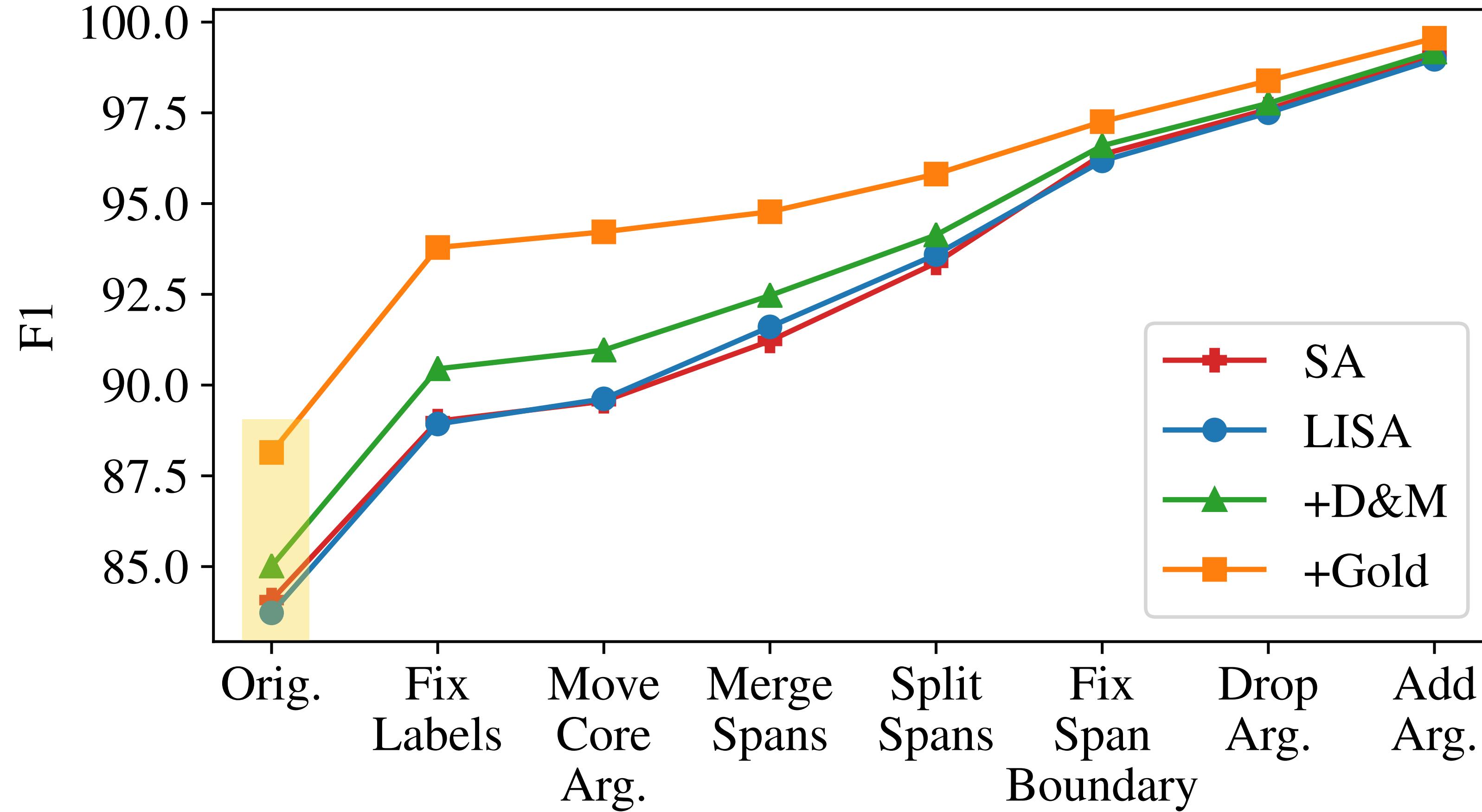
	GloVe	ELMo		
	in-domain	out-of-domain	in-domain	out-of-domain
He et al. 2017	82.7	70.1	---	---
He et al. 2018	82.5	70.8	86.0	76.1
SA	83.72	71.51	86.09	76.35
LISA	83.61	71.91	86.55	78.05
+D&M	94.9 UAS 84.99 F1 <i>+2.49 F1</i>	90.3 UAS 74.66 F1 <i>+3.86 F1</i>	96.3 UAS 86.90 F1 <i>+0.9 F1</i>	93.4 UAS 78.25 F1 <i>+2.15 F1</i>

?

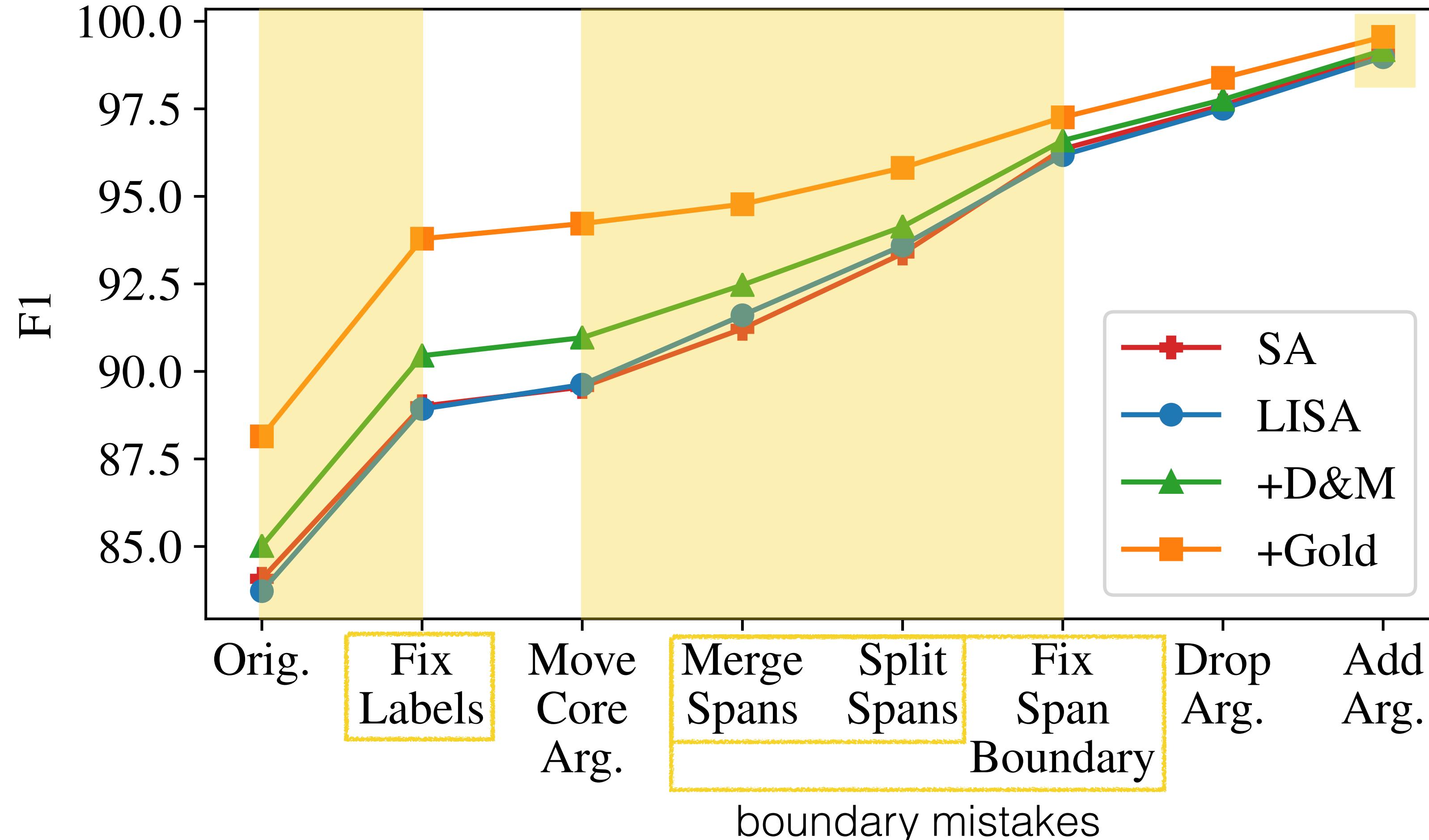
Experimental results: CoNLL-2005



Experimental results: Analysis



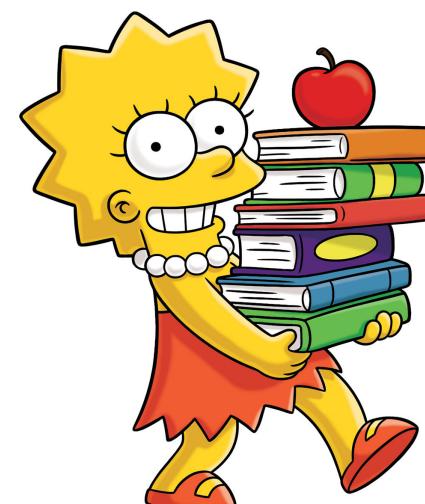
Experimental results: Analysis



Summary

Thank you!

- **LISA**: Multi-task learning + multi-head self attention trained to attend to syntactic parents
 - Achieves state-of-the-art F1 on PropBank SRL
 - Linguistic structure improves generalization
 - Fast: encodes sequence *only once* to predict predicates, parts-of-speech, labeled dependency parse, SRL
- Everyone wants to run NLP on the entire web:
 - **accuracy, robustness, computational efficiency.**



Models & Code: <https://github.com/strubell/LISA>